eman ta zabal zazu

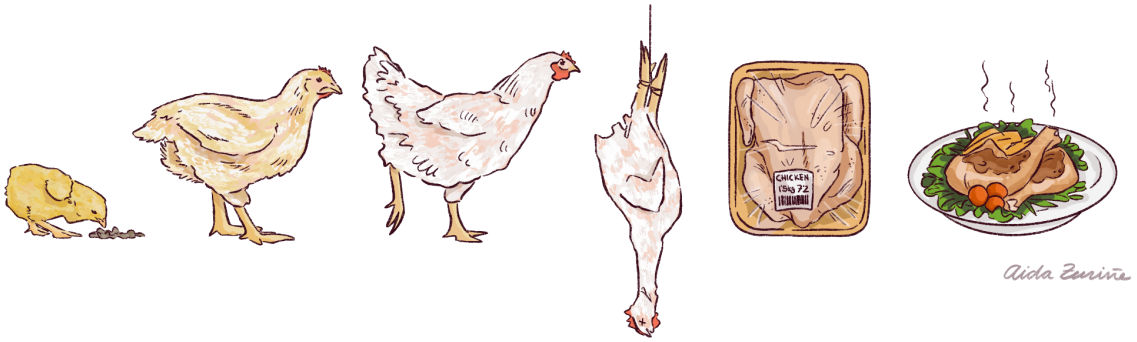
Universidad del País Vasco

Euskal Herriko Unibertsitatea

A holo-omic approach towards a more sustainable broiler chicken production

University of the Basque Country

Leioa, 2023



PhD Thesis by

Sofia Marcos Basagoiti

eman ta zabal zazu

Universidad del País Vasco

Euskal Herriko Unibertsitatea

A holo-omic approach towards a more sustainable broiler chicken production

PhD Thesis
Sofia Marcos Basagoiti

Supervisors
Andone Estonba Rekalde & Antton Alberdi Estibaritz

This thesis has been submitted to the doctoral programme in Molecular Biology and Biomedicine at the University of the Basque Country.

Leioa, 2023

Acknowledgements

I would like to thank my two supervisors, Andone and Antton, for your dedication in the development of this doctoral thesis and HoloFood. You trusted me when I had just turned 23 to participate in such an interdisciplinary and demanding project as HoloFood. Without your involvement in planning, analysis and interpretation of this thesis, it would not have come this far. Especially Antton, for your innate optimistic vision, effortless creativity and 24/7 availability.

I would also like to give special thanks to Joan, Sarah, Melanie, Iñaki and Jorge, who have supervised me throughout my PhD journey. Thanks to Joan, for teaching me so many things in the experimental phase of the project and for always being willing to collaborate and supervise my texts. Gràcies per preocupar-te també per la meva salut. To Sarah, for teaching me that every second in the lab is crucial, but also reminding me how important it is to make 5 min breaks. To Melanie, for teaching me the basics of programming. Special thanks to Iñaki, for your eternal patience and dedication in the last phase of data analysis. Eskerrik asko ere komunitate zientifikoan lan erritmo, bizi eredu eta izakera anitz daudela gogorarazteagatik. To Jorge, all the clarity in my code is because of you.

To HoloFood partners, and specially to the Holochicken team, for making long sampling, laboratory, and meeting days entertaining and exciting. To all the colleagues that at some point helped in the project. Specially to Garazi Martin and Louisa Pless, for the laboratory work, as well as to Germana Baldi and Varsha Kale, for the computer work. To Eray, for accepting to perform the preliminary analyses of host transcriptomics. To Sigrid (et *al*.), for helping me label and fill more than 10,000 tubes for the samplings. To Anna, for guidance. Without this massive effort, this thesis would simply not exist.

Mange tak to the first and actual generation of colleagues from Alberdi lab, for welcoming me so warmly into the group. To Jacob, for teaching me so much stuff. To Lasse, for sharing his wise thoughts and to Adam, for all the support that first year when I felt so baby yet. To the section of Evolutionary Genomics, for the fun plans during my stay in Copenhagen.

To the Applied Genomics and Bioinformatics lab, to those who left and those who continue, for putting up with me on those days when I can't even stand myself. To the sporadic meals with the colleagues from the EHU genetics department before the pandemic and the colleagues from Maria Goyri for the fun lunch and coffee breaks.

Eskerrak eta barkamenak lagunei, urte hauetan nire gehiegizko falta edo presentzia jasan duzuen arren beti hor egoteagatik. Kuadrilakoei, anitz eta berezi, baina berezi denak merezi. Laukote feministari, hizketaldi luzeengatik. Andreari, asteburuengatik. Iruneri, mendi txangoegatik.

Gurasoei, beti hobetzeko aukera dagoela erakusteagatik. Ireneri, tesi honek amaiera zuela ikusarazigatik. Aitorreri, nork esango lukelako nik amaituko nuela biologia ikasten, txikitan zu baitzinen etxeko naturalista.

This research was funded by the European Union's Horizon Research and Innovation Programme under grant agreement No. 817729.

My PhD was supported by the Basque Government doctoral fellowship No. PRE_2019_1_0264 for 4 years.

The cover illustration was done by Aida Zuriñe.

Table of contents

Summary

The coming decades will see a substantial increase in food production to meet growing demand, raising the importance of prioritising sustainable and safe production practices on a global scale. In this context, chicken emerges as the most efficient meat production system with the lowest impact on climate change. However, the intensification of poultry meat production faces the challenge of achieving rapid and predictable animal growth without compromising public health, animal welfare and the environment. Antibiotics are administered as growth promoters, but due to concerns about their overuse and the increasing awareness of the pivotal role of host-associated microorganisms in the biological functions of animals, feed additives that modulate gut microbiota have been proposed as alternative options. Nevertheless, their efficacy varies considerably due to our limited knowledge of the dynamics of the resident gut microbiota in farm animals, and their interaction with the host and dietary treatments. These limitations can now be addressed from a hologenomic perspective by analysing multiple omic layers of host-microbiota domains.

In this thesis, I introduce a holo-omic perspective to study host-microbiota interactions in intensively produced broiler chickens in relation to performance. I report the results from 3 identical experimental trials we conducted in 2019, in which chickens from 2 genetic lines and both sexes were grown under 3 dietary treatments, all closely monitored throughout their production period. The dissertation encompasses 3 original research articles that summarise the collaborative work I have carried out with the rest of partners within the H2020 project HoloFood. We initially assess the impact of the experimental groups on chicken performance, finding negligible effects of the tested dietary additives. However, we identify a high inter-individual variation in chicken growth performance for the first two trials, and an opportunistic colonisation of *Campylobacter spp.* linked to a reduced chicken growth in the third trial. The dissertation revolves around the hypothesis that both performance trends can be partly explained by the interplay of the host and associated gut microbial communities. Accordingly, the following chapters apply holo-omics to respond to the mentioned questions. In the second article, we delve into characterising the functional dynamics of genome-resolved bacterial strains in the gut microbiome development by using metagenomics and metatranscriptomics. Our findings reveal that gut microbiome transitions from high functional capacity to low-capacity bacteria as chickens grow, with a more pronounced transition positively associated with chicken growth. In the third article, we examine the microbial dynamics and host response to *Campylobacter* colonisation to unveil the molecular reasons of the reduced weight gain using metagenomics and (meta)transcriptomics. We discover that an early *B. fragilis_A* spread facilitates *Campylobacter* colonisation by changing the functional profile of the microbial community with a higher and more specialised metabolic activity that prevents the host from gaining weight.

This thesis collectively advances our understanding of host-microbiota interactions in broiler chickens, opening up new avenues in the search for microbe-based solutions to enhance animal production in a sustainable manner.

Laburpena

Klima-aldaketaren ondorioz, gizarteak natura-baliabideen kontsumoaren eta ingurumeneko kontserbazio-lanen arteko oreka bilatzeko premia dauka. Oreka horren lorpenak lehantasuna dauka bereziki elikagai-industrian. Izan ere, giza-populazioaren hazkuntzak haragi kontsumoaren handipena eragin du, eta joera hori mantendu edo areagotuko da hurrengo urteetan. Haragi eskaerari aurre egiteko bermatzen den abeltzaintza intentsiboa bereziki kutsatzailea da elikagaien ekoizpenaren barruan. Beraz, garrantzitsua da mundu mailako ekoizpen-praktika jasangarri eta seguruak gureganatzea.

Azken hamarkadetan abere-ekoizpenean aurrerapen esangarriak egin dira zaintza, dieta eta osasun praktiketan eginiko hobekuntzekin. Hala ere, aldakortasun handia nabaritzen da errendimenduan nahiz eta populazio bereko indibiduoekin ihardun. Badakigu animalien digestio-aparatuaren osasuna funtsezkoa dela bere errendimendu egokirako. Beraz, interbentzio dietetikoak proposatu izan dira, antibiotikoak barne.

Hasiera batean antibiotikoak bakterio patogenoak mugatzeko garatu ziren arren, aparteko baliagarritasuna aurkitu zitzaien abeltzaintza intentsiboan, ikerketek erakutsi baitute dosi baxuagoetan hazkuntza-sustatzaile bezala jokatzen dutela. Nahiz eta ezezaguna izan nola bermatzen duten animalien errendimendua, uste da hesteko mikroorganismoen dentsitatea murrizten dutela, edota hesteko inflamazioa ekiditen. Baina haien gehiegizko erabilerak abereen ekoizpenean, antibiotikoekiko erresistentzia kasuak ekarri ditu gizartera, osasun publikorako arazo bat bilakatuz. Ondorioz, Europak hazkuntza-sustatzaile bezala erabiltzen diren antibiotikoak debekatu zituen 2006. urtean.

Antibiotikoen ordezko gisa, bestelako elikagai-gehigarriak proposatu izan dira, hots, animaliak digeritu ezin dituen baina haren osasun eta errendimendurako onuragarriak diren produktu naturalak. Haien artean fitobiotikoak, probiotikoak eta prebiotikoak aurkitzen dira. Fitobiotikoak eta prebiotikoak landareengandik deribaturiko produktuak dira. Probiotikoak, aldiz, ostalariarentzat onuragarriak diren mikroorganismo andui batez edo ugariz osaturik daude. Fitobiotikoak orokorrean antibiotikoak bezala jokatzen dute, biek propietate antimikrobianoak baitituzte. Probiotikoak eta prebiotikoak, aldiz, hesteko mikrobiota modulatzaile kontsideratzen dira. Azken gehigarri hauen erabilera izugarri orokortu da abere-ekoizpen industrian, azken hamarkadetan hesteko mikrobiotak ostalarian dituen onurak detaile gehiagoz ezagutu baitira.

Baina elikagai-gehigarri hauen eraginkortasuna oso aldakorra produkzio-sistema batetik bestera. Izan ere, hesteko mikrobiotaren dinamika eta harek ostalariarekin eta dietarekin dituen elkarreraginen inguruan dagoen ulermen mugatuak gehigarri eraginkorren diseinua galarazten du. Muga horiei aurre egiteko, hesteko ingurumen-baldintzetan gertatzen diren

mikrobio-mikrobio eta ostalari-mikrobioen arteko interakzioak aztertu behar dira tratamentu-mikrobiota-ostalari ardatza sakonki ezagutzeko, eta horixe da tesi honen **helburu** nagusia. Zehazki, tratamentuek eraginiko aldaketek bakterio komunitatean eta azken honek ostalarien errendimenduan duten eragina aztertzea.

Izan ere, hesteko bakterioak ostalariaren hainbat prozesu fisiologikoetan parte hartzen dute, hala nola, erantzun immunearen moldapenean eta elikagaien xurgapenaren erregulazioan. Bi prozesu hauek animaliaren garapen prozesuan zehar finkatzen dira, eta animaliaren osasuna eta errendimendua baldintzatzen dute haren ekoizpen zikloaren amaieran. Mikrobiotak ostalarian duen eragina zelula-zelula bidezko zuzeneko interakzioaren bidez eman daiteke, baina baita bakterioek sortzen dituzten metabolitoen bidez. Izan ere, ostalariaren hesteetan kokatutako zelula mota desberdinek hartzaile mota ugari aurkezten dituzte, bai zelularen mintzean zein zitoplasman. Bakterioen egiturak edo hauek ekoitzitako metabolitoak identifikatzean, seinale bideak aktibatzen dira, zeinak erreakzio lokalak edota sistemikoak sorrarazten dituzten, era horretan ostalariaren metabolismoan eragiten.

Animalien hesteetan gertatzen diren interakzio molekularrak ezagutzeko, aurrerakuntza teknologikoak funtsezkoak izan dira. Gaur egun, gai gara metodo ez-zuzenduen bidez izaki bizidunen geruza omiko ugari (genoma, transkriptoma, proteoma, metaboloma, etab) aldi berean aztertzeko. Honetaz gain, geruza multiple horiek sistema biologikoen bi domeinuetan, ostalari eta hari asoziaturiko mikrobiota, aldi berean azter daitezke sortu berri den holo-omika hurbilketa metodologikoari jarraiturik. Hala ere, metodologia berritzaile honen zailtasun nagusietako bat sorturiko datu guztiak analizatzerakoan esangarritasun biologikoa lortzea da.

Beraz, holo-omika aplikatzeko sistema aproposa aukeratzea ezinbestekoa da. Giza-kontsumorako erabiltzen diren abereen artean, oiloa garrantzitsuenetarikoa da mundu mailan. Gehien kontsumitzen den aberea izateaz gain, haren efizientzia energetiko altuarengatik aukerako proteina-iturri bilakatu da abeltzaintzak praktika jasangarriagoetara trantsizionatzeko saiakeretan. Ezaugarri genetikoei dagokionez, ekoizpen intentsiborako erabiltzen diren haragirako oilo leinu genetikoen populazioek hautespen artifizialaren ondorioz botila-lepo ugari jasan dituzte. Orokorrean dibertsitate baxuko populazioak bilakatu dira, leinu genetikoen artean ere desberdintasun txikiak egonik. Ekoizpen intentsiboan hazitako oiloen mikrobiota ere, murritza da ingurumen-baldintza kontrolatuetan hazten baitira txitak, helduengandik banaturik daudelarik haien bizi-zikloaren hasieratik. Haien heste mikrobiotaren konposizioa ingurumenetik eta dietatik jasotzen dituzten bakterioetara mugatuta dago.

Aipaturiko arrazoiengatik, tesi honetan aurkezten den hipotesia da holo-omika metodo eraginkorra dela elikagai-mikrobiota-ostalari ardatzak oiloen errendimenduan duen eragina hobeto ulertzeko. Horrela, ikerkuntza honetan sorturiko ezagutzak etorkizunean bakterioetan oinarritutako soluzio jasangarriagoak diseinatzeko erabilgarria izan daiteke.

Uste hauek testatzeko, 2019an zehar HoloFood H2020 proiektuaren barruan egin genituen 3 saiakuntza esperimental identikoetan jasotako oiloen hesteko laginak, zein oiloen errendimenduari eta osasunari lotutako hainbat parametro aztertu ditugu. Saiakuntza horietan, 2 leinu genetiko eta 2 sexuetako oiloak, 3 tratamendu dietetiko desberdinekin hazi genituen 35 egunez. Ekoizpen-aldian zehar, 3 laginketa gauzatu ziren 7., 21. eta 35. egunetan. Animalien itsuko mukosa eta digesta laginak jaso ziren oiloaren (meta)genoma eta (meta)transkriptoma aztertzeko. Oiloen errendimendua eta osasunaren jarraipena egiteko ere hainbat parametro gehigarri neurtu ziren. Esperimentu horietan jasotako emaitzak hiru ikerketa artikuluetan bildu dira.

Lehen artikuluak, *Perspektiba hologenomiko bat ostalari-mikrobiota elkarrekintzak ezagutuz oiloen errendimendua eta osasuna hobetzeko* deiturikoa, gauzaturiko esperimentuak deskribatzen eta oiloen errendimenduari eta osasunari loturiko parametroak aztertzen ditu. Artikuluaren helburu nagusia talde esperimentalen, hots oilo leinu genetikoen, sexuen, eta tratamenduen artean errendimenduan egon zitezkeen desberdintasunak aztertzea zen. Baina espero ez bezala, talde esperimentalen arteko desberdintasunak txikiagoak diren kaiola bereko eta esperimentu desberdinetakoetako indibiduoen arteko desberdintasunak baino. Izan ere, tratamenduek ez zuten eraginik izan oiloen errendimenduan. Horren arrazoia izan liteke diseinu esperimentalak, inplementatu zen moduan, ez zituela tratamenduen ondorioak azaleratu, eta, aitzitik, nabarmen areagotu zuela oiloen errendimenduaren aldakortasuna. Metodo estandarrez neurturiko osasun parametroek ere (plasmako haptoglobina eta lipopolisakaridoak, eta lumetako kortikosterona) ez zuten oiloen pisuan ikusiriko aldeak azaltzen.

Indibiduoen artean ikusiriko aldakortasun handia ziur aski ostalari-mikrobiota arteko interakzio molekularren ondoriozkoa izan zitekeen. Honekin bat, hirugarren esperimentuko emaitzek erakutsi zuten oiloak gutxiago hazi zirela 35 egun horietan, esperimentuen arteko aldakortasuna handituz. Azken esperimentuko oiloen batez besteko pisu jaitsiera *Campylobacter* spp. kolonizazio oportunistari loturik egon zitekeen. Hain zuzen, 21. egunetik aurrera laginduriko indibiduo guztiek *Campylobacter* positiboa eman baitzuten.

Beraz, errendimenduan ikusitako bi joerak ostalariaren eta hari lotutako hesteko mikrobio-komunitateen arteko elkarrekintzaren ondorioz azal daitezkeela proposatzen dugu. Horrenbestez, tesiaren hurrengo atalek holo-omika erabiltzen dute errendimenduan ikusitako

aldakortasunen atzean dauden ostalari-mikrobiota interakzioak ulertzeko. Hau baieztatzeko, itsuko laginen bidez bertako bakterioen genoma katalogoa eraiki zen, 822 bakteriodun katalogoa osatu zelarik. Katalogo honekin, sekuentziaturiko 613 indibiduoetako bakoitzak zuen mikrobioma konposizioa eta bakterio bakoitzaren, zein komunitatearen, ezaugarri funtzionalak aztertu genituen. Zehazki ostalarian eragina izan dezaketen funtzioetan interesaturik geundenez, bakterioen genomatik anotaturiko milioi bat gene baino gehiago, 170 ezaugarri funtzionaletan filtratu eta multzokatu genituen. Ezaugarri funtzional hauekin, bakterio bakoitzari kapazitate funtzional bat esleitu genion, eta bakterio bakoitzak ezaugarri funtzional horiek burutzeko espresio genikoa ere lortu genuen.

Bigarren artikuluan, *Hesteko mikrobiotaren ahalmen metaboliko txikia haragirako oiloaren hazkundearekin lotuta dago* deiturikoa, lehen eta bigarren esperimentuetako oiloen hesteko mikrobioma aztertu genuen denboran zehar. Ohartu ginen dibertsitate funtzionala gora zihoan ahala kapazitate funtzionala beherantza zihoala. Joera hau denboran zehar mikrobiomaren konposizioa aldatzen zelako ematen zen, nahiz eta bakterio gehiago agertu oiloak hazi ahala, hauen kapazitate funtzionalak murritzagoak ziren. Gainera, kapazitate baxuko bakterio hauen abundantzia positiboki erlazionatuta zegoen oiloen pisuarekin esperimentuen amaierarako. Ahalmen txikiko bakterio taxoi hauek konkretuki, hala nola TANB77, RF39, RF32 eta UBA1242, gutxi ezagutzen dira ez baitira laborategietan isolatu. Hare gehiago, beste ornodun askotan aurkitu izan dira eta baita bestelako osasun aspektuekin erlazionatu dira. Taxoi hauek genoma murritzak izateak eta haien abundantziak eragina izateak ostalarian gainera sinbiosi erlazio estu bat iradokitzen du ostalariarekin.

Bakterioen datu multi-omikoetan oinarrituta, erakutsi dugu oraindik ez ditugula guztiz ulertzen oiloaren hesteko mikrobiomaren dinamika tenporalak, ezta haren konposizioak eta profil funtzionalak izan ditzaketen eragina ostalarian. Mikrobiomaren ezaugarri funtzionalak aztertzeak bide berriak irekitzen ditu bi domeinuen arteko elkarreragin mekanistikoak ulertzeko.

Hirugarren artikuluak, *Lehentasunezko efektuak eta bakterioen arteko elikadura gurutzatuak forma ematen diote agente zoonotikoen kolonizazioari oiloetan* deiturikoa, *Campylobacter* positibo eman zuten oiloak *Campylobacter* negatibo ziren oiloekin konparatzen ditu. Datu multi-omikoek hirugarren esperimentuko oiloetan *Campylobacter*-en presentzia konfirmatu zuten, baina Campylobacterales taxoiko 3 bakterio desberdin identifikatu genituen, hots *C. jejuni*, *C. coli* eta *Helicobacter pullorum,* bakoitzak dinamika tenporal desberdina erakusten zuelarik 21. egunetik aurrera. Gainera, hirugarren esperimentuko oiloen mikrobiomak 7.egunetik bereizten hasia ziren beste bi esperimentuetako oiloen mikrobiomengandik. Desberdintasun azpimarragarriena *Bacteroides fragilis_A*-ren hedapen goiztiarra zen. Bakterioen genometan oinarrituz sare metabolikoak inferitu genituen bakterioen arteko

erlazioak aztertzeko. Horrela *B. fragilis_A*-ren hedapen goiztiarrak *Campylobacter*-en kolonizazioa errazten zuela ikusi genuen. Are gehiago, denboran zehar nagusitu ziren aldeak mikrobiomen konposizioan, mikrobiomen aktibitatean ere ematen zirela baieztatu genuen. Hazkuntza zikloaren amaierarako enterotipo desberdina aminoazidoen degradazioan espezializatu zen, ostalariaren elikagaien xurgapena zailtzen duelarik. Horrek hirugarren esperimentuko oiloen pisu galera azaldu ahalko luke. Izan ere, ez genuen ostalariaren hesteko transkriptoman inolako erantzun immunerik nabaritu.

Aurkikuntza hauek bidea errazten dute markatzaile bakterioano goiztiarrak identifikatzeko *B. fragilis_A* bezala, eta horrela mikrobiomaren manipulazioa agente zoonotikoen agerpena baino lehenago emateko, *Campylobacter*-en kolonizazioa saihesteko helburuarekin. Hala ere, beste hainbat bakteriek erraztu ahalko lukete *Campylobacter*-en agerpena eta hedapena. Lan honek erakusten du nola kolonizatzailea edo patogenoa eta ostalaria ikertzeaz gain, zein garrantzitsua den hesteko mikrobiota osoa aztertzea.

Laburbilduz, holo-omika tresna ahaltsua bilakatu da, bestelako metodoen bidez ikusezinak izango liratekeen interakzio biologikoak ezagutarazteko. Hala ere, ikuspegi horren eraginkortasunari aurre egiten diote datu-sorkuntzaren kostu handiek eta informatikan eta estatistikan hainbat geruza omiko integratzeko gaitasun aurreratuek. Beraz, estrategikoa izan daiteke metadatuen aurretiazko explorazioa eta talde mugatu bati buruzko atariko analisiak egitea, holo-omiken bideragarritasuna ziurtatzeko. Erronka horiek gorabehera, gure ikerketek nabarmentzen dute proba eta erroreko esperimentuetatik jakintzan oinarrituriko saiakeretara trantsizionatzeko beharra dagoela.

Disertazio honek oiloen mikrobiomak ostalariaren ikuspuntutik aztertzen ditu, eta arreta jartzen du haren denboran zeharreko joeretan. Analisietan zehar, ostalariaren energia-orekan garrantzia izan duten hainbat taxoi bakterioano nabarmendu dira. Eskura dauden baliabideak aprobetxatuz, etorkizuneko ikerketek sortutako datuak erabili ditzakete taxoi hauetan gehiago sakontzeko.

Introduction

Chapter 1

Sustainable solutions for the food production crisis

The food production crisis

Amidst the ongoing climate change, humanity faces numerous challenges to maintain a long-term balance between consuming natural resources and preserving the environment (Reid et al., 2010). Intensive livestock farming is particularly harmful for the environment, contributing to 14.5% of global anthropogenic greenhouse gas emissions (Gerber et al., 2013). Furthermore, global meat consumption is expected to increase due to the population and economic growth (Godfray et al., 2018). High-income countries have witnessed stable or even declining trends mainly in red meat intake due to transitioning to more sustainable diets (Kozicka et al., 2023) and an increased awareness of higher risks for certain chronic diseases (Willett et al., 2019); but animal-based foods are becoming a relevant source of nutrients in low- and middle-income countries (Godfray et al., 2018). Under these circumstances, competition with other land and water uses is intensifying, resulting in less available resources for animal production (Campbell et al., 2017). Therefore, the scientific community and policy makers agree that it is urgent to exploit available technologies and improve management practices, for transitioning towards a more sustainable, efficient, and safe animal production model (Bowles et al., 2019; Springmann et al., 2018).

Significant progress in production efficiency has been mainly due to selective breeding programmes and advances in genetics, followed to a lesser extent by improved husbandry, sanitary, and diet strategies (Thornton, 2010). Despite significant advances already achieved, a considerable degree of variability is found in production parameters such as feed efficiency (Cantalapiedra-Hijar et al., 2018; He et al., 2023; Quan et al., 2018), body weight gain (Tarsani et al., 2019; H. Zhou et al., 2022) and nonuniformity of performance (Lundberg et al., 2021; Vasdal et al., 2019) between farms (Ricciardi et al., 2021), flocks (Vasdal et al., 2019) and even individuals (Shah et al., 2019), creating a gap between the theoretical potential of animals and the actual performance achieved under practical conditions. Maintaining optimal functioning of the digestive tract, which plays a crucial role in digesting and absorbing nutrients, is essential for achieving a balanced energy state and preventing the proliferation of harmful microorganisms in the intestinal tract (Kogut & Arsenault, 2016). Accordingly, numerous dietary interventions were developed to tackle animal poor performance in recent decades, antibiotics being a widespread option (Celi et al., 2017).

Antibiotics were originally administered to animals as treatment for specific pathogenic bacteria, but later was discovered that they could also be used for sub-therapeutic purposes such as improving growth, increasing efficiency, and preventing disease (Gorbach, 2001). The positive impact of antibiotics on animal growth is attributed to two main causes: the reduction of competition for nutrients with bacteria due to the reduced microbial biomass (Dibner & Richards, 2005; Feighner & Dashkevicz, 1987; Gaskins et al., 2002), as well as a decrease of energy expenditure due to a lowered inflammatory response (Niewold, 2007). The use of antibiotics as growth promoters has become an integral practice in intensive animal production systems since its discovery in the 1940s (Stokstad & Jukes, 1950). To get an estimate, animals accounted for 63% of antibiotic consumption worldwide by 2010 (Van Boeckel et al., 2015). This overuse contributes to the emergence of antibiotic resistance, and

is a significant concern for public health and the environment (Laxminarayan et al., 2013; Manyi-Loh et al., 2018). To address this situation, the European Union banned antibiotics as growth promoters in 2006 (Parliament and Council, 2003). A measure that only a few countries outside Europe have adopted since then (New Zealand and Republic of Korea), whereas the rest have merely implemented more restrictive regulations (Okocha et al., 2018; Robles-Jimenez et al., 2021). With the growing concern about antibiotics, a number of alternatives to improve production efficiency have been proposed (Gadde et al., 2017; Verstegen & Williams, 2002).

Alternatives to antibiotics

According to the European Commission, feed additives are non-nutritive natural products used in animal nutrition for improving the quality of feed and the quality of food from animal origin or to improve the animals' performance and health (Parliament and Council, 2003). In addition to antibiotics in subtherapeutic doses, these include phytochemicals, probiotics and prebiotics, among others (Gadde et al., 2017; Y. Liu et al., 2018). In terms of their functionality, the four additives mentioned above are grouped among animal growth and production enhancers (Gadde et al., 2017; Pandey et al., 2019). Phytogenic chemicals are a wide range of plant-derived products such as herbs, essential/aromatic oils, and oleoresins (Gadde et al., 2017). Prebiotics are non-digestible feed components for the host that support the growth of beneficial gut microorganisms (Gibson et al., 2017), while probiotics might refer to a single microorganism, or a mixture of several live microorganisms, that is directly inoculated to the gastrointestinal tract (Fao/who, 2002). In general, phytobiotics are attributed antimicrobial properties like antibiotics, whereas prebiotics and probiotics are known to enhance microbial communities (Kim & Lillehoj, 2019; Kogut, 2019). These gut microbiota modulators have become immensely popular in the last two decades, due to the increasing understanding of the pivotal role of microorganisms in host biological functions (Allen et al., 2013; Gaggìa et al., 2010). They require, however, a marketing authorisation, which is obtained through a scientific evaluation that demonstrates they are beneficial to the target animal without being harmful to animals and humans (EFSA Panel on Additives and Products or Substances used in Animal Feed (FEEDAP) et al., 2018).

Probiotics and prebiotics are tested with *in vitro* and *in vivo* experiments, probiotics having the strictest controls. *In vitro* assays for probiotics may comprise strain cell-cultures with different physico-chemical conditions and cell surface cultures of eukaryotic epithelial or other cell line monolayers (de Melo Pereira et al., 2018). Strains can also be genetically characterised in this first step (de Melo Pereira et al., 2018; Vinderola et al., 2017). *In vitro* essays for prebiotics are divided in two evaluations, namely digestibility in the upper gastrointestinal tract and microbial activity in the distal gastrointestinal tract. For this second test, the growth of popularly known beneficial bacteria such as *Lactobacillus spp.* and *Bifidobacterium spp.* is evaluated (Bajury et al., 2018). If the *in vitro* experiments pass all the quality controls needed, *in vivo* trials with target hosts are performed afterwards. These large-scale trials consist of a comparison of one or more treatments with a control group, in which animals are sampled for performance and health measures (Markowiak & Śliżewska, 2018).

Despite the many diverse experimental procedures to verify their action, producers, end-users and the scientific community have consistently reported significant variations in the efficacy of these additives (Barba-Vidal et al., 2019; Gaggìa et al., 2010; Soccol et al., 2010; Zommiti et al., 2020). In addition, the correlation between *in vitro* and *in vivo* experiments remains unclear (Vinderola et al., 2017), as the response of additives *in vivo* can differ significantly from that observed in *in vitro* studies due to varying physicochemical conditions and possible interactions with numerous other microbial strains and the host organism (Suez et al., 2019). The fundamental reason behind this inconsistency is that these experiments do not consider microbe-microbe and host-microbe interactions within the complex gut environment (Suez et al., 2019). These systems involve various biological components beyond the animal being produced, with host-associated microorganisms being particularly important for ensuring optimal biological function in most animals (McFall-Ngai et al., 2013). Hence, gaining an in-depth understanding of the biological mechanisms underlying animal production systems is a crucial step towards finding microbe-based solutions in order to optimise farming practices (Messerli et al., 2019).

Unravelling host-microbiota interactions

Gut microbiota

The study and manipulation of host associated gut microbiota has become global, particularly in the context of agricultural production (Małyska et al., 2019; Sessitsch & Mitter, 2015). The gut microbial community can include viruses, prokaryotes (archaea and bacteria) and eukaryotes (fungi, protozoa and helminths) (Alberdi et al., 2021), but the terms “microbiota” and “microorganisms” will refer only to bacteria hereafter. The gut microbiota plays a crucial role in many physiological host functions, such as regulating nutrient absorption (Diaz Carrasco et al., 2019), shaping immune and inflammatory processes (H. Zhou et al., 2020), and affecting the host's systemic growth parameters (Fraune & Bosch, 2010). Microorganisms colonise the upper mucus layer of the gastrointestinal tract as soon as an animal becomes exposed to the environment (Sprockett et al., 2018), and form communities with complex temporal and spatial dynamics that continuously interact with the host animal (Ansari et al., 2020; Khan et al., 2019). This interaction is essential for the development and maturation of the immune cells, ensuring the tolerance of commensal microorganisms, the recognition of pathogens, and the control of potential pathogenic commensal microorganisms (Broom & Kogut, 2018; Kogut et al., 2020). Gut microbiota contributes to the development of the intestinal mucosa, which is the first physical barrier composed of a mucus layer and epithelial monolayer, by producing metabolites that increase the epithelial surface (Oakley et al., 2014). In addition, metabolite recognition can regulate hormonal signals release. Gut hormones have a wide range of targets in the whole body and play systemic physiological roles especially in the control of metabolism, but also can act as locally in epithelial cells by generating hormonal signals that reflect dietary intake, microbial composition and epithelial integrity (Gribble & Reimann, 2019). These effects result from direct cell contact and the influence of metabolites, either directly produced by bacteria or generated by the host and then metabolised by bacteria (Agus et al., 2018).

Sensing of bacteria

Recognition of bacteria activates a host response that can occur at two different scales. The innate immune response, considered a local-scale reaction, is initiated through recognition of molecular structures shared by multiple microbes by pattern recognition receptors (PRRs), which are broadly expressed in different cell types of the gut, including immune cells, stromal cells, and neuronal cells (Wan et al., 2023). Bacteria are recognised by identifying certain structural elements such as lipopolysaccharides, flagellin or peptidoglycans (Zhao & Elson, 2018). These PRRs are classified based on their portrait domain homology, and include Toll-like receptors (TLRs), NOD-like receptors (NLRs), retinoic acid-inducible gene I-like receptors (RLRs) and C-type lectin receptors (CLRs), among others (Wan et al., 2023). TLRs and CLRs are expressed in the cellular membrane but can also be located in endosomes and lysosomes, whereas NLRs and RLRs are cytoplasmic sensors (Wan et al., 2023). The intracellular signalling causes the expression of various immune response genes that can either realise interferons and cytokines or suppress an inflammatory response (Kogut et al., 2020). The adaptive response is a systemic reaction that initiates with the recognition of antigens by T and B lymphocytes to then produce antibodies resulting in microbial killing and provide a long-term specific protection against subsequent infections with a pathogen bearing the same antigens (Kogut et al., 2020). T-cells, B-cells, and immunoglobulin repertoires and responses can vary greatly depending on the gut microbes' colonisation niche and metabolic properties (Zhao & Elson, 2018).

Sensing of metabolites

The host cells can also sense and respond to metabolites derived either from the anaerobic fermentation of undigested dietary components or from compounds that are *de novo* synthesised by the microbes (Krautkramer et al., 2021). Thus, microbial products such as fatty acids (short- (SCFAs), medium- (MCFAs) or long-chained (LCFAs), lipopolysaccharide (LPS) and secondary bile acids, serve as interaction molecules (Wan et al., 2023). These microbial metabolites are recognised by G protein-coupled receptors (GPCRs), aryl hydrocarbon receptors (AHRs) and nuclear receptors (Wan et al., 2023). While CPCRs are membrane receptors, AHRs and nuclear receptors are transcription factors located in cell nucleus that induce downstream signalling pathways (Wan et al., 2023). Naturally, all these processes are also influenced by microbe-microbe interactions, which mutually influence functional activities through both agonistic and antagonistic interactions (Culp & Goodman, 2023). Antagonistic interactions result from competition for resources or space, and can produce specific metabolic functions to be majorly conducted by different bacteria at different time points. In contrast, agonistic interactions imply bacteria use molecules produced by other bacteria (Venturelli et al., 2018), thus scaffolding into more complex metabolic networks.

This body of evidence underlines the necessity of considering the host-associated gut microbiota in combination with the host when studying animal's health and disease in food production contexts. Technological advances have led to the rapid expansion of this research area that is dedicated to answer biological questions taking host-microbiota interactions into

account (Alberdi et al., 2021). This integrated approach has been coined hologenomics (Alberdi et al., 2021), reflecting the coupled study of the genomes of both the host and its associated microorganisms, known as the hologenome (Bordenstein & Theis, 2015; Zilber-Rosenberg & Rosenberg, 2008).

Feasibility of a holo-omics approach

Hologenomics

The myriad evidence of host-microbiota symbioses has led the scientific community to devote attention to these relationships in multiple areas of basic and applied research (Alberdi et al., 2021; Nyholm et al., 2020). However, debates have arisen due to the specific terminology for this area (Suárez, 2018). The term “holobiont” or “metaorganism” are commonly used to describe the symbiotic assemblies of a multicellular organism and its associated microorganisms (Gilbert et al., 2012; Jaspers et al., 2019). Lynn Margulis introduced the term “holobiont” limiting it to cases of heredity symbiosis (Suárez, 2018), but the “hologenome theory of evolution” broadened the concept to encompass the host with all its microbes as a unit of evolutionary selection (Zilber-Rosenberg & Rosenberg, 2008). Likewise, the sum of the genetic information of the holobionts was coined “hologenome” (Bordenstein & Theis, 2015; Zilber-Rosenberg & Rosenberg, 2013). The theory is defended by multiple authors (Doolittle & Booth, 2017; Gilbert et al., 2012), but also has many detractors (Douglas & Werren, 2016; Moran & Sloan, 2015) likely due to diverging conceptions about biological individuality (Suárez, 2018). But the adoption of the terms “hologenome” and “hologenomics” in this dissertation is instrumental, and reflects an approach to study the intricate host-microbiota relationships (Alberdi et al., 2021).

Recent advances in high-throughput DNA sequencing and mass spectrometry technologies, coupled with more powerful bioinformatic tools and increased computing capacity, are revolutionising this field (Graw et al., 2021). Today, the hologenome can be studied on various levels of resolution, ranging from marker genes to obtain information on host population/microbial taxonomy to the use of whole-genome sequencing for both domains (Alberdi et al., 2021). However, host-microbiota associations are dynamic, bidirectional relations that are influenced by various external and internal physico-chemical and biological factors (Y. Liu et al., 2019; Rasmussen et al., 2022; X. Wang et al., 2019). To gain insights into momentary interactions, it is essential to rely on multiple other omic layers beyond genomics, such as (meta)transcriptomics, (meta)proteomics or (meta)metabolomics (Limborg et al., 2018; Nyholm et al., 2020). Nevertheless, the addition of various omic layers from both domains involves higher resources to collect, process and sequence samples, as well as generate, store, analyse data (Nyholm et al., 2020). Thus, holo-omic studies focus on a reduced part of the entire network of interactions (Rasmussen et al., 2022; Shah et al., 2019; Wu et al., 2022). This dissertation, in particular, details the characteristics of non-targeted sequencing of the two nucleic acids from both domains.

Data generation

Sample processing and data quality check

Currently, shotgun sequencing is the most popular method for sequencing nucleic acids, with the laboratory and initial data quality check processes being similar but featuring some particularities. First, DNA extracted from a sample is fragmented, and then sequencing libraries are built, indexed using platform-specific sequencing adaptors, and pooled for sequencing (Ekblom & Wolf, 2014; Quince et al., 2017). Sequencing of short-reads is usually performed based on second-generation sequencing instruments (Pervez et al., 2022). RNA is more unstable than DNA, demanding meticulous handling and preservation methods until it is converted to complementary DNA (cDNA) (Westermann & Vogel, 2021). In addition, the highly abundant ribosomal RNA (rRNA) needs to be removed, as it often constitutes more than 80% of the RNA content of bacterial and eukaryotic cells (Westermann & Vogel, 2021). For host transcriptomics, it is usually removed by relying on poly-A tails (Westermann & Vogel, 2021), whereas for metatranscriptomics direct depletion strategies are needed because prokaryotes lack poly-A tails (Tan et al., 2023). Once nucleic acids are sequenced, raw reads undergo a quality filtering process that entails removing duplicate reads and trimming adaptors (Quince et al., 2017). Depending on the aims of the study as well as host/microbial DNA proportions, different approaches exist. These can be broadly split between read-based and assembly-based analyses (Y.-X. Liu et al., 2021; Quince et al., 2017). Given the abundance of already assembled reference genomes (Rhie et al., 2021), and the high diversity of bacteria yet to be explored (Avila Santos et al., 2023), the host genome is usually resequenced while bacterial genomes are assembled (Eisenhofer et al., 2023). In case of RNA, while assembly-based methods are feasible (Langa et al., 2021), transcripts from both domains are usually aligned to their respective annotated genomes (Ojala et al., 2023).

Bacterial metagenome

Host-associated metagenomics involves the non-targeted sequencing of all DNA within an environmental sample, including bacterial genomes (Quince et al., 2017). Through genome-resolved metagenomics, bacterial genes and genomes can be reconstructed and quantified, enabling better taxonomic assignment, phylogenetic reconstruction and direct functional annotations than 16S rRNA amplicon sequencing (Durazzi et al., 2021).

Reference-based methods align clean reads to curated databases to obtain taxonomic and functional information using MetaPhlAn2 or Kraken 2 (Y.-X. Liu et al., 2021; Quince et al., 2017). In fact, curated gut metagenome catalogues for farm animals such as chicken (Gilroy et al., 2021), pig (Xiao et al., 2016) and cow (Stewart et al., 2019) are already available. This method can be preferred when the relative amounts of microbial DNA are low. Its main limitation is that previously uncharacterised microbes are difficult to profile (Y.-X. Liu et al., 2021) and that potential biases depend on the characteristics of the reference database (Avila Santos et al., 2023).

Assembly-based methods assemble reads into larger sequences called contigs with tools such as MEGAHIT and metaSPAdes (Quince et al., 2017). These contigs can be directly

used to build a non-redundant gene catalogue (Huang et al., 2018), to map against an existing reference database or to assemble a *de novo* bacterial genomes (Quince et al., 2017). To study novel bacterial strains, contigs are grouped into bins to recover partial or complete bacterial genomes through the process known as contig binning (Quince et al., 2017). Popular binning tools include CONCOCT, MaxBin2 and MetaBAT2, which cluster contigs based on tetra-nucleotide (also called k-mer) frequency and contig coverage. Then, low-quality bins can be removed and the catalogue can be de-replicated to avoid redundancy (Quince et al., 2017). But genome-resolved metagenomics also has its own limitations, as it often misses a considerable portion of the existing diversity and usually does not recover circular genomes (Y.-X. Liu et al., 2021). Long-reads obtained with platforms such PacBio and Oxford Nanopore or hybrid strategies can be considered depending on a study-specific basis (Eisenhofer et al., 2023).

Host genome

Host genome contains genetic characteristics unlikely to change over the typical timescales of holo-omic studies, with variations across individuals and populations (Alberdi et al., 2021). It enables accurate estimates of population genetic parameters and potentially identifies gene-level variation between populations (Bourgeois & Warren, 2021). The main advantage of using whole-genome resequencing instead of targeted approaches such as single-nucleotide (SNP) chips for genotyping is that it captures a wide range of variation specific to the population of interest, such as rare variants, structural variations and copy number variations (CNVs) (Jones & Wilson, 2022).

If a reference genome is available, clean reads are directly aligned (Ekblom & Wolf, 2014). While primarily used for characterising the genomic architecture of microbial communities, metagenomic data generated from host-associated samples can also be used for extracting genomic information of the animal host (Blekhman et al., 2015). Indeed, modern host DNA derived from metagenomics has many applications as a non-invasive approach, such as the study of endangered species (Ang et al., 2020) or human diseases (Jiang et al., 2020). However, in animal studies the sequencing depth of the host is highly unpredictable, as the proportion of host DNA can vary substantially between samples (Blekhman et al., 2015). If enough DNA is obtained, variants can even be directly genotyped (Blekhman et al., 2015). If a very low amount of DNA is obtained, genotype likelihoods can be calculated or missing data can be imputed with tools like ANGSD, Beagle or IMPUTE2, taking advantage of recent advances in ancient DNA studies (Orlando et al., 2021; Parejo et al., 2020).

Although reference-based approaches are the most common option for studying the host genome, assembly-based methods are becoming more affordable (Jung et al., 2020). An example of it is the international effort to generate complete genomes for many species (Rhie et al., 2021). However, generating high-quality genomes requires an approach that combines multiple sequencing strategies to resolve the complex structures of eukaryotic genomes (Amarasinghe et al., 2020). The actual optimal standard procedure combines long-read sequencing, such as Pac-Bio HiFi data, for genome assembly with scaffolding methods like Illumina Hi-C data to obtain nearly complete genomes (Amarasinghe et al., 2020).

Bacterial metatranscriptome and host transcriptome

Metatranscriptomics provides information about the microbial community's transcriptional regulation of active genes at the time the sample was taken, while intestinal transcriptomics provides information on local host response (Nyholm et al., 2020). Host transcriptome covers all types of transcripts, including messenger RNAs (mRNAs), microRNAs(miRNAs) and long noncoding RNAs (lncRNAs) (Westermann & Vogel, 2021). Microbial and host RNA sample collection usually is spatially divided, as the bacterial diversity in the mucosal part of the intestine is lower compared to that in the gut contents (Nyholm et al., 2022), while the host response is more appreciable in the intestinal mucosa or epithelium itself (Lichtman et al., 2015).

Mapping to a reference genome of (meta)transcriptomics is specific for each domain, since eukaryotic RNA molecules contain introns, which prokaryotic cells do not, requiring splice-aware reads mappers (Westermann & Vogel, 2021). For metatranscriptomics, reference-based methods can directly map reads to databases. This works well when community member species are largely known and databases are well characterised. But if bacterial species are unknown, mapping to a *de novo* metagenome-resolved catalogue might be preferred (Ojala et al., 2023).

Data integration

Omics-based studies are facilitating a new, more holistic understanding of systems biology, but the analysis of each layer generates big data files, making the integration and biological interpretation of these multidimensional omics data challenging (Q. Wang et al., 2019). Multi-omic integration tools, which can be categorised as multi-staged and meta-dimensional, employ traditional statistical methods for supervised and unsupervised analyses (Graw et al., 2021). Among unsupervised methods, analyses can be divided into clustering or Principal component methods that aim to identify groups, and association or network methods that are used to identify relationships between omic and sample characteristics (Santiago-Rodriguez & Hollister, 2021). In contrast, supervised analyses attempt to model features that can be used to predict traits. These methods include regression and multivariate analyses (Santiago-Rodriguez & Hollister, 2021). The multi-staged approach analyses each layer in various steps, while in meta-dimensional analyses multiple omic layers are pooled in a single analysis (Graw et al., 2021). Meta-dimensional approaches are becoming more popular with the advances of artificial intelligence. Alongside user-friendly web-based tools, more complex R packages such as mixOmics, mCIA, MOFA and MoCluster, support microbiome or single organism datasets (Santiago-Rodriguez & Hollister, 2021). But multi-layered approaches are computationally challenging, as well as difficult to display and to comprehend visually (Graw et al., 2021). In the case of host-microbiota studies, multi-staged approaches, enable to acknowledge the distinct nature of the host (a single eukaryotic organism) and microbiota (a community of prokaryotes) omics (Kwoji et al., 2023; Q. Wang et al., 2019). Species communities are assembled following the basic processes of selection, drift, dispersal and speciation, which determines the community composition, functional profile and biodiversity of the microbiome (Costello et al., 2012; Coyte et al., 2015). For instance, making inferences

from metagenomic data may benefit from specific statistical frameworks such as joint species distribution modelling (Tikhonov et al., 2020).

The application of such complex methodological approaches require not only interdisciplinary knowledge, but also specialised infrastructure and methodologies. The European Union concentrated a large part of its research and innovation activities in the Horizon 2020 (H2020) Framework Programme in the period of 2014 and 2020. This thesis is developed within HoloFood, a H2020 project granted in 2019 to study host-microbiota interactions with a holo-omic approach for a sustainable animal production (HoloFood Consortium, 2019).

Research framework

HoloFood

Holofood ran from 1 January 2019 to 30 June 2023, and brought together 10 European industrial, research and academic partners, the University of the Basque Country being one of them (HoloFood Consortium, 2019). HoloFood's aim was to implement a holo-omic framework to improve feed additives and diet formulations, as well as to optimise meat production efficiency. To achieve this, different stages of the production line were thoroughly examined in two critically important farmed animal systems (salmon and chicken) that were raised under different dietary treatments. The purpose behind selecting these two animal models was to highlight the most significant contrast between the two systems, given that salmon inhabits aquatic environments while chicken is a terrestrial animal. The chosen holo-omic framework comprises the characterisation of the associated gut microorganisms' genomes, transcriptomes and metabolomes; and the genomes, intestinal transcriptomes and intestinal immune processes of the mentioned two animal systems. All this in relation to animal performance and welfare parameters.

The chicken scientific work was coordinated between The University of the Basque Country (EHU) and the Center for Evolutionary Hologenomics (CEH), The GLOBE Institute, Denmark. In short, chicken experiments were conducted in the Institute of Agrifood Research and Technology (IRTA), Spain. IRTA was responsible for coordinating the experiments, as well as collecting and analysing animal performance and welfare parameters. Bacterial metabolites, host cytokines, and mucosa integrity from chicken intestinal samples were analysed by the Institute of Animal Nutrition, Freie Universität Berlin (FUB), Germany. Omic data for both domains, chicken and its microbiota, was generated by CEH. Metagenome-resolved genomes were assembled by the European Bioinformatics Institute (EMBL-EBI), United Kingdom. Data was analysed by CEH and EHU.

This dissertation explores host-microbiota interactions, using (meta)genomics and (meta)transcriptomics analyses obtained from chicken caecum, which are complemented by animal performance and welfare parameters.

Broiler chicken

Poultry, and especially chicken (*Gallus gallus domesticus*), is the main source of animal meat consumed by humans worldwide and it is forecasted to undergo the largest production increase in the near future (Alexandratos et al., 2012). In 2020, 130 million tons of chicken meat were produced (*OECD-FAO Agricultural Outlook 2017-2026 - En - OECD*, n.d.). But its relevance in the meat industry has been rather recent. Domestication of chicken from its wild ancestor, the red junglefowl, started about 8,000 years ago in South-Eastern Asia (M.-S. Wang et al., 2020). Its adaptability to a range of environments and manageability facilitated the spread of the chicken from the jungles of Southeast Asia to all parts of the world (Lawal & Hanotte, 2021). But scientists suspect that they were not used as food resources when introduced to Europe, rather they were rendered as exotic animals for spiritual practices (Doherty et al., 2021; Sykes, 2012). Over time, chickens became highly valued for their eggs, whereas their meat was considered a by-product (Siegel, 2014). It was only in the late 1940s that genetic selection programs for modern meat chicken breeds, also called broilers, were developed (Havenstein et al., 1994). Genetic selection has proven highly effective, leading to a complete transformation of their physiology from that of their shared ancestor. Productivity gains have been achieved through intensive selection on production traits over generations of purebred populations, followed by cross-breeding strategies (Fad, 2013; van der Most et al., 2011). Their short generation interval and the high heritability of production parameters facilitated obtaining market weights in younger ages (Havenstein et al., 2003a, 2003b). Today, their feed conversion ratio (FCR) is around 1.6-2.0 (Havenstein et al., 2003b), which means that a 2.44 kg broiler chicken can be produced in approximately 35 days using 3.66 kg of feed (Siegel, 2014), a period of time in which birds do not reach sexual maturity (Rychlik, 2020).

The domestication bottleneck (M.-S. Wang et al., 2021) and subsequent genetic selection programs for production (Siegel, 2014) has left a number of genetic adaptations in modern broiler chickens. At present, primary broiler genetic lines used in global intensive production are controlled by three major companies: Cobb-Vantress, Aviagen, and Hubbard (Fad, 2013). The most popular genetic lines from these companies could be the broilers Ross 308 (Aviagen) and Cobb 550 (Cobb-Vantress). These genetic lines are close to each other, probably due to their common ancestor, the Cornish breed (Qanbari et al., 2019; Rubin et al., 2010). They show a smaller proportion of rare alleles compared with wild chickens, presumably caused by genetic selection that reduces frequencies of slightly deleterious mutations (Qanbari et al., 2019). They also show higher inbreeding signs than wild chickens, but not as much as chickens that are used for eggs (Talebi et al., 2020). Hence, all these characteristics highlight the little genomic complexity of broilers, which renders them a good system for studying host-associated microbiota variations (Alberdi et al., 2021)

Chicken gut microbiota

Intensification of chicken meat production has introduced several changes to the natural behaviour of chickens that affect gut microbial dynamics. The newly hatched chicks are exposed to non-avian, environmental sources of bacteria, instead of acquiring resident gut bacteria through the direct contact with the mother hen (Rychlik, 2020). Therefore, the intestinal colonisation patterns are highly variable and unstable, with delayed gut microbiota development likely following multiple microbial trajectories that are governed by environmental and host-associated factors (Proszkowiec-Weglarz et al., 2022).

Microbial communities are specific to each intestinal section, with higher microbial mass and diversity of anaerobic bacteria in the large intestine (caecum and colorectum) compared to the small intestine (duodenum, jejunum and ileum) (Huang et al., 2018). This microbial compartmentalization is due the specific characteristics of each gastrointestinal section. The small intestine is the main section for chemical digestion and absorption, while the caecum is the main microbial fermentation chamber (Svihus, 2014). Despite the jejunum being the preferent site for nutrient absorption in the small intestine, it is suggested that it continues to happen to a certain extent in the ileum and even in the caecum (Svihus, 2014; Svihus & Choct, 2013). But whereas the absorption is continuous during digesta transit in the small intestine, it resembles batch cultivation in the caecum (Rychlik, 2020). Unabsorbed feed in the small intestine is then passed to the colon, from where digesta moves by retrograde antiperistalsis to the caeca, two blind-ending distal segments (Duke, 1989; Svihus & Choct, 2013). Considering that chickens have a shorter gastrointestinal tract and faster digesta transit compared to other food animals, which is about 3.5h, caeca is the most important section for an extended digestion since the emptying of the caecum takes place every 24 to 48 hours (Bindari & Gerber, 2022; Pan & Yu, 2014).

Chicken caecal microbiota is the most studied microbiome in chickens (Bindari & Gerber, 2022; Borda-Molina et al., 2018; Oakley et al., 2014; Rychlik, 2020). Caecal microbial communities diversify according to chicken age, and once the community is established, it is primarily dominated by Firmicutes and Bacteroidota phylums (Jurburg et al., 2019; Q. Zhou et al., 2021). There are expected differences between chicken genetic lines (Richards et al., 2019; Schokker et al., 2015), as well as a significant interindividual variation that can affect chicken growth performance (Shah et al., 2019). Bacterial gene catalogues have improved our ability to comprehend how dietary treatments influence the functional traits of the microbiome (Huang et al., 2018), while bacterial genome catalogues provide the landscape of unique functional attributes of individual microbial strains (Feng et al., 2021; Gilroy et al., 2021; Glendinning et al., 2020; Segura-Wang et al., 2021; Zhang et al., 2022). Consequently, the complex caecal microbiome shows a temporal succession which can be monitored during chicken growth period, and even most likely modulated by dietary treatments (Alberdi et al., 2021).

References

Agus, A., Planchais, J., & Sokol, H. (2018). Gut Microbiota Regulation of Tryptophan Metabolism in Health and Disease. *Cell Host & Microbe*, *23*(6), 716–724. https://doi.org/10.1016/j.chom.2018.05.003

Alberdi, A., Andersen, S. B., Limborg, M. T., Dunn, R. R., & Gilbert, M. T. P. (2021). Disentangling host–microbiota complexity through hologenomics. *Nature Reviews. Genetics*, 1–17. https://doi.org/10.1038/s41576-021-00421-0

Alexandratos, N., Bruinsma, J., & Others. (2012). *World agriculture towards 2030/2050: the 2012 revision*. ESA Working paper FAO, Rome. http://large.stanford.edu/courses/2014/ph240/yuan2/docs/ap106e.pdf

Allen, H. K., Levine, U. Y., Looft, T., Bandrick, M., & Casey, T. A. (2013). Treatment, promotion, commotion: antibiotic alternatives in food-producing animals. *Trends in Microbiology*, *21*(3), 114–119. https://doi.org/10.1016/j.tim.2012.11.001

Amarasinghe, S. L., Su, S., Dong, X., Zappia, L., Ritchie, M. E., & Gouil, Q. (2020). Opportunities and challenges in long-read sequencing data analysis. *Genome Biology*, *21*(1), 1–16. https://doi.org/10.1186/s13059-020-1935-5

Ang, A., Roesma, D. I., Nijman, V., Meier, R., Srivathsan, A., & Rizaldi. (2020). Faecal DNA to the rescue: Shotgun sequencing of non-invasive samples reveals two subspecies of Southeast Asian primates to be Critically Endangered species. *Scientific Reports*, *10*(1), 9396. https://doi.org/10.1038/s41598-020-66007-8

Ansari, I., Raddatz, G., Gutekunst, J., Ridnik, M., Cohen, D., Abu-Remaileh, M., Tuganbaev, T., Shapiro, H., Pikarsky, E., Elinav, E., Lyko, F., & Bergman, Y. (2020). The microbiota programs DNA methylation to control intestinal homeostasis and inflammation. *Nature Microbiology*, *5*(4), 610–619. https://doi.org/10.1038/s41564-019-0659-3

Avila Santos, A. P., Kabiru Nata'ala, M., Kasmanas, J. C., Bartholomäus, A., Keller-Costa, T., Jurburg, S. D., Tal, T., Camarinha-Silva, A., Saraiva, J. P., Ponce de Leon Ferreira de Carvalho, A. C., Stadler, P. F., Sipoli Sanches, D., & Rocha, U. (2023). The AnimalAssociatedMetagenomeDB reveals a bias towards livestock and developed countries and blind spots in functional-potential studies of animal-associated microbiomes. *Animal Microbiome*, *5*(1), 48. https://doi.org/10.1186/s42523-023-00267-3

Bajury, D. M., Nashri, S. M., King Jie Hung, P., & Sarbini, S. R. (2018). Evaluation of potential prebiotics: a review. *Food Reviews International*, *34*(7), 639–664. https://doi.org/10.1080/87559129.2017.1373287

Barba-Vidal, E., Martín-Orúe, S. M., & Castillejos, L. (2019). Practical aspects of the use of probiotics in pig production: A review. *Livestock Science*, *223*, 84–96. https://doi.org/10.1016/j.livsci.2019.02.017

Bindari, Y. R., & Gerber, P. F. (2022). Centennial Review: Factors affecting the chicken gastrointestinal microbial composition and their association with gut health and productive performance. *Poultry Science*, *101*(1), 101612. https://doi.org/10.1016/j.psj.2021.101612

Blekhman, R., Goodrich, J. K., Huang, K., Sun, Q., Bukowski, R., Bell, J. T., Spector, T. D., Keinan, A., Ley, R. E., Gevers, D., & Clark, A. G. (2015). Host genetic variation impacts microbiome composition across human body sites. *Genome Biology*, *16*, 191. https://doi.org/10.1186/s13059-015-0759-1

Borda-Molina, D., Seifert, J., & Camarinha-Silva, A. (2018). Current Perspectives of the Chicken Gastrointestinal Tract and Its Microbiome. *Computational and Structural Biotechnology Journal*, *16*, 131–139. https://doi.org/10.1016/j.csbj.2018.03.002

Bordenstein, S. R., & Theis, K. R. (2015). Host Biology in Light of the Microbiome: Ten Principles of Holobionts and Hologenomes. *PLoS Biology*, *13*(8), e1002226. https://doi.org/10.1371/journal.pbio.1002226

Bourgeois, Y. X. C., & Warren, B. H. (2021). An overview of current population genomics methods for the analysis of whole-genome resequencing data in eukaryotes. *Molecular Ecology*, *30*(23), 6036–6071. https://doi.org/10.1111/mec.15989

Bowles, N., Alexander, S., & Hadjikakou, M. (2019). The livestock sector and planetary boundaries: A "limits to growth" perspective with dietary implications. *Ecological Economics: The Journal of the International Society for Ecological Economics*, *160*, 128–136. https://doi.org/10.1016/j.ecolecon.2019.01.033

Broom, L. J., & Kogut, M. H. (2018). The role of the gut microbiome in shaping the immune system of chickens. *Veterinary Immunology and Immunopathology*, *204*, 44–51. https://doi.org/10.1016/j.vetimm.2018.10.002

Campbell, B. M., Beare, D. J., Bennett, E. M., Hall-Spencer, J. M., Ingram, J. S. I., Jaramillo, F., Ortiz, R., Ramankutty, N., Sayer, J. A., & Shindell, D. (2017). Agriculture production as a major driver of the Earth system exceeding planetary boundaries. *Ecology and Society*, *22*(4). https://ora.ox.ac.uk/objects/uuid:25c20aa7-88b3-4457-9e4d-8ffbf57fc943

Cantalapiedra-Hijar, G., Abo-Ismail, M., Carstens, G. E., Guan, L. L., Hegarty, R., Kenny, D. A., McGee, M., Plastow, G., Relling, A., & Ortigues-Marty, I. (2018). Review: Biological determinants of between-animal variation in feed efficiency of growing beef cattle. *Animal: An International Journal of Animal Bioscience*, *12*(s2), s321–s335. https://doi.org/10.1017/S1751731118001489

Celi, P., Cowieson, A. J., Fru-Nji, F., Steinert, R. E., Kluenter, A.-M., & Verlhac, V. (2017). Gastrointestinal functionality in animal nutrition and health: New opportunities for sustainable animal production. *Animal Feed*

Science and Technology, *234*, 88–100. https://doi.org/10.1016/j.anifeedsci.2017.09.012

Costello, E. K., Stagaman, K., Dethlefsen, L., Bohannan, B. J. M., & Relman, D. A. (2012). The application of ecological theory toward an understanding of the human microbiome. *Science*, *336*(6086), 1255–1262. https://doi.org/10.1126/science.1224203

Coyte, K. Z., Schluter, J., & Foster, K. R. (2015). The ecology of the microbiome: Networks, competition, and stability. *Science*, *350*(6261), 663–666. https://doi.org/10.1126/science.aad2602

Culp, E. J., & Goodman, A. L. (2023). Cross-feeding in the gut microbiome: Ecology and mechanisms. *Cell Host & Microbe*, *31*(4), 485–499. https://doi.org/10.1016/j.chom.2023.03.016

de Melo Pereira, G. V., de Oliveira Coelho, B., Júnior, A. I. M., Thomaz-Soccol, V., & Soccol, C. R. (2018). How to select a probiotic? A review and update of methods and criteria. *Biotechnology Advances*. https://www.sciencedirect.com/science/article/pii/S0734975018301605

Diaz Carrasco, J. M., Casanova, N. A., & Fernández Miyakawa, M. E. (2019). Microbiota, Gut Health and Chicken Productivity: What Is the Connection? *Microorganisms*, *7*(10). https://doi.org/10.3390/microorganisms7100374

Dibner, J. J., & Richards, J. D. (2005). Antibiotic growth promoters in agriculture: history and mode of action. *Poultry Science*, *84*(4), 634–643. https://doi.org/10.1093/ps/84.4.634

Doherty, S. P., Foster, A., Best, J., Hamilton-Dyer, S., Morris, J., Sadler, P., Skelton, C., Smallman, R., Woldekiros, H., Thomas, R., & Sykes, N. (2021). Estimating the age of domestic fowl (Gallus gallus domesticus L. 1758) cockerels through spur development. *International Journal of Osteoarchaeology*, *31*(5), 770–781. https://doi.org/10.1002/oa.2988

Doolittle, W. F., & Booth, A. (2017). It's the song, not the singer: an exploration of holobiosis and evolutionary theory. *Biology & Philosophy*, *32*(1), 5–24. https://doi.org/10.1007/s10539-016-9542-2

Douglas, A. E., & Werren, J. H. (2016). Holes in the Hologenome: Why Host-Microbe Symbioses Are Not Holobionts. *mBio*, *7*(2), e02099. https://doi.org/10.1128/mBio.02099-15

Duke, G. E. (1989). Relationship of cecal and colonic motility to diet, habitat, and cecal anatomy in several avian species. *The Journal of Experimental Zoology. Supplement: Published under Auspices of the American Society of Zoologists and the Division of Comparative Physiology and Biochemistry / the Wistar Institute of Anatomy and Biology*, *3*, 38–47. https://doi.org/10.1002/jez.1402520507

Durazzi, F., Sala, C., Castellani, G., Manfreda, G., Remondini, D., & De Cesare, A. (2021). Comparison between 16S rRNA and shotgun sequencing data for the taxonomic characterization of the gut microbiota. *Scientific Reports*, *11*(1), 3030. https://doi.org/10.1038/s41598-021-82726-y

EFSA Panel on Additives and Products or Substances used in Animal Feed (FEEDAP), Rychen, G., Aquilina, G., Azimonti, G., Bampidis, V., Bastos, M. de L., Bories, G., Chesson, A., Cocconcelli, P. S., Flachowsky, G., Gropp, J., Kolar, B., Kouba, M., López-Alonso, M., López Puente, S., Mantovani, A., Mayo, B., Ramos, F., Saarela, M., … Martino, L. (2018). Guidance on the assessment of the efficacy of feed additives. *EFSA Journal*, *16*(5), e05274. https://doi.org/10.2903/j.efsa.2018.5274

Eisenhofer, R., Nesme, J., Santos-Bay, L., Koziol, A., Sørensen, S. J., Alberdi, A., & Aizpurua, O. (2023). A comparison of short-read, HiFi long-read, and hybrid strategies for genome-resolved metagenomics. In *bioRxiv* (p. 2023.10.04.560907). https://doi.org/10.1101/2023.10.04.560907

Ekblom, R., & Wolf, J. B. W. (2014). A field guide to whole-genome sequencing, assembly and annotation. *Evolutionary Applications*, *7*(9), 1026–1042. https://doi.org/10.1111/eva.12178

Fad, P. P. (2013). *POULTRY INDUSTRY MANUAL*. Citeseer. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=ffeacae51cb565b846aa2b3d3c196140b4a6615a

Fao/who. (2002). *Joint FAO/WHO (Food and Agriculture Organization/World Health Organization) working group report on drafting guidelines for the evaluation of probiotics in food*. FAO/WHO London.

Feighner, S. D., & Dashkevicz, M. P. (1987). Subtherapeutic levels of antibiotics in poultry feeds and their effects on weight gain, feed efficiency, and bacterial cholyltaurine hydrolase activity. *Applied and Environmental Microbiology*, *53*(2), 331–336. https://doi.org/10.1128/aem.53.2.331-336.1987

Feng, Y., Wang, Y., Zhu, B., Gao, G. F., Guo, Y., & Hu, Y. (2021). Metagenome-assembled genomes and gene catalog from the chicken gut microbiome aid in deciphering antibiotic resistomes. *Communications Biology*, *4*(1), 1–9. https://doi.org/10.1038/s42003-021-02827-2

Fraune, S., & Bosch, T. C. G. (2010). Why bacteria matter in animal development and evolution. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, *32*(7), 571–580. https://doi.org/10.1002/bies.200900192

Gadde, U., Kim, W. H., Oh, S. T., & Lillehoj, H. S. (2017). Alternatives to antibiotics for maximizing growth performance and feed efficiency in poultry: a review. *Animal Health Research Reviews / Conference of Research Workers in Animal Diseases*, *18*(1), 26–45. https://doi.org/10.1017/S1466252316000207

Gaggìa, F., Mattarelli, P., & Biavati, B. (2010). Probiotics and prebiotics in animal feeding for safe food production. *International Journal of Food Microbiology*, *141 Suppl 1*, S15–S28.

https://doi.org/10.1016/j.ijfoodmicro.2010.02.031

Gaskins, H. R., Collier, C. T., & Anderson, D. B. (2002). Antibiotics as growth promotants: mode of action. *Animal Biotechnology*, *13*(1), 29–42. https://doi.org/10.1081/ABIO-120005768

Gerber, P.J., Steinfeld, H., Henderson, B., Mottet, A., Opio, C., Dijkman, J., Falcucci, A. & Tempio, G. (2013). Tackling climate change through livestock: a global assessment of emissions and mitigation opportunities. *Food and Agriculture Organization of the United Nations (FAO),Rome*. https://www.cabdirect.org/cabdirect/abstract/20133417883

Gibson, G. R., Hutkins, R., Sanders, M. E., Prescott, S. L., Reimer, R. A., Salminen, S. J., Scott, K., Stanton, C., Swanson, K. S., Cani, P. D., Verbeke, K., & Reid, G. (2017). Expert consensus document: The International Scientific Association for Probiotics and Prebiotics (ISAPP) consensus statement on the definition and scope of prebiotics. *Nature Reviews. Gastroenterology & Hepatology*, *14*(8), 491–502. https://doi.org/10.1038/nrgastro.2017.75

Gilbert, S. F., Sapp, J., & Tauber, A. I. (2012). A symbiotic view of life: we have never been individuals. *The Quarterly Review of Biology*, *87*(4), 325–341. https://doi.org/10.1086/668166

Gilroy, R., Ravi, A., Getino, M., Pursley, I., Horton, D. L., Alikhan, N.-F., Baker, D., Gharbi, K., Hall, N., Watson, M., Adriaenssens, E. M., Foster-Nyarko, E., Jarju, S., Secka, A., Antonio, M., Oren, A., Chaudhuri, R. R., La Ragione, R., Hildebrand, F., & Pallen, M. J. (2021). Extensive microbial diversity within the chicken gut microbiome revealed by metagenomics and culture. *PeerJ*, *9*(e10941), e10941. https://doi.org/10.7717/peerj.10941

Glendinning, L., Stewart, R. D., Pallen, M. J., Watson, K. A., & Watson, M. (2020). Assembly of hundreds of novel bacterial genomes from the chicken caecum. *Genome Biology*, *21*(1), 34. https://doi.org/10.1186/s13059-020-1947-1

Godfray, H. C. J., Aveyard, P., Garnett, T., Hall, J. W., Key, T. J., Lorimer, J., Pierrehumbert, R. T., Scarborough, P., Springmann, M., & Jebb, S. A. (2018). Meat consumption, health, and the environment. *Science*, *361*(6399). https://doi.org/10.1126/science.aam5324

Gorbach, S. L. (2001). Antimicrobial Use in Animal Feed — Time to Stop. *The New England Journal of Medicine*, *345*(16), 1202–1203. https://doi.org/10.1056/NEJM200110183451610

Graw, S., Chappell, K., Washam, C. L., Gies, A., Bird, J., Robeson, M. S., 2nd, & Byrum, S. D. (2021). Multi-omics data integration considerations and study design for biological systems and disease. *Molecular Omics*, *17*(2), 170–185. https://doi.org/10.1039/d0mo00041h

Gribble, F. M., & Reimann, F. (2019). Function and mechanisms of enteroendocrine cells and gut hormones in metabolism. *Nature Reviews. Endocrinology*, *15*(4), 226–237. https://doi.org/10.1038/s41574-019-0168-8

Havenstein, G. B., Ferket, P. R., & Qureshi, M. A. (2003a). Carcass composition and yield of 1957 versus 2001 broilers when fed representative 1957 and 2001 broiler diets. *Poultry Science*, *82*(10), 1509–1518. https://doi.org/10.1093/ps/82.10.1509

Havenstein, G. B., Ferket, P. R., & Qureshi, M. A. (2003b). Growth, livability, and feed conversion of 1957 versus 2001 broilers when fed representative 1957 and 2001 broiler diets. *Poultry Science*, *82*(10), 1500–1508. https://doi.org/10.1093/ps/82.10.1500

Havenstein, G. B., Ferket, P. R., Scheideler, S. E., & Larson, B. T. (1994). Growth, livability, and feed conversion of 1957 vs 1991 broilers when fed “typical” 1957 and 1991 broiler diets. *Poultry Science*, *73*(12), 1785–1794. https://doi.org/10.3382/ps.0731785

He, Z., Liu, R., Wang, M., Wang, Q., Zheng, J., Ding, J., Wen, J., Fahey, A. G., & Zhao, G. (2023). Combined effect of microbially derived cecal SCFA and host genetics on feed efficiency in broiler chickens. *Microbiome*, *11*(1), 198. https://doi.org/10.1186/s40168-023-01627-6

HoloFood Consortium. (2019). *Holistic solution to improve animal food production through deconstructing the biomolecular interactions between feed, gut microorganisms and animals in relation to performance parameters*. CORDIS. https://cordis.europa.eu/project/id/817729/es

Huang, P., Zhang, Y., Xiao, K., Jiang, F., Wang, H., Tang, D., Liu, D., Liu, B., Liu, Y., He, X., Liu, H., Liu, X., Qing, Z., Liu, C., Huang, J., Ren, Y., Yun, L., Yin, L., Lin, Q., … Zeng, J. (2018). The chicken gut metagenome and the modulatory effects of plant-derived benzylisoquinoline alkaloids. *Microbiome*, *6*(1), 211. https://doi.org/10.1186/s40168-018-0590-5

Jaspers, C., Fraune, S., Arnold, A. E., Miller, D. J., Bosch, T. C. G., Voolstra, C. R., & Consortium of Australian Academy of Science Boden Research Conference Participants. (2019). Resolving structure and function of metaorganisms through a holistic framework combining reductionist and integrative approaches. *Zoology* , *133*, 81–87. https://doi.org/10.1016/j.zool.2019.02.007

Jiang, P., Lai, S., Wu, S., Zhao, X.-M., & Chen, W.-H. (2020). Host DNA contents in fecal metagenomics as a biomarker for intestinal diseases and effective treatment. *BMC Genomics*, *21*(1), 348. https://doi.org/10.1186/s12864-020-6749-z

Jones, H. E., & Wilson, P. B. (2022). Progress and opportunities through use of genomics in animal production. *Trends in Genetics: TIG*, *38*(12), 1228–1252. https://doi.org/10.1016/j.tig.2022.06.014

Jung, H., Ventura, T., Chung, J. S., Kim, W.-J., Nam, B.-H., Kong, H. J., Kim, Y.-O., Jeon, M.-S., & Eyun, S.-I.

(2020). Twelve quick steps for genome assembly and annotation in the classroom. *PLoS Computational Biology*, *16*(11), e1008325. https://doi.org/10.1371/journal.pcbi.1008325

Jurburg, S. D., Brouwer, M. S. M., Ceccarelli, D., van der Goot, J., Jansman, A. J. M., & Bossers, A. (2019). Patterns of community assembly in the developing chicken microbiome reveal rapid primary succession. *MicrobiologyOpen*, *8*(9), e00821. https://doi.org/10.1002/mbo3.821

Khan, M. A. W., Stephens, W. Z., Mohammed, A. D., Round, J. L., & Kubinak, J. L. (2019). Does MHC heterozygosity influence microbiota form and function? *PloS One*, *14*(5), e0215946. https://doi.org/10.1371/journal.pone.0215946

Kim, W. H., & Lillehoj, H. S. (2019). Immunity, immunomodulation, and antibiotic alternatives to maximize the genetic potential of poultry for growth and disease response. *Animal Feed Science and Technology*, *250*, 41–50. https://doi.org/10.1016/j.anifeedsci.2018.09.016

Kogut, M. H. (2019). The effect of microbiome modulation on the intestinal health of poultry. *Animal Feed Science and Technology*, *250*, 32–40. https://doi.org/10.1016/j.anifeedsci.2018.10.008

Kogut, M. H., & Arsenault, R. J. (2016). Editorial: Gut Health: The New Paradigm in Food Animal Production. *Frontiers in Veterinary Science*, *3*, 71. https://doi.org/10.3389/fvets.2016.00071

Kogut, M. H., Lee, A., & Santin, E. (2020). Microbiome and pathogen interaction with the immune system. *Poultry Science*, *99*(4), 1906–1913. https://doi.org/10.1016/j.psj.2019.12.011

Kozicka, M., Havlík, P., Valin, H., Wollenberg, E., Deppermann, A., Leclère, D., Lauri, P., Moses, R., Boere, E., Frank, S., Davis, C., Park, E., & Gurwick, N. (2023). Feeding climate and biodiversity goals with novel plant-based meat and milk alternatives. *Nature Communications*, *14*(1), 5316. https://doi.org/10.1038/s41467-023-40899-2

Krautkramer, K. A., Fan, J., & Bäckhed, F. (2021). Gut microbial metabolites as multi-kingdom intermediates. *Nature Reviews. Microbiology*, *19*(2), 77–94. https://doi.org/10.1038/s41579-020-0438-4

Kwoji, I. D., Aiyegoro, O. A., Okpeku, M., & Adeleke, M. A. (2023). "Multi-omics" data integration: applications in probiotics studies. *NPJ Science of Food*, *7*(1), 25. https://doi.org/10.1038/s41538-023-00199-x

Langa, J., Huret, M., Montes, I., Conklin, D., & Estonba, A. (2021). Transcriptomic dataset for Sardina pilchardus: Assembly, annotation, and expression of nine tissues. *Data in Brief*, *39*, 107583. https://doi.org/10.1016/j.dib.2021.107583

Lawal, R. A., & Hanotte, O. (2021). Domestic chicken diversity: Origin, distribution, and adaptation. *Animal Genetics*. https://doi.org/10.1111/age.13091

Laxminarayan, R., Duse, A., Wattal, C., Zaidi, A. K. M., Wertheim, H. F. L., Sumpradit, N., Vlieghe, E., Hara, G. L., Gould, I. M., Goossens, H., Greko, C., So, A. D., Bigdeli, M., Tomson, G., Woodhouse, W., Ombaka, E., Peralta, A. Q., Qamar, F. N., Mir, F., … Cars, O. (2013). Antibiotic resistance—the need for global solutions. *The Lancet Infectious Diseases*, *13*(12), 1057–1098. https://doi.org/10.1016/S1473-3099(13)70318-9

Lichtman, J. S., Sonnenburg, J. L., & Elias, J. E. (2015). Monitoring host responses to the gut microbiota. *The ISME Journal*, *9*(9), 1908–1915. https://doi.org/10.1038/ismej.2015.93

Limborg, M. T., Alberdi, A., Kodama, M., Roggenbuck, M., Kristiansen, K., & Gilbert, M. T. P. (2018). Applied Hologenomics: Feasibility and Potential in Aquaculture. *Trends in Biotechnology*, *36*(3), 252–264. https://doi.org/10.1016/j.tibtech.2017.12.006

Liu, Y., Espinosa, C. D., Abelilla, J. J., Casas, G. A., Lagos, L. V., Lee, S. A., Kwon, W. B., Mathai, J. K., Navarro, D. M. D. L., Jaworski, N. W., & Stein, H. H. (2018). Non-antibiotic feed additives in diets for pigs: A review. *Animal Nutrition (Zhongguo Xu Mu Shou Yi Xue Hui)*, *4*(2), 113–125. https://doi.org/10.1016/j.aninu.2018.01.007

Liu, Y.-X., Qin, Y., Chen, T., Lu, M., Qian, X., Guo, X., & Bai, Y. (2021). A practical guide to amplicon and metagenomic analysis of microbiome data. *Protein & Cell*, *12*(5), 315–330. https://doi.org/10.1007/s13238-020-00724-8

Liu, Y., Zheng, Z., Yu, L., Wu, S., Sun, L., Wu, S., Xu, Q., Cai, S., Qin, N., & Bao, W. (2019). Examination of the temporal and spatial dynamics of the gut microbiome in newborn piglets reveals distinct microbial communities in six intestinal segments. *Scientific Reports*, *9*(1), 3453. https://doi.org/10.1038/s41598-019-40235-z

Lundberg, R., Scharch, C., & Sandvang, D. (2021). The link between broiler flock heterogeneity and cecal microbiome composition. *Animal Microbiome*, *3*(1), 54. https://doi.org/10.1186/s42523-021-00110-7

Małyska, A., Markakis, M. N., Pereira, C. F., & Cornelissen, M. (2019). The Microbiome: A Life Science Opportunity for Our Society and Our Planet. *Trends in Biotechnology*, *37*(12), 1269–1272. https://doi.org/10.1016/j.tibtech.2019.06.008

Manyi-Loh, C., Mamphweli, S., Meyer, E., & Okoh, A. (2018). Antibiotic Use in Agriculture and Its Consequential Resistance in Environmental Sources: Potential Public Health Implications. *Molecules* , *23*(4). https://doi.org/10.3390/molecules23040795

Markowiak, P., & Śliżewska, K. (2018). The role of probiotics, prebiotics and synbiotics in animal nutrition. *Gut Pathogens*, *10*, 21. https://doi.org/10.1186/s13099-018-0250-0

McFall-Ngai, M., Hadfield, M. G., Bosch, T. C. G., Carey, H. V., Domazet-Lošo, T., Douglas, A. E., Dubilier, N., Eberl, G., Fukami, T., Gilbert, S. F., Hentschel, U., King, N., Kjelleberg, S., Knoll, A. H., Kremer, N., Mazmanian, S. K., Metcalf, J. L., Nealson, K., Pierce, N. E., … Wernegreen, J. J. (2013). Animals in a bacterial world, a new imperative for the life sciences. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(9), 3229–3236. https://doi.org/10.1073/pnas.1218525110

Messerli, P., Murniningtyas, E., & Eloundou-Enyegue, P. (2019). *Global sustainable development report 2019: the future is now–science for achieving sustainable development*. pure.iiasa.ac.at. http://pure.iiasa.ac.at/id/eprint/16067/1/24797GSDR_report_2019.pdf

Moran, N. A., & Sloan, D. B. (2015). The Hologenome Concept: Helpful or Hollow? *PLoS Biology*, *13*(12), e1002311. https://doi.org/10.1371/journal.pbio.1002311

Niewold, T. A. (2007). The nonantibiotic anti-inflammatory effect of antimicrobial growth promoters, the real mode of action? A hypothesis. *Poultry Science*, *86*(4), 605–609. https://doi.org/10.1093/ps/86.4.605

Nyholm, L., Koziol, A., Marcos, S., Botnen, A. B., Aizpurua, O., Gopalakrishnan, S., Limborg, M. T., Gilbert, M. T. P., & Alberdi, A. (2020). Holo-Omics: Integrated Host-Microbiota Multi-omics for Basic and Applied Biological Research. *iScience*, *23*(8), 101414. https://doi.org/10.1016/j.isci.2020.101414

Nyholm, L., Odriozola, I., Martin Bideguren, G., Aizpurua, O., & Alberdi, A. (2022). Gut microbiota differences between paired intestinal wall and digesta samples in three small species of fish. *PeerJ*, *10*, e12992. https://doi.org/10.7717/peerj.12992

Oakley, B. B., Lillehoj, H. S., Kogut, M. H., Kim, W. K., Maurer, J. J., Pedroso, A., Lee, M. D., Collett, S. R., Johnson, T. J., & Cox, N. A. (2014). The chicken gastrointestinal microbiome. *FEMS Microbiology Letters*, *360*(2), 100–112. https://doi.org/10.1111/1574-6968.12608

OECD-FAO Agricultural Outlook 2017-2026 - en - OECD. (n.d.). Retrieved July 13, 2023, from https://www.oecd.org/development/oecd-fao-agricultural-outlook-19991142.htm

Ojala, T., Häkkinen, A.-E., Kankuri, E., & Kankainen, M. (2023). Current concepts, advances, and challenges in deciphering the human microbiota with metatranscriptomics. *Trends in Genetics: TIG*, *39*(9), 686–702. https://doi.org/10.1016/j.tig.2023.05.004

Okocha, R. C., Olatoye, I. O., & Adedeji, O. B. (2018). Food safety impacts of antimicrobial use and their residues in aquaculture. *Public Health Reviews*, *39*, 21. https://doi.org/10.1186/s40985-018-0099-2

Orlando, L., Allaby, R., Skoglund, P., Der Sarkissian, C., Stockhammer, P. W., Ávila-Arcos, M. C., Fu, Q., Krause, J., Willerslev, E., Stone, A. C., & Warinner, C. (2021). Ancient DNA analysis. *Nature Reviews Methods Primers*, *1*(1), 1–26. https://doi.org/10.1038/s43586-020-00011-0

Pandey, A. K., Kumar, P., & Saxena, M. J. (2019). Feed Additives in Animal Health. In R. C. Gupta, A. Srivastava, & R. Lall (Eds.), *Nutraceuticals in Veterinary Medicine* (pp. 345–362). Springer International Publishing. https://doi.org/10.1007/978-3-030-04624-8_23

Pan, D., & Yu, Z. (2014). Intestinal microbiome of poultry and its interaction with host and diet. *Gut Microbes*, *5*(1), 108–119. https://doi.org/10.4161/gmic.26945

Parejo, M., Wragg, D., Henriques, D., Charrière, J.-D., & Estonba, A. (2020). Digging into the Genomic Past of Swiss Honey Bees by Whole-Genome Sequencing Museum Specimens. *Genome Biology and Evolution*, *12*(12), 2535–2551. https://doi.org/10.1093/gbe/evaa188

Parliament and Council, E. (2003). Regulation (EC) No 1831/2003 of the European Parliament and of the Council of 22 September 2003 on additives for use in animal nutrition. *Off. J. Eur. Union*.

Pervez, M. T., Hasnain, M. J. U., Abbas, S. H., Moustafa, M. F., Aslam, N., & Shah, S. S. M. (2022). A Comprehensive Review of Performance of Next-Generation Sequencing Platforms. *BioMed Research International*, *2022*, 3457806. https://doi.org/10.1155/2022/3457806

Proszkowiec-Weglarz, M., Miska, K. B., Ellestad, L. E., Schreier, L. L., Kahl, S., Darwish, N., Campos, P., & Shao, J. (2022). Delayed access to feed early post-hatch affects the development and maturation of gastrointestinal tract microbiota in broiler chickens. *BMC Microbiology*, *22*(1), 206. https://doi.org/10.1186/s12866-022-02619-6

Qanbari, S., Rubin, C.-J., Maqbool, K., Weigend, S., Weigend, A., Geibel, J., Kerje, S., Wurmser, C., Peterson, A. T., Brisbin, I. L., Jr, Preisinger, R., Fries, R., Simianer, H., & Andersson, L. (2019). Genetics of adaptation in modern chicken. *PLoS Genetics*, *15*(4), e1007989. https://doi.org/10.1371/journal.pgen.1007989

Quan, J., Cai, G., Ye, J., Yang, M., Ding, R., Wang, X., Zheng, E., Fu, D., Li, S., Zhou, S., Liu, D., Yang, J., & Wu, Z. (2018). A global comparison of the microbiome compositions of three gut locations in commercial pigs with extreme feed conversion ratios. *Scientific Reports*, *8*(1), 4536. https://doi.org/10.1038/s41598-018-22692-0

Quince, C., Walker, A. W., Simpson, J. T., Loman, N. J., & Segata, N. (2017). Shotgun metagenomics, from sampling to analysis. *Nature Biotechnology*, *35*(9), 833–844. https://doi.org/10.1038/nbt.3935

Rasmussen, J. A., Villumsen, K. R., Ernst, M., Hansen, M., Forberg, T., Gopalakrishnan, S., Gilbert, M. T. P., Bojesen, A. M., Kristiansen, K., & Limborg, M. T. (2022). A multi-omics approach unravels metagenomic and metabolic alterations of a probiotic and synbiotic additive in rainbow trout (Oncorhynchus mykiss). *Microbiome*, *10*(1), 21. https://doi.org/10.1186/s40168-021-01221-8

Reid, W. V., Chen, D., Goldfarb, L., Hackmann, H., Lee, Y. T., Mokhele, K., Ostrom, E., Raivio, K., Rockström, J., Schellnhuber, H. J., & Whyte, A. (2010). Earth System Science for Global Sustainability: Grand Challenges. *Science*, *330*(6006), 916–917. https://doi.org/10.1126/science.1196263

Rhie, A., McCarthy, S. A., Fedrigo, O., Damas, J., Formenti, G., Koren, S., Uliano-Silva, M., Chow, W., Fungtammasan, A., Kim, J., Lee, C., Ko, B. J., Chaisson, M., Gedman, G. L., Cantin, L. J., Thibaud-Nissen, F., Haggerty, L., Bista, I., Smith, M., … Jarvis, E. D. (2021). Towards complete and error-free genome assemblies of all vertebrate species. *Nature*, *592*(7856), 737–746. https://doi.org/10.1038/s41586-021-03451-0

Ricciardi, V., Mehrabi, Z., Wittman, H., James, D., & Ramankutty, N. (2021). Higher yields and more biodiversity on smaller farms. *Nature Sustainability*, *4*(7), 651–657. https://doi.org/10.1038/s41893-021-00699-2

Richards, P., Fothergill, J., Bernardeau, M., & Wigley, P. (2019). Development of the Caecal Microbiota in Three Broiler Breeds. *Frontiers in Veterinary Science*, *6*, 201. https://doi.org/10.3389/fvets.2019.00201

Robles-Jimenez, L. E., Aranda-Aguirre, E., Castelan-Ortega, O. A., Shettino-Bermudez, B. S., Ortiz-Salinas, R., Miranda, M., Li, X., Angeles-Hernandez, J. C., Vargas-Bello-Pérez, E., & Gonzalez-Ronquillo, M. (2021). Worldwide Traceability of Antibiotic Residues from Livestock in Wastewater and Soil: A Systematic Review. *Animals : An Open Access Journal from MDPI*, *12*(1). https://doi.org/10.3390/ani12010060

Rubin, C.-J., Zody, M. C., Eriksson, J., Meadows, J. R. S., Sherwood, E., Webster, M. T., Jiang, L., Ingman, M., Sharpe, T., Ka, S., Hallböök, F., Besnier, F., Carlborg, O., Bed'hom, B., Tixier-Boichard, M., Jensen, P., Siegel, P., Lindblad-Toh, K., & Andersson, L. (2010). Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature*, *464*(7288), 587–591. https://doi.org/10.1038/nature08832

Rychlik, I. (2020). Composition and Function of Chicken Gut Microbiota. *Animals*. https://www.mdpi.com/2076-2615/10/1/103

Santiago-Rodriguez, T. M., & Hollister, E. B. (2021). Multi 'omic data integration: A review of concepts, considerations, and approaches. *Seminars in Perinatology*, *45*(6), 151456. https://doi.org/10.1016/j.semperi.2021.151456

Schokker, D., Veninga, G., Vastenhouw, S. A., Bossers, A., de Bree, F. M., Kaal-Lansbergen, L. M. T. E., Rebel, J. M. J., & Smits, M. A. (2015). Early life microbial colonization of the gut and intestinal development differ between genetically divergent broiler lines. *BMC Genomics*, *16*, 418. https://doi.org/10.1186/s12864-015-1646-6

Segura-Wang, M., Grabner, N., Koestelbauer, A., Klose, V., & Ghanbari, M. (2021). Genome-Resolved Metagenomics of the Chicken Gut Microbiome. *Frontiers in Microbiology*, *12*, 726923. https://doi.org/10.3389/fmicb.2021.726923

Sessitsch, A., & Mitter, B. (2015). 21st century agriculture: integration of plant microbiomes for improved crop production and food security. *Microbial Biotechnology*, *8*(1), 32–33. https://doi.org/10.1111/1751-7915.12180

Shah, T. M., Patel, J. G., Gohil, T. P., Blake, D. P., & Joshi, C. G. (2019). Host transcriptome and microbiome interaction modulates physiology of full-sibs broilers with divergent feed conversion ratio. *NPJ Biofilms and Microbiomes*, *5*, 24. https://doi.org/10.1038/s41522-019-0096-3

Siegel, P. B. (2014). Evolution of the modern broiler and feed efficiency. *Annual Review of Animal Biosciences*, *2*, 375–385. https://doi.org/10.1146/annurev-animal-022513-114132

Soccol, C. R., de Souza Vandenberghe, L. P., Spier, M. R., Medeiros, A. B. P., Yamaguishi, C. T., De Dea Lindner, J., Pandey, A., & Thomaz-Soccol, V. (2010). The potential of probiotics: a review. *Food Technology and Biotechnology*, *48*(4), 413–434. https://hrcak.srce.hr/file/92463.html

Springmann, M., Clark, M., Mason-D'Croz, D., Wiebe, K., Bodirsky, B. L., Lassaletta, L., de Vries, W., Vermeulen, S. J., Herrero, M., Carlson, K. M., Jonell, M., Troell, M., DeClerck, F., Gordon, L. J., Zurayk, R., Scarborough, P., Rayner, M., Loken, B., Fanzo, J., … Willett, W. (2018). Options for keeping the food system within environmental limits. *Nature*, *562*(7728), 519–525. https://doi.org/10.1038/s41586-018-0594-0

Sprockett, D., Fukami, T., & Relman, D. A. (2018). Role of priority effects in the early-life assembly of the gut microbiota. *Nature Reviews. Gastroenterology & Hepatology*, *15*(4), 197–205. https://doi.org/10.1038/nrgastro.2017.173

Stewart, R. D., Auffret, M. D., Warr, A., Walker, A. W., Roehe, R., & Watson, M. (2019). Compendium of 4,941 rumen metagenome-assembled genomes for rumen microbiome biology and enzyme discovery. *Nature Biotechnology*, *37*(8), 953–961. https://doi.org/10.1038/s41587-019-0202-3

Stokstad, E. L. R., & Jukes, T. H. (1950). Further Observations on the "Animal Protein Factor." *Proceedings of the Society for Experimental Biology and Medicine. Society for Experimental Biology and Medicine* , *73*(3), 523–528. https://doi.org/10.3181/00379727-73-17731

Suárez, J. (2018). "The importance of symbiosis in philosophy of biology: an analysis of the current debate on biological individuality and its historical roots." *Symbiosis* , *76*(2), 77–96. https://doi.org/10.1007/s13199-018-0556-1

Suez, J., Zmora, N., Segal, E., & Elinav, E. (2019). The pros, cons, and many unknowns of probiotics. *Nature Medicine*. https://www.nature.com/articles/s41591-019-0439-x

Svihus, B. (2014). Function of the digestive system1 1Presented as a part of the Informal Nutrition Symposium

"From Research Measurements to Application: Bridging the Gap" at the Poultry Science Association's annual meeting in San Diego, California, July 22–25, 2013. *The Journal of Applied Poultry Research*, *23*(2), 306–314. https://doi.org/10.3382/japr.2014-00937

Svihus, B., & Choct, M. (2013). Function and nutritional roles of the avian caeca: a review. *World's Poultry Science Journal*, *69*(2), 249–264. https://doi.org/10.1017/S0043933913000287

Sykes, N. (2012). A social perspective on the introduction of exotic animals: the case of the chicken. *World Archaeology*, *44*(1), 158–169. https://doi.org/10.1080/00438243.2012.646104

Talebi, R., Szmatoła, T., Mészáros, G., & Qanbari, S. (2020). Runs of Homozygosity in Modern Chicken Revealed by Sequence Data. *G3* , *10*(12), 4615–4623. https://doi.org/10.1534/g3.120.401860

Tan, A., Murugapiran, S., Mikalauskas, A., Koble, J., Kennedy, D., Hyde, F., Ruotti, V., Law, E., Jensen, J., Schroth, G. P., Macklaim, J. M., Kuersten, S., LeFrançois, B., & Gohl, D. M. (2023). Rational probe design for efficient rRNA depletion and improved metatranscriptomic analysis of human microbiomes. *BMC Microbiology*, *23*(1), 299. https://doi.org/10.1186/s12866-023-03037-y

Tarsani, E., Kranis, A., Maniatis, G., Avendano, S., Hager-Theodorides, A. L., & Kominakis, A. (2019). Discovery and characterization of functional modules associated with body weight in broilers. *Scientific Reports*, *9*(1), 9125. https://doi.org/10.1038/s41598-019-45520-5

Thornton, P. K. (2010). Livestock production: recent trends, future prospects. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *365*(1554), 2853–2867. https://doi.org/10.1098/rstb.2010.0134

Tikhonov, G., Opedal, Ø. H., Abrego, N., Lehikoinen, A., de Jonge, M. M. J., Oksanen, J., & Ovaskainen, O. (2020). Joint species distribution modelling with the r-package Hmsc. *Methods in Ecology and Evolution / British Ecological Society*, *11*(3), 442–447. https://doi.org/10.1111/2041-210X.13345

Van Boeckel, T. P., Brower, C., Gilbert, M., Grenfell, B. T., Levin, S. A., Robinson, T. P., Teillant, A., & Laxminarayan, R. (2015). Global trends in antimicrobial use in food animals. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(18), 5649–5654. https://doi.org/10.1073/pnas.1503141112

van der Most, P. J., de Jong, B., Parmentier, H. K., & Verhulst, S. (2011). Trade‐off between growth and immune function: a meta‐analysis of selection experiments. *Functional Ecology*, *25*(1), 74–80. https://doi.org/10.1111/j.1365-2435.2010.01800.x

Vasdal, G., Granquist, E. G., Skjerve, E., de Jong, I. C., Berg, C., Michel, V., & Moe, R. O. (2019). Associations between carcass weight uniformity and production measures on farm and at slaughter in commercial broiler flocks. *Poultry Science*, *98*(10), 4261–4268. https://doi.org/10.3382/ps/pez252

Venturelli, O. S., Carr, A. C., Fisher, G., Hsu, R. H., Lau, R., Bowen, B. P., Hromada, S., Northen, T., & Arkin, A. P. (2018). Deciphering microbial interactions in synthetic human gut microbiome communities. *Molecular Systems Biology*, *14*(6), e8157. https://doi.org/10.15252/msb.20178157

Verstegen, M. W. A., & Williams, B. A. (2002). Alternatives to the use of antibiotics as growth promoters for monogastric animals. *Animal Biotechnology*, *13*(1), 113–127. https://doi.org/10.1081/ABIO-120005774

Vinderola, G., Gueimonde, M., Gomez-Gallego, C., Delfederico, L., & Salminen, S. (2017). Correlation between in vitro and in vivo assays in selection of probiotics from traditional species of bacteria. *Trends in Food Science & Technology*, *68*, 83–90. https://doi.org/10.1016/j.tifs.2017.08.005

Wang, M.-S., Thakur, M., Peng, M.-S., Jiang, Y., Frantz, L. A. F., Li, M., Zhang, J.-J., Wang, S., Peters, J., Otecko, N. O., Suwannapoom, C., Guo, X., Zheng, Z.-Q., Esmailizadeh, A., Hirimuthugoda, N. Y., Ashari, H., Suladari, S., Zein, M. S. A., Kusza, S., … Zhang, Y.-P. (2020). 863 genomes reveal the origin and domestication of chicken. *Cell Research*, *30*(8), 693–701. https://doi.org/10.1038/s41422-020-0349-y

Wang, M.-S., Zhang, J.-J., Guo, X., Li, M., Meyer, R., Ashari, H., Zheng, Z.-Q., Wang, S., Peng, M.-S., Jiang, Y., Thakur, M., Suwannapoom, C., Esmailizadeh, A., Hirimuthugoda, N. Y., Zein, M. S. A., Kusza, S., Kharrati-Koopaee, H., Zeng, L., Wang, Y.-M., … Zhang, Y.-P. (2021). Large-scale genomic analysis reveals the genetic cost of chicken domestication. *BMC Biology*, *19*(1), 118. https://doi.org/10.1186/s12915-021-01052-x

Wang, Q., Wang, K., Wu, W., Giannoulatou, E., Ho, J. W. K., & Li, L. (2019). Host and microbiome multi-omics integration: applications and methodologies. *Biophysical Reviews*, *11*(1), 55–65. https://doi.org/10.1007/s12551-018-0491-7

Wang, X., Tsai, T., Deng, F., Wei, X., Chai, J., Knapp, J., Apple, J., Maxwell, C. V., Lee, J. A., Li, Y., & Zhao, J. (2019). Longitudinal investigation of the swine gut microbiome from birth to market reveals stage and growth performance associated bacteria. *Microbiome*, *7*(1), 109. https://doi.org/10.1186/s40168-019-0721-7

Wan, T., Wang, Y., He, K., & Zhu, S. (2023). Microbial sensing in the intestine. *Protein & Cell*, *14*(11), 824–860. https://doi.org/10.1093/procel/pwad028

Westermann, A. J., & Vogel, J. (2021). Cross-species RNA-seq for deciphering host–microbe interactions. *Nature Reviews. Genetics*, *22*(6), 361–378. https://doi.org/10.1038/s41576-021-00326-y

Willett, W., Rockström, J., Loken, B., Springmann, M., Lang, T., Vermeulen, S., Garnett, T., Tilman, D., DeClerck, F., Wood, A., Jonell, M., Clark, M., Gordon, L. J., Fanzo, J., Hawkes, C., Zurayk, R., Rivera, J. A., De Vries,

W., Majele Sibanda, L., … Murray, C. J. L. (2019). Food in the Anthropocene: the EAT–Lancet Commission on healthy diets from sustainable food systems. *The Lancet*, *393*(10170), 447–492. https://doi.org/10.1016/S0140-6736(18)31788-4

Wu, J.-J., Zhu, S., Tang, Y.-F., Gu, F., Liu, J.-X., & Sun, H.-Z. (2022). Microbiota-host crosstalk in the newborn and adult rumen at single-cell resolution. *BMC Biology*, *20*(1), 280. https://doi.org/10.1186/s12915-022-01490-1

Xiao, L., Estellé, J., Kiilerich, P., Ramayo-Caldas, Y., Xia, Z., Feng, Q., Liang, S., Pedersen, A. Ø., Kjeldsen, N. J., Liu, C., Maguin, E., Doré, J., Pons, N., Le Chatelier, E., Prifti, E., Li, J., Jia, H., Liu, X., Xu, X., … Wang, J. (2016). A reference gene catalogue of the pig gut microbiome. *Nature Microbiology*, 16161. https://doi.org/10.1038/nmicrobiol.2016.161

Zhang, Y., Jiang, F., Yang, B., Wang, S., Wang, H., Wang, A., Xu, D., & Fan, W. (2022). Improved microbial genomes and gene catalog of the chicken gut from metagenomic sequencing of high-fidelity long reads. *GigaScience*, *11*. https://doi.org/10.1093/gigascience/giac116

Zhao, Q., & Elson, C. O. (2018). Adaptive immune education by gut microbiota antigens. *Immunology*, *154*(1), 28–37. https://doi.org/10.1111/imm.12896

Zhou, H., Ni, W.-J., Meng, X.-M., & Tang, L.-Q. (2020). MicroRNAs as Regulators of Immune and Inflammatory Responses: Potential Therapeutic Targets in Diabetic Nephropathy. *Frontiers in Cell and Developmental Biology*, *8*, 618536. https://doi.org/10.3389/fcell.2020.618536

Zhou, H., Yang, L., Ding, J., Xu, K., Liu, J., Zhu, W., Zhu, J., He, C., Han, C., Qin, C., Luo, H., Chen, K., Zheng, Y., Honaker, C. F., Zhang, Y., Siegel, P. B., & Meng, H. (2022). Dynamics of Small Non-coding RNA Profiles and the Intestinal Microbiome of High and Low Weight Chickens. *Frontiers in Microbiology*, *13*. https://doi.org/10.3389/fmicb.2022.916280

Zhou, Q., Lan, F., Li, X., Yan, W., Sun, C., Li, J., Yang, N., & Wen, C. (2021). The Spatial and Temporal Characterization of Gut Microbiota in Broilers. *Frontiers in Veterinary Science*, *8*, 712226. https://doi.org/10.3389/fvets.2021.712226

Zilber-Rosenberg, I., & Rosenberg, E. (2008). Role of microorganisms in the evolution of animals and plants: the hologenome theory of evolution. *FEMS Microbiology Reviews*, *32*(5), 723–735. https://doi.org/10.1111/j.1574-6976.2008.00123.x

Zilber-Rosenberg, I., & Rosenberg, E. (2013). *The Hologenome Concept- Human, Animal and Plant Microbiota*. Springer International Publishing Switzerland 2013. https://doi.org/10.1007/978-3-319-04241-1

Zommiti, M., Chikindas, M. L., & Ferchichi, M. (2020). Probiotics—Live Biotherapeutics: a Story of Success, Limitations, and Future Prospects—Not Only for Humans. *Probiotics and Antimicrobial Proteins*, *12*(3), 1266–1289. https://doi.org/10.1007/s12602-019-09570-5

Thesis roadmap, hypothesis, aims and outline

Chapter 2

Thesis roadmap and my contribution to HoloFood

During my four and a half years being part of HoloFood, I participated in numerous experimental, laboratory and computational tasks that were developed within the project as part of my PhD. Initially, I was offered a position in 2019 to work as a researcher in HoloFood within the Applied Genomics and Bioinformatics group, directed by Andone Estonba. A year later I was granted a predoctoral fellowship by the Basque Government in the same group, under which this dissertation has been developed. Given the complexity of developing a doctoral thesis as part of a European Project, the following paragraphs attempt to detail i) the steps followed in HoloFood, ii) my involvement in various of those steps, and iii) the development of the research papers that constitute the chapters of this thesis.

Chicken experimental trials were carried out at the Institute of Agrifood Research and Technology (IRTA), Spain. The trials were performed between February and October 2019, and during this period, I made 14 trips (of 3-4 days each) to IRTA. I participated in the design of the protocols for the sampling procedures together with Dr. Antton Alberdi and Dr. Joan Tarradas, researcher from the Animal Nutrition Department (IRTA), under the supervision of Dr. Enric Esteve-Garcia. During the sampling procedures, I had the opportunity to learn different protocols for collecting a variety of samples from multiple organs, as we collected more than 30 samples per animal, for a total of 1,296 chicken individuals. Nevertheless, it must be noted that the experimental and sampling work would not be possible without all the researchers, farm workers and laboratory technicians that participated in these tasks.

Between 2019 and 2020, I did 3 internships at the Animal-microbiota interactions group, Center for Evolutionary Hologenomics (CEH) Denmark, led by co-supervisor Antton Alberdi. The internships responded to the decision of processing all samples for multi-omic data generation at the University of Copenhagen. During those internships, I worked on protocol optimization and processing of ileal and caecal (meta)genomic and (meta)transcriptomic samples, under the supervision of the laboratory manager Dr. Sarah Mak. I must note that the laboratory work of processing more than 800 samples would not be possible without the work of laboratory technicians Garazi Martin and Louisa Pless. During these internships, I had the opportunity to discuss technological advances to study host-microbiota interactions with PhD students (now both doctors) Lasse Nyholm and Adam Koziol, as well as other colleagues, with whom I ended up publishing a perspective article that can be found in Annex 1.

Once chicken ileal and caecal metagenomic samples were sequenced, data were handed over EMBL-EBI's Microbiome Informatics team, who assembled the metagenomic data and reconstructed metagenome-assembled microbial genomes (MAGs). This task was performed when the Covid-19 pandemic hit, which forced us to stay home. I took advantage of this pause to work on bioinformatic methods to recover host genomic data from metagenomic samples, in collaboration with Dr. Melanie Parejo. During 2020 we compared genotyping and imputation methods which yielded the publication of the research article that constitutes Annex 2.

Covid-19 also hindered the sequencing and *de novo* genome assembly of the two broiler breeds used in HoloFood by Vertebrate Genome Laboratory, The Rockefeller University (EEUU). This delay required a change of management and directionality of the research, which is why it ended up constituting Annex 3 of this thesis.

Simultaneously, Dr. Nuria Tous (IRTA), Joan Tarradas (IRTA), and myself analysed pen-level and animal-level chicken performance and welfare parameters under the supervision of Antton Alberdi. We observed that the impact of dietary treatments on chicken performance was negligible. However, we identified a high interindividual variation in chicken growth for the first two trials, and an opportunistic colonisation of *Campylobacter spp.* associated with a reduced chicken growth in the third trial. These observations sparked long discussions on how holo-omic data layers could be used for explaining the observed trends, aiming to understand the interplay between hosts and associated gut microbial communities. These discussions resulted in a research article submitted for publication in late 2021, which embodies Chapter 3 of this thesis.

When EBI-EMBL partners released the MAG catalogues, researchers from CEH and EHU began the coordinated task of analysing the data. Dr. Iñaki Odriozola, Dr. Jorge Langa, Dr. Ostaizka Aizpurua, and I started discussing the optimal ways of analysing shotgun sequenced metagenomic data, under the supervision of Antton Alberdi. During this period, I learnt unsupervised and supervised statistical analyses for metagenomics and (meta)transcriptomics under the supervision of Iñaki Odriozola and Jorge Langa. While the quality of the data collected from cecum samples was very good, the data quality from ileum samples was proven insufficient to address our scientific inquiries. For this reason, we decided to focus on the caecum only.

The first two experimental replicates enabled us to characterise the functional dynamics of the caecal microbiota development during chicken growth in 2022. This work yielded the research article that embodies Chapter 4.

The third experimental replicate enabled us to examine the microbial dynamics and host response to *Campylobacter* colonisation. The analysis of these data allowed me to learn about zoonotic agents and the implications of *Campylobacte*r in chickens' performance and welfare in 2023. The result was the research publication that is included in Chapter 5.

Thesis hypothesis

The main hypothesis of this thesis is that:

> A holo-omic framework is an effective approach for unveiling host-microbiota interactions with which to explain the variation in chicken performance caused by the feed-microbiome-host axis

This general hypothesis can be further elucidated through the following hypotheses:

1. Specific biological hypotheses
   - 1.1. Chicken growth performance is affected, or even conditioned, by the interplay between the host and its associated gut microbiota
   - 1.2. Dietary additives modulate chicken-microbiota interplay, and thus improve chicken performance and welfare
   - 1.3. The efficacy of dietary additives varies depending on the genomic background of the chicken and its associated gut microbiota

2. Specific technological hypotheses
   - 2.1. Holo-omics enable to screen the biochemical landscape of intestinal host-microbe interactions
   - 2.2. Genome-resolved metagenomics offers direct insights of the entire functional characteristics and functional dynamics of the chicken gut microbiome
   - 2.3. Genome-resolved metatranscriptomics provides information on the gene expression of every bacterial strain identified at each sampling time
   - 2.4. Chicken intestinal transcriptomics enables monitoring host local responses to the developing microbiota in the gastrointestinal tract
   - 2.5. The integration of multiple omic layers allows unveiling biological patterns hidden to single omic data

Thesis aims

The main objective of this thesis is to:

> Provide novel insights into host-microbiota dynamics in broiler chickens subjected to different dietary treatments in an intensive production context by applying holo-omic methodologies on intestinal samples

The specific objectives of each chapter are listed below:

1) Identify knowledge gaps in standard and targeted methodologies used in the animal production to assess chicken performance and welfare
    - Describe the experimental work conducted in HoloFood
    - Asses dietary treatments effect on chicken performance
    - Present main pen-level performance, and individual-level inflammation and stress biomarkers results of broiler chickens
    - Propose multi-omic approaches to cover host-microbiota interactions that could potentially explain described biological trends

2) Model the functional development of the chicken gut microbiota and test whether it is associated with chicken growth performance
    - Describe the functional landscape of the entire caecal microbial catalogue
    - Identify the main sources of variation for microbiota composition and function
    - Explore the functional dynamics of the microbiome across productive lifespan of the chicken
    - Correlate functional capabilities with functional activity of the microbiota
    - Associate body weight with the microbiome

3) Unravel the potential role of microbe-microbe and host-microbe interactions in *Campylobacter spp.* colonisation
    - Confirm *Campylobacter* presence using genome-resolved metagenomics
    - Compare microbial compositional and functional progression of chickens that did and did not undergo an opportunistic *Campylobacter* colonisation
    - Search for metabolites produced by the microbiome that could facilitate *campylobacter* colonisation
    - Explore whether *Campylobacter* triggered an immune response in the host
    - Establish a connection between diminished chicken growth performance and either the host immune response or alterations in the gut microbiota

Thesis outline

This thesis consists of six chapters comprising a comprehensive introduction, a second chapter on the thesis roadmap, the proposed hypotheses and objectives together with an outline of the dissertation, followed by three original research papers that culminate in a final discussion. The thesis also includes an annex section, with three additional research articles. The contents of the three original articles are detailed below.

In Chapter 3, named *Novel strategies to improve chicken performance and welfare by unveiling host-microbiota interactions through hologenomics*, we draft an article halfway between a research article and a perspective. We describe the study design, the experimental work, the sampling procedures carried out, as well as the samples available at HoloFood. We present pen-level and individual-level performance and welfare parameters. Chicken body weight variance is hierarchically decomposed to understand the main sources of variation. We hypothesise that in-depth knowledge based on hologenomics can help us better understand animal performance and welfare.

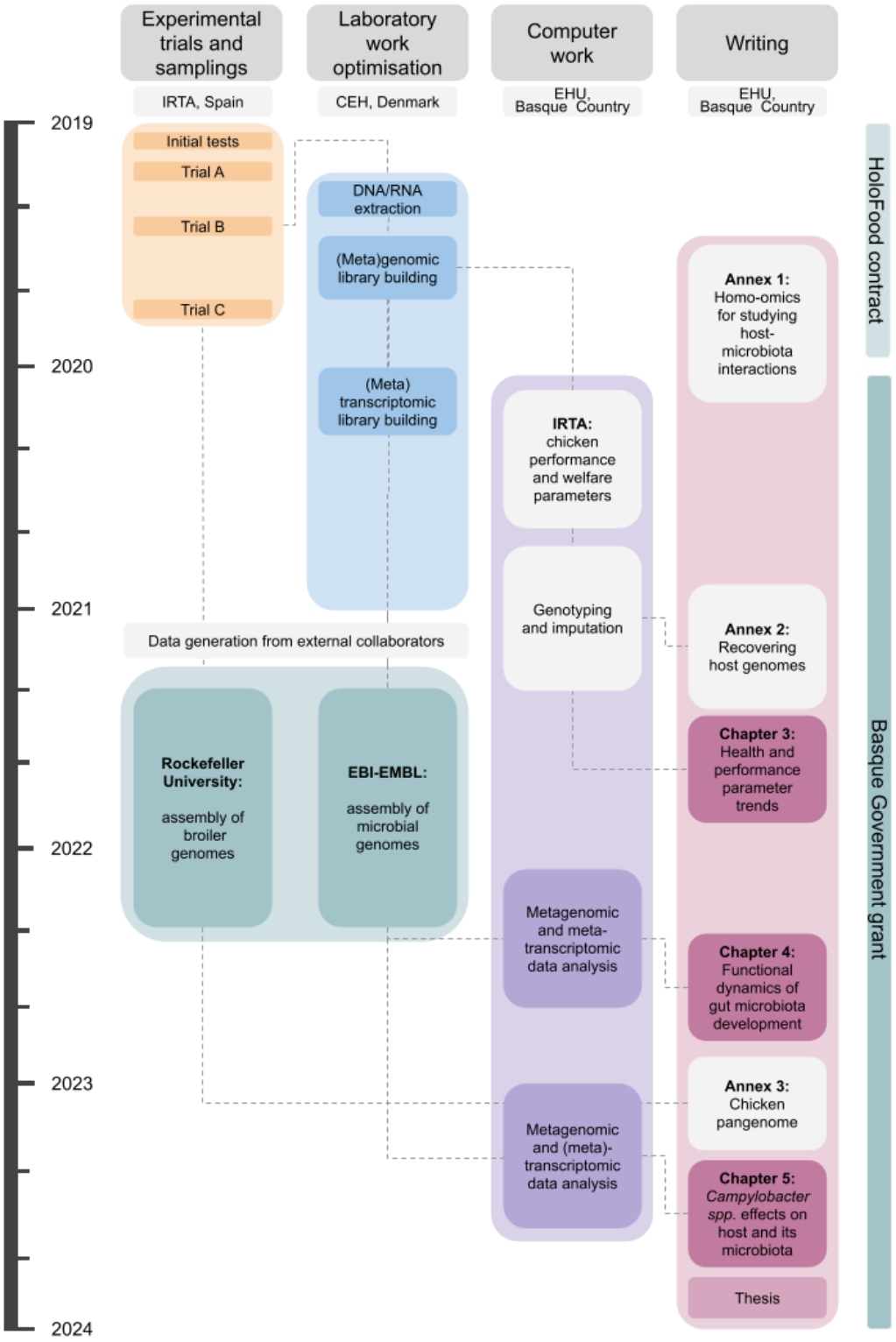
Chapter 4, titled *Reduced metabolic capacity of the gut microbiota is associated with host growth in broiler chickens*, characterises the compositional and functional features of the gut microbiota of chickens during their productive lifespan. We use various metrics of alpha and beta diversities, and hierarchical linear models to examine microbial dynamics over time, categorising microbes as increasers and decreasers based on temporal trends. We then associate these two microbial groups with chicken body weight. We show that, although all metrics of microbiome diversity increase with chicken age, the overall capacity and activity of the microbiome to perform metabolic functions decreases. This pattern is due to the spread of strains with low metabolic capacities as chickens grow older. A more pronounced transition from high-capacity to low-capacity bacteria is positively associated with chicken growth, suggesting the potential to improve animal production by modulating the microbiome towards communities dominated by low-capacity bacteria.

For Chapter 5, designated *Priority effects and microbial cross-feeding shape zoonotic agent spread in broiler chickens*, we study how microbial communities contribute to *Campylobacter spp.* colonisation, and how the distinct microbiota developed after *Campylobacter* colonisation affect chicken body weight. We define two microbial enterotypes with the Dirichlet multinomial mixture model and identify microbial dependencies with genome-scale metabolic networks. We show that the unusual relative abundance of *Bacteroides fragilis_A* can facilitate *Campylobacter* colonisation by providing essential metabolites that are key in central metabolic processes. By performing differential gene expression analyses, we also observe that the reduced growth performance of chickens in trial C is due to an increased microbial activity that might alter nutrient availability to the host, rather than due to a host immune response.

The concluding discussion provides a comprehensive summary of the thesis. It focuses on highlighting the achieved goals but also explores the many challenges and limitations encountered in the process. The chapter ends with a concise reflection on the future of holo-omic studies and their potential for microbe-based solutions.

Thesis timeline

This timeline aims to illustrate where and when the different tasks were carried out. The coloured boxes indicate tasks from my thesis, while the grey boxes indicate tasks on which I collaborated. Boxes in light green indicate data generation processes performed by external collaborators.



Novel strategies to improve chicken performance and welfare by unveiling host-microbiota interactions through hologenomics

Chapter 3

Abstract

Fast optimisation of farming practices is essential to meet environmental sustainability challenges. Hologenomics, the joint study of the genomic features of animals and the microbial communities associated with them, opens new avenues to obtain in-depth knowledge on how host-microbiota interactions affect animal performance and welfare, and in doing so, improve the quality and sustainability of animal production. Here, we introduce the animal trials conducted with broiler chickens in the H2020 project HoloFood, and our strategy to implement hologenomic analyses in light of the initial results, which despite yielding negligible effects of tested feed additives, provide relevant information to understand how host genomic features, microbiota development dynamics and host-microbiota interactions shape animal welfare and performance. We report the most relevant results, propose hypotheses to explain the observed patterns, and outline how these questions will be addressed through the generation and analysis of animal-microbiota multi-omic data during the HoloFood project.

Publication

Núria Tous^[1], Sofia Marcos^[2], Farshad Goodrazi Boroojeni[3], Ana Pérez de Rozas[4], Jürgen Zentek[3], Andone Estonba[2], Dorthe Sandvang[5], MTP Gilbert[6,7], Enric Esteve-Garcia[1], Robert Finn[8], Antton Alberdi[6], Joan Tarradas[1]

[1]Animal Nutrition, Institut de Recerca i Tecnologia Agroalimentàries (IRTA), Centre Mas Bové, Constantí, Spain
[2]Applied Genomics and Bioinformatics, University of the Basque Country (UPV/EHU), Leioa, Bilbao, Spain
[3]Institute of Animal Nutrition, Department of Veterinary Medicine, Freie Universität Berlin (FUB), Berlin, Germany
[4]Centre de Recerca en Sanitat Animal (CReSA), Institut de Recerca i Tecnologia Agroalimentàries (IRTA), Campus de la Universitat Autònoma de Barcelona, Spain
[5]Chr. Hansen A/S, Animal Health Innovation, Hoersholm, Denmark
[6]Center for Evolutionary Hologenomics, The GLOBE Institute, University of Copenhagen, Copenhagen, Denmark
[7]University Museum, Norwegian University of Science and Technology (NTNU), Trondheim, Norway
[8]European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Hinxton, UK

Publication status: Published
Link to publication: ttps://www.frontiersin.org/articles/10.3389/fphys.2022.884925/full

Introduction

With the ever-increasing human population on Earth, humanity is facing several major challenges to ensure long-term balance between natural resource use and environmental conservation (Reid et al., 2010). Many of the current agricultural practices are not sustainable due to excessive carbon emission, resource consumption, and waste production (Agovino et al., 2019). Hence, there is an urgent need to transition into a more resilient and sustainable agriculture model, in which the efficiency of production is improved, the use of antibiotics is reduced, and the welfare of animals is ensured (Eyhorn et al., 2019).

A fundamental step to optimise farming practices is to obtain in-depth understanding on the biological functioning of the animal production systems (Messerli et al., 2019). These systems often include biological elements beyond the actual animal that is being produced, among which host-associated microorganisms stand out due to their relevance for the optimal biological functioning of most animals (McFall-Ngai et al., 2013). Intestinal microorganisms not only modulate nutrient intake (Diaz Carrasco et al., 2019), but also shape intestinal immune and inflammatory processes (Zhou et al., 2021), intervene on host systemic growth parameters (Fraune & Bosch, 2010), and even influence host behaviour (Johnson & Foster, 2018). Microorganisms colonise the animal gut as soon as it is exposed to the environment (Sprockett et al., 2018), and develop communities with complex spatial and temporal dynamics (Debray et al., 2021), which continuously interact with the host animal (Ansari et al., 2020; Khan et al., 2019).

Animal-microbiota interactions have so far remained largely unexplored because of the limited capacity of scientists to properly characterise and analyse the key elements partaking in this interplay due to their excessive complexity (Alberdi et al., 2021). However, the development of high-throughput DNA sequencing and mass spectrometry technologies, linked to higher computing capacity and development of powerful bioinformatic tools, is changing this scenario (Graw et al., 2021). Today we are not only able to characterise the entire genetic information of animals and their associated microorganisms (namely the hologenome), but we can also quantify how genes are expressed, which proteins are synthesised and what metabolites result from enzymatic reactions happening in the gut (Nyholm et al., 2020). Such a holo-omic approach that considers multiple omic layers of both animals and associated microorganisms, is starting to unveil biological features and patterns that have remained hidden so far (Alberdi et al., 2021).

This technological revolution is likely to change animal science practices, because many sources of variability that have been so far attributed to background noise, such as host microgenetic and microbiota variation among individuals, can be surfaced and included in the analyses (Alberdi et al., 2021). This requires an increased attention on the biological processes happening in each individual animal, rather than considering animals just as units that contribute to pen or tank statistic averages. The first attempts to implement individual-based multi-omic strategies in farm animals have provided detailed understanding of feed-microbiota-animal interactions (Andersen et al., 2021), by for example demonstrating which bacteria with which genes are able to degrade which carbohydrates

(Michalak et al., 2020). We anticipate that such a mechanistic understanding of biological processes will contribute to generating refined hypotheses, predictive models and experimental treatments that will lead to a reduction of animals employed for research and an ultimate development of more optimal farming strategies.

HoloFood (HoloFood Consortium, 2019) is a multi-partner H2020 project that is pioneering such an approach, among other farming systems, on broiler chickens. The project aims at developing and implementing joint multi-omic analyses of animals and their associated microorganisms to generate in-depth knowledge of animal-microbiota interactions, and in doing so, improve the quantity, quality and safety of the produced food, the sustainability of the production process and the welfare of the produced animals. To set the baseline to an upcoming series of publications that implement such multi-omic analyses, here we introduce the strategic vision, and experimental work conducted to generate biological samples and associated performance results of broiler chickens in the project HoloFood. We present the main performance results, discuss their relevance, and relate them to future multi-omic analyses that HoloFood partners will conduct to address the variety of biological questions raised from the initial screening of animal performance.

Results

The study was designed in a way that would maximise variability, to surface interactions between animal and microbial features. The design consisted of three identical experiments (A, B and C) with a randomised pen design in which broilers from two genetic lines (Ross308® and Cobb500®) and both sexes were grown under three dietary treatments: i) basal diet (BD), ii) BD plus a probiotic additive (PR), and iii) BD plus a phytobiotic additive (PH) (full details in Methods section, overview in Fig. 1). The BD had a high content of wheat (rich in non-starch polysaccharides - NSPs) and was not supplemented with enzymes (Table S1), aiming to induce a mild inflammatory process in the intestine (Raza et al., 2019), and thus maximise the beneficial effects of feed additives (Bortoluzzi et al., 2019; Whelan et al., 2019). The probiotic contained a blend of three *Bacillus* strains and the phytobiotic was based on polyphenols (78%) and procyanidins (22%) from white grapes. Both additives are associated with promotion of a healthy gut microbiota, improved nutrient digestion and absorption, enhanced intestinal morphology and immunomodulation (Chamorro et al., 2019; Tarradas et al., 2020).

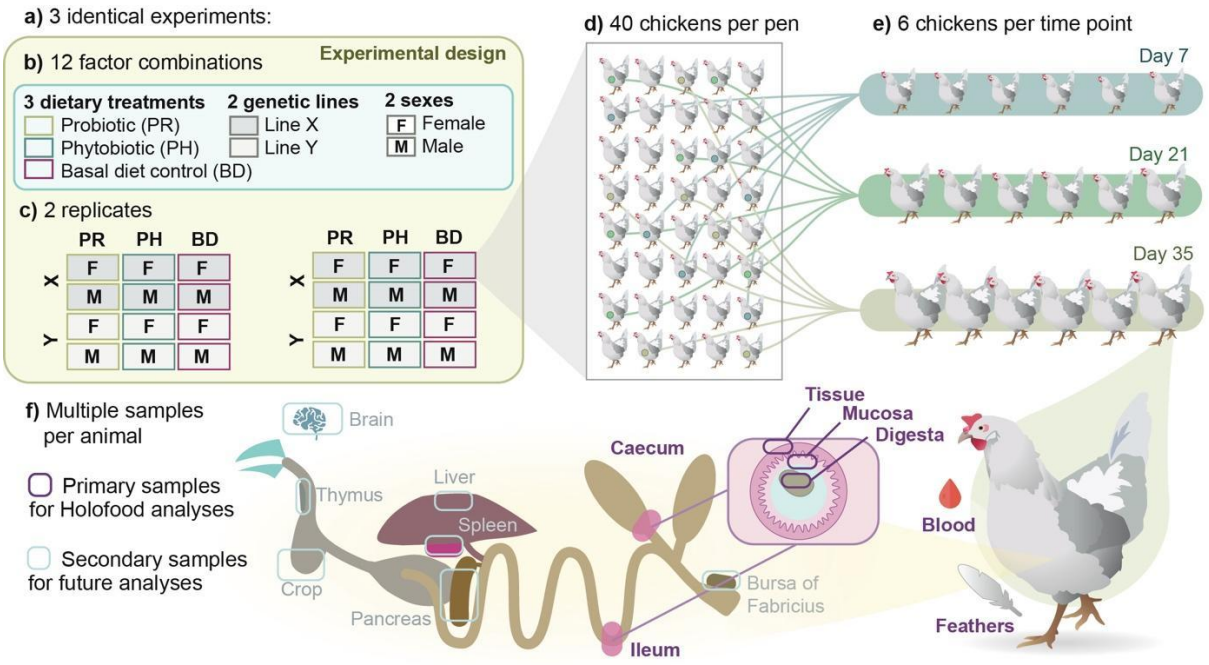


Figure 1. Experimental, sampling and analytical design of the HoloFood study focused on broiler chickens. a) Three experiments with identical study design were conducted in 2019. **b)** Each experiment contained 12 factor combinations of three dietary treatments, two genetic lines and two sexes. **c)** Each combination was replicated twice per experiment, yielding 24 pens per experiment and 72 pens in total. **d)** Each pen contained 40 chickens in the beginning of the experiment, 18 experimental animals that were randomly selected and tagged the first day of the experiment, and 22 more animals that provided commercial-like density conditions. **e)** Six chickens were euthanised at day 7, six more at day 21 and six more at day 35 for collecting samples for performance, multi-omic and complementary analyses. **f)** From each animal, 14 tissue and digesta samples were collected for a variety of analyses (see Table 1 for details), and complementary organ samples were also collected for future analyses.

Growth performance was assessed at pen level at three time points (7, 21 and 35 days) in terms of average body weight (BW), daily gain (ADG), daily feed intake (ADFI), feed conversion ratio (FCR), and European production efficiency factor (EPEF). Moreover, a total of 1296 animals (from the 72 pens; 24 pens per experiment) were individually weighted (iBW) and euthanised (6 animals per pen and day). A total of 14 samples were collected from each animal to measure individual key performance indicators (KPIs), as well as to generate multi-omic and complementary analyses (Table 1). Measured KPIs aimed at quantifying quantity, quality and safety of the produced food, the sustainability of the production process and the welfare of the produced animals.

Table 1. Overview of the biological samples collected from each individual animal within the project HoloFood, with their preservation strategy and corresponding multi-omic and complementary analyses.

Sample	Analyses
Blood	Lipopolysaccharide and acute phase proteins
Ileum tissue	Mucus production Histology Targeted amplification of inflammatory markers Shotgun chicken transcriptomics
Ileum mucosa	Shotgun chicken transcriptomics 16S amplicon sequencing
Ileum content	Shotgun metagenomics Shotgun metatranscriptomics 16S amplicon sequencing Chicken genomics Metabolomics
Caecum tissue	Mucus production Histology Targeted amplification of inflammatory markers Shotgun chicken transcriptomics
Caecum mucosa	Shotgun chicken transcriptomics 16S amplicon sequencing
Caecum content	Shotgun metagenomics Shotgun metatranscriptomics 16S amplicon sequencing Chicken genomics Metabolomics
Feathers	Corticosterone measurement

The data on each individual animal derived from the experiments were analysed at three different levels. The first level included the analyses of measurements obtained in the farm (e.g., iBW, footpad dermatitis). The second level comprised analyses conducted *a posteriori* from samples obtained during the trials using techniques regularly applied in animal sciences (e.g., ELISA for protein quantification, qPCR for pathogen detection). The third level included multi-omic analyses, which characterises the animal-microbiota system at the highest level of breadth and resolution. These analyses include whole-genome sequencing of chicken genomes, deep metagenomics of microbial communities, (meta)transcriptomics and metabolomics.

In the following, we present the main results obtained per pen, and the first two levels of the individual analyses sorted by topic. We explain their biological relevance based on current knowledge, and we project the potential of multi-omic analyses that will be conducted on top of these results to address the most relevant pending questions.

Effect of dietary treatments

No significant effect of administered dietary additives was observed in the performance of the animals (Fig. 2a; Tables S2-12). Regarding the probiotic, *Bacillus* spp. are commonly used as additives in broiler production (Irta, 2015), and the specific probiotic strains tested in our study have been previously shown to improve performance and physiological traits in broilers

(Goodarzi Boroojeni et al., 2018; Molnár et al., 2011). However, contrasting observations that align with our results have also been reported (C.-L. Li et al., 2018), which could be explained by varying experimental conditions, such as the specific *Bacillus* strain employed, the dose, or the basal diet. The supplementation of poultry diets with multi-strain *Bacillus* probiotic products is in general associated with competitive exclusion of common pathogens, improved nutrient digestion and absorption through the production of exogenous enzymes, enhanced intestinal morphology, and the modulation of relevant immune system pathways (Ramlucken et al., 2020; Tarradas et al., 2020). Regarding the phytobiotic, multiple modes of action have been attributed to additives containing polyphenols and procyanidins, including antioxidant and anti-inflammatory properties, promotion of beneficial bacteria in the gut, and enhancement of nutrient absorption through binding of dietary proteins and carbohydrates (Chamorro et al., 2019; Hasted et al., 2021).

As the beneficial effect of feed additives is usually not evident when animals grow under optimal conditions (Vilà et al., 2010), we deliberately induced a challenging condition through increasing the amount of dietary soluble NSPs. These compounds are known to have deleterious effects on the bird's health and performance through increasing intestinal viscosity and hampering nutrient digestibility (Raza et al., 2019), and can thus maximise the beneficial effects of feed additives (Bortoluzzi et al., 2019; Whelan et al., 2019). Accordingly, the overall performance was 13.8% lower than the expected from reference performance tables (Aviagen, 2019; Cobb-Vantress, 2018), yet with no differences between treatments. This could be in part explained because the experiments were designed to maximise variability.

Dietary treatments did not induce any significant change in the acute phase protein values measured in plasma (Figure 2F, Table S7). The only parameter affected by dietary treatments was corticosterone (COR) (Table S6). COR measured in feathers is used as a biomarker of accumulative stress (Bortolotti et al., 2008), as chronic levels of COR are associated with detrimental effects on growth and related biological traits (Scanes, 2016). COR levels increased with the PH diet at day 7, but these levels were not prolonged over time nor decreased compared to the BD. This observation contrasts with previous studies that reported a reduction of serum and feather COR in broilers fed with polyphenol extracts (Gong et al., 2018; Gopi et al., 2020).

To delve into the reasons under, among other questions, the lack of positive effect of the feed additives on commonly assessed nutritional parameters, HoloFood will generate whole genome sequences of probiotic strains used through hybrid short- and long-read DNA sequencing (Wick et al., 2017), and perform a pangenome analysis with other sequenced and annotated *Bacillus* strains to identify bacterial genes that could confer beneficial functional capabilities to the strains. HoloFood will also generate deep genome-resolved metagenomic datasets (Almeida et al., 2019; Pasolli et al., 2019) from the ileal and caecal content. While most chicken-associated microbiota research is being conducted using targeted sequencing (Jurburg et al., 2019; Mohd Shaufi et al., 2015; Ocejo et al., 2019), the first shotgun-sequencing based studies have recently been published (Gilroy et al., 2021; Glendinning et al., 2020). The metagenome-assembled genome (MAG) catalogue of bacteria

associated with broiler chickens generated in HoloFood, will not only complement such efforts for the high-resolution characterisation of chicken-associated microbiomes, but will also enable in-depth study of strain-level microbiota dynamics in the analysed production context through combining taxonomic and direct functional inferences. The reconstructed bacterial genomes will be functionally annotated, thus directly inferring the metabolic capacities (e.g., complex polysaccharide degradation, short-chain fatty acid (SCFA) production) of strains, and acknowledging the aggregated functional landscape of the entire community present in each animal (Shaffer et al., 2020). HoloFood will also generate whole genome sequences of the probiotic strains through hybrid short- and long-read DNA sequencing to ensure highest-quality genome reconstructions of the tested probiotics (Wick et al., 2017), and perform a pangenome analysis with other sequenced and annotated Bacillus strains to identify bacterial genes that could confer beneficial functional capabilities to each of the strains. This will enable us to understand the specific means of action through which each strain can interact with the microbiota and various intestinal features of the host. All these analyses will enable ascertaining the relative abundances of Bacillus probiotics in different intestinal segments, and measuring whether the additives trigger broad-as well as fine-scale taxonomic and/or functional changes in the microbiota of broilers, that could contribute to explain the observed results.

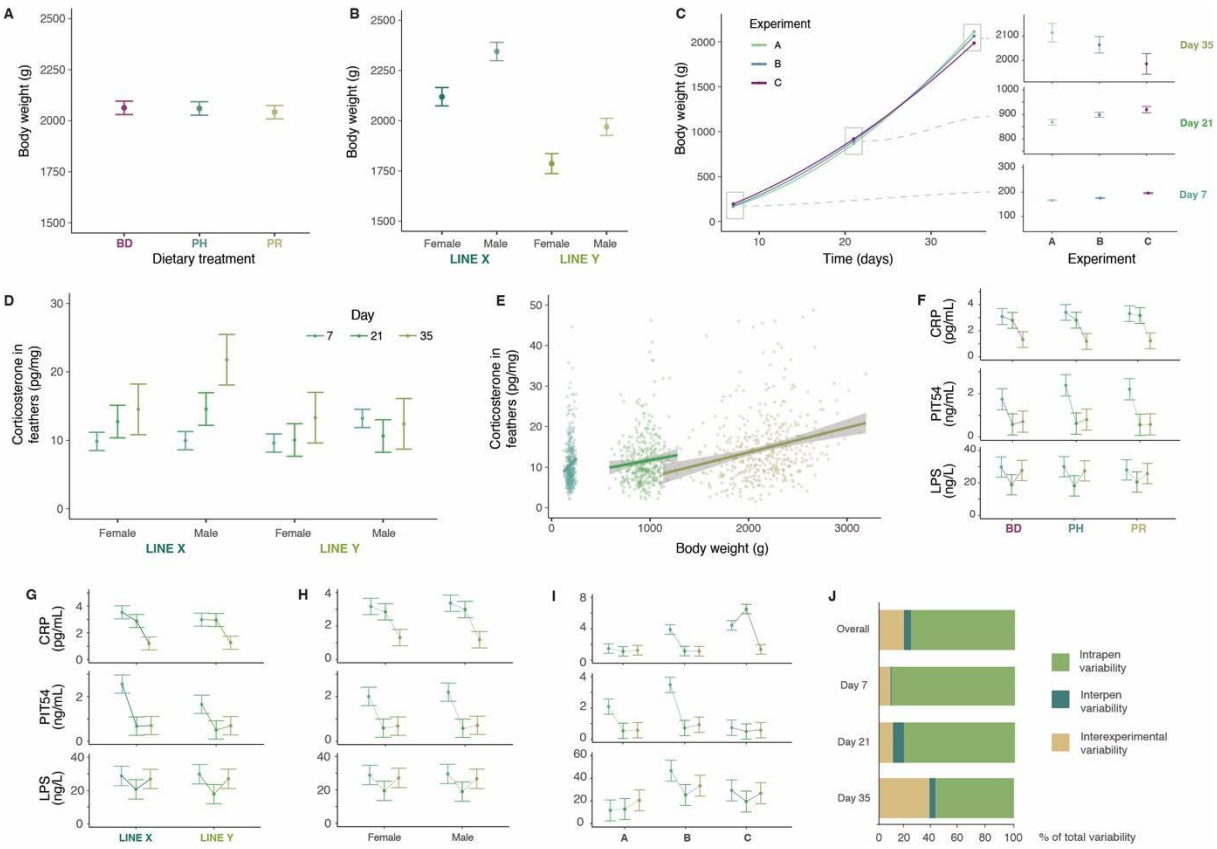


Figure 2. Overview of main results. **a)** Body weight differences across dietary treatments, namely basal diet (BD), BD plus probiotic (PR) and BD plus phytobiotic (PH), at day 35. **b)** Body weight differences across lines and sexes at day 35. **c)** Body weight progression of the

three experiments, with detailed overview of days 7, 21 and 35. **d)** Corticosterone (COR) levels measured in feathers at different days, sexes, and genetic lines. **e)** Linear correlation between COR levels and body weight at the three time points. **f)** C-reactive protein (CRP), avian haptoglobin-like protein (PIT54) and lipopolysaccharide (LPS) levels in plasma across time points in different dietary treatments. **g)** CRP, PIT54 and LPS levels in plasma across time points in different genetic lines. **h)** CRP, PIT54 and LPS levels in plasma across time points in different sexes. **i)** CRP, PIT54 and LPS levels in plasma across time points in different experiments. **j)** Hierarchical decomposition of observed body weight variation within pens, among pens and among experiments.

Effect of broiler line

Although theoretically both broiler lines tend to perform similarly in terms of growth and final BW (Aviagen, 2019; Cobb-Vantress, 2018; Livingston et al., 2020), broiler line X reached significantly higher BW, ADG, and ADFI and lower FCR than line Y at day 35 (2232 g and 1878 g for BW (SE=18.8), 62.6 g and 52.4 g for ADG (SE=0.54), 96.6 g and 81.1 g for ADFI (SE=0.80), and 1.497 and 1.542 for FCR (SE=0.015), respectively; Fig. 2b and Table S2). The differences observed on performance between lines could partially be explained by the higher initial BW of line X (44.5 g) at day 0 than line Y (40.8 g) (SE=0.31; Table S2). A retrospective analysis of the breeders' features showed that the age of breeders was higher for line X (49.6±10.1 weeks) than for line Y (44.1±11.4 weeks), which probably caused the observed difference between the initial BW (Iqbal et al., 2017). However, the differences between lines at day 35 did not disappear even when BW at day 0 was included as a covariate in the statistical model, suggesting that other factors contributed to shape the differences observed between both lines. Therefore, performance results suggest that line X could exhibit a higher resistance to NSPs or a more active feeding behaviour than line Y.

The accumulation of COR in feathers was significantly higher for line X (13.91 pg/mg) than for line Y (11.54 pg/mg) (standard error (SE)=0.914; p<0.001) (Fig. 2d, Table S6), in contrast to what would be expected from a stress indicator (Carbajal et al., 2014). Individual COR levels were also higher in animals with higher iBW (Fig. 2e), which could be explained by an increased growth rate promoting deposition of COR in feathers (Jimeno et al., 2018).

PIT54 levels in plasma were also higher in line X at days 7 (p=0.0687) and 21 (p=0.001) than in line Y (+35.5% and +25.4%, respectively) (Figure 2G, Table S7b) and for the whole experiment (p=0.0354; +27.5%). Chicken haptoglobin-like protein (PIT54) is an acute phase protein with an important inhibitory role in inflammation processes (Wicher and Fries, 2006), which is rapidly increased in the blood as a response to infectious agents or physiological stressors (O'reilly and Eckersall, 2014). PIT54 in the chicken plasma binds free haemoglobin to inhibit haemoglobin-mediated oxidation of lipid and protein (Ahn et al., 2019). Antinutritional effects of NSP are related to a reduction of BW and FCR, and can trigger a mild chronic inflammation in the gut (Cardoso Dal Pont et al., 2020). Host gut inflammatory response produces reactive oxygen species (ROS), which cause oxidative stress and have the potential to damage host tissue (Costantini & Møller, 2009). The increased antioxidant

activity through the high levels of PIT54 in line X could explain the better performance compared with line Y through the amelioration of antinutritional effects of NSP.

The observed differences between broiler lines point to differences in systemic responses to the pro-inflammatory diet, which probably yielded the BW and associated KPI differences between the two genetic lines. The whole-genome analyses we will conduct in HoloFood will enable deepening into these observations through unveiling genome-wide differences between both lines. We will perform whole genome resequencing to generate single nucleotide polymorphism (SNP) profiles of all individuals (R. Li et al., 2009). This will enable testing whether both lines have genetic differences in key genes related to inflammatory responses induced by a high dietary concentration of NSPs, as well as key metabolic pathways such as steroid hormone biosynthesis (Kanehisa et al., 2021). In addition, we will also generate *de-novo* reference genomes (Baker, 2012) and Hi-C maps (Lieberman-Aiden et al., 2009) of both lines, to explore the effects of structural genome variants (e.g., copy number variations (CNVs), translocations, inversions) in the performance differences observed between both lines. Such chicken genomic data will be coupled with the aforementioned microbial metagenomic information, which will allow exploring whether and how host-microbiota interactions orchestrate different physiological responses to nutritional stress in the two broiler lines, as previously reported in other taxa (Ma et al., 2019).

Effect of biological sex

Male chickens exhibited larger average BW than females at all time points, reaching a 8.7% larger BW at day 35 (2157 and 1953 g, respectively (SE=13.8); Fig. 2b). The rest of performance traits were also improved in males compared to females (ADG, ADFI, FCR, and EPEF) (Tables S2). These differences are expected (Aviagen, 2019; Cobb-Vantress, 2018), as growth patterns of male chickens outperform that of females (Aggrey, 2002; Hausman et al., 2014; Livingston et al., 2020). However, unlike in previous studies (Carbajal et al., 2014), males showed higher COR levels than females in the period 0-35 days (13.8 vs 11.7 pg/mg; $p<0.001$; SE = 0.91;) and also at day 7 and at day 35 ($p<0.001$ and $p=0.028$, respectively) (Fig. 2d, Table S6), supporting again that COR cannot directly be negatively associated with welfare and performance of animals (Jimeno et al., 2018).

In the last period of the experiment (at day 35), the levels of C-Reactive protein (CRP) in plasma were higher in females than males (1.31 vs 1.18 ng/mg; SE=0.08; $p=0.045$) (Fig. 2h, Table S7b). CRP is an acute phase protein used as a highly sensitive marker of inflammation and tissue damage (O'reilly & Eckersall, 2014). Some immune salivary markers including CRP are influenced by biological sex in other production animals such as pigs (Gutiérrez et al., 2018). However, the reasons behind biological sex differences for inflammatory markers are still unknown.

In HoloFood we will aim at further understanding these differences in growth between both sexes, as well as other questions mentioned above, by generating whole-genome transcriptomic data (RNAseq) to identify gene expression differences in the intestinal tissues. While intestinal expression of targeted genes is routinely measured in animal sciences

(Farahat et al., 2021; Slawinska et al., 2019), how genome-wide gene expression varies across sexes and intestinal sections is still largely unknown. As intestinal gene expression is known to be modulated by the microbiota (Volf et al., 2017), and biological sex also contributes to shaping microbial communities (Lee et al., 2017), we will aim at detecting associations between host expression patterns and microbial communities to ascertain host-microbiota interactions related to sex differences. For instance, we will analyse expression of host genes involved in cholesterol (precursor of steroids, and thus related to COR) absorption, such as NPC1L1 and ABCG5/ABCG8, which have been shown to be modulated by the microbiota in rodents (Zhong et al., 2015). Sample collection in HoloFood was extended to organs beyond the gastrointestinal tract, including liver and brain, which are involved in appetite regulation and other processes related to the gut-brain axis (Cryan et al., 2019). The analysis of gene expression in such organs along with gut processes, will enable delving into the relationship between host genetics and microbiota factors with nutrient metabolism in shaping feeding behaviour and related performance differences in broiler chickens.

Effect of age and development

Animal development was linked to multiple changes in analysed metrics. For instance, COR accumulation in feathers increased as animals grew (Table S6), mirroring previous observations (Nordquist et al., 2020) (Fig. 2e). The levels of PIT54 and CRP in plasma peaked at day 7, and decreased through time (Table S7b), exhibiting a trend that could be linked to vaccination as well as to microbiota development (Figs. 2f-g). On the one hand, vaccines are known to increase concentration of acute phase proteins and stress markers during the first days after administration (Kaab et al., 2018), and all chickens in the experiment were vaccinated against Avian Infectious Bronchitis and Gumboro diseases at birth. On the other hand, early microbial colonisation of the intestine is also known to boost the development of the immune system (Broom & Kogut, 2018) through, for example, the production of the intestinal mucus layer (Duangnumsawang et al., 2021), which provides the first protective shield preventing a direct access of pathogenic bacteria to the epithelial surface.

HoloFood will also generate and analyse microbiota-wide gene expression in the chicken intestine, as well as metabolites that play essential roles in the host-microbiota interplay, such as SCFAs (van der Hee & Wells, 2021). Our study design, which entails euthanising animals at each sampling point, prioritises spatial resolution of intestinal sections over temporal development, which complicates tracking the temporal development of individual animals. However, sampling at three different time points will provide an overview of how much microbiota development varies across individuals (Ballou et al., 2016; Debray et al., 2021; Jurburg et al., 2019; Sprockett et al., 2018). Shotgun metatranscriptomic data will complement the metagenomic information, thus providing not only an overview of the relative abundances of different bacterial taxa in different time points, but also displaying the gene expression patterns of the bacteria. In addition, metabolomic data will enable validating whether the activated metabolic pathways are translated into different levels of SCFA concentrations. We will measure how the expression of microbial genes involved in

carbohydrate metabolism and SCFA production vary across development, and how these changes are associated with animal growth and changes observed in acute phase protein levels.

Effect of zoonotic pathogens

Targeted detection of three common zoonotic pathogens, namely *Salmonella*, *Clostridium* and *Campylobacter* spp. was performed to detect whether natural colonisation of the chicken intestines occurred during the trials. Only one animal in trial A and another in trial B were positive for Clostridium, being all the animals negative for Salmonella and Campylobacter. However, in trial C, the analyses revealed that 13.7% of the sampled animals at day 7 were positive for *Salmonella*, but the colonisation vanished as the animals grew. Moreover, 99% of the birds sampled from day 21 onwards in trial C presented *Campylobacter* colonisation, with no prevalence differences between dietary treatments. *Campylobacter* and *Salmonella* are usually considered mere commensals in poultry, but they cause the highest numbers of foodborne diseases in humans globally (European Food Safety Authority & European Centre for Disease Prevention and Control, 2021). Recent studies have nevertheless shown that *Campylobacter* colonisation in chickens can cause gut microbiota alterations and intestinal damage that occasionally facilitates bacterial colonisation of extraintestinal organs, which may eventually lead to a reduced animal performance and welfare (Awad et al., 2015, 2018). In accordance with these observations, the detection of *Campylobacter* was correlated with a drop in BW of the animals from day 21 onwards compared to the two previous trials (Fig. 2c). In addition, a peak of CRP values in plasma was detected in *Campylobacter* colonised animals at day 21 (Fig. 2i), most probably indicating a systemic reaction to colonisation (Liu et al., 2019; Zhang et al., 2020). Other opportunistic pathogens from the Campylobacterales order, such as *Helicobacter brantae,* which may be present along with *Campylobacter* but undetected with targeted approaches, might likewise be involved in performance drop (Kollarcikova et al., 2019).

The bacterial genome catalogue we will build in HoloFood will not only enable us to ascertain whether the *Campylobacter* colonisation was due to a single or multiple strains (Chaloner et al., 2014), but also to characterise the entire catalogues of genes of these strains and thus identify potential virulence factors that could have triggered the inflammatory response. We will also be able to ascertain whether the *Campylobacter* colonisation triggers any systemic change in the microbiota and in the intestinal response of the animals, through combining gene expression data of chickens and microorganisms as well as metabolomic information. *Campylobacter* induces the expression of various host pro-inflammatory cytokines through the activation of Toll-like receptor 4 (TLR4) and TLR21 signalling pathways (de Zoete et al., 2010). Our analyses will enable measuring changes in the expression levels of genes involved in these signalling pathways between Campylobacter positive and negative animals, to deepen into the effect of Campylobacter in chicken welfare and performance.

Hierarchical variance decomposition of chicken body weight

Due to our interest in generating systemic characterisation of individual chickens, we explored how the variance of iBW data for each combination of factors (i.e., treatment,

biological sex and genetic line) was distributed across the three hierarchical levels of the study: i) variation across the six animals sampled in each pen at each time point, ii) variation between the two pen replicates within each experiment, and iii) variation among the three experiments.

The average coefficient of variation for iBW of the six sampled animals within each pen was 11.08%, with maximum values reaching 22.5% (Table S12). At day 35, the average BW difference between the largest and smallest animal sampled in each pen was over 31% of the mean value. The mentioned variability was higher than previously reported (Lundberg et al., 2021; Vasdal et al., 2019), although chicken BW variation can increase if animals are subjected to challenging diets (Gous, 2018). The intrapen variability explained most (76.3%) of the variance observed within combinations of factors (Figure 2J). However, its weight respect to experiment factor decreased with the age of animals, probably due to augmenting environmental effects and potentially microbiota development factors. Although the three experiment replicates were identical, and abiotic conditions were controlled in the farm, these were conducted in spring, summer, and autumn, which could have entailed slight differences in the temperature and humidity of the barn.

The whole-genome analyses we will perform in HoloFood will enable ascertaining whether the interindividual genetic variability could be related to the observed dispersion of the data. Genotype-phenotype association studies in commercial chicken lines have identified several important genomic regions that explain a percentage of the BW variation (Dadousis et al., 2021; Tarsani et al., 2019; Wang et al., 2020). In addition, the variability in the intestinal microbiota can intensify differences in performance between individuals from the same population (Shah et al., 2019; Wen et al., 2021; Yan et al., 2017). Ultimately, we will aim at identifying chicken genetic variants associated with microbiota changes with noteworthy impact on performance. These chicken genomic and microbial metagenomic analyses will enable ascertaining to which degree the observed intrapen and interexperiment variation can be attributed to differences in the genetic features of chickens and microbial communities.

Discussion

The range of analytical approaches displayed in this article showcases the strength of implementing new multi-omic approaches to address relevant questions for farming practices. Many questions that would remain unanswered by employing traditional techniques can now be addressed using these new technologies, and most importantly, new questions that were not so far set out (e.g., the reasons for intra-pen variability) can be now proposed. We believe that the hologenomic approach being implemented in HoloFood will help us move from “Does factor X affect KPI Y?” to “How and why does factor X affect KPI Y?”. That is to say, transitioning from a trial-and-error approach to a knowledge-based strategy in which understanding biological processes that underlie the administration of feeds, additives or drugs, as well as the observed interindividual variation, is prioritised. We will address all the questions outlined in this article and more through the collaboration of

multiple academic and industrial partners, aimed at pioneering the large-scale implementation of hologenomics in animal farming (Fig. 3). Although HoloFood will generate and analyse one of the largest multi-omic datasets in farm animals with characterisation of hundreds of specimens, we acknowledge the mathematical challenges of analysing such a complex and hyper- dimensional dataset. The dimensionality of the data will be reduced by leveraging the hierarchical structure of biology itself (e.g., enzymes embedded within metabolic pathways), as well as using the most advanced feature selection approaches to identify the most relevant molecular elements. Ongoing data analyses, which will be published in upcoming articles, will show us how far we can reach, what are the limitations of this novel approach, and how the field can best advance to make the most of the new technologies for a more secure, ethical and sustainable food production.

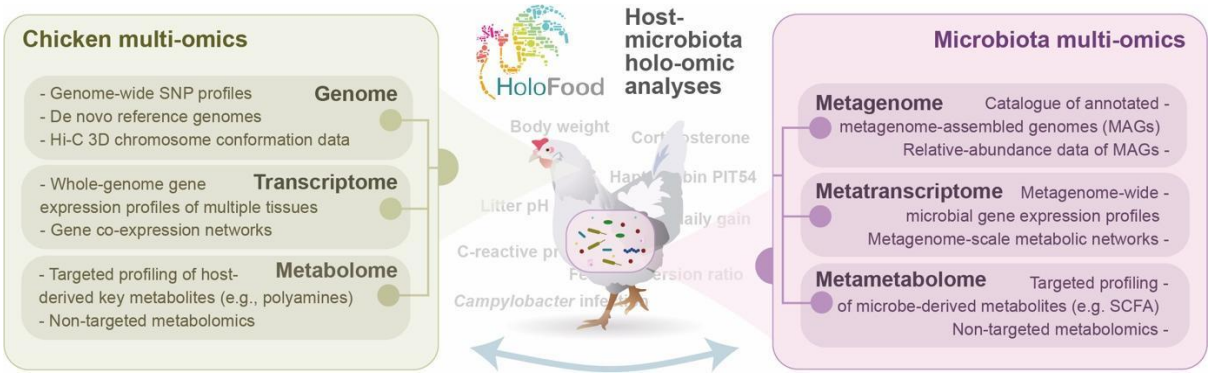


Figure 3. Overview of holo-omic analyses that will be conducted in the H2020 project HoloFood to deepen into the results outlined in this manuscript and address the questions raised and beyond.

Methods

Animal welfare

This experiment followed the EU principles for animal care and experimentation and experimental procedures approved by Ethical Committees of IRTA and Generalitat de Catalunya, Spain (Proceeding number 10226).

Animals and housing

The study consisted of three identical experiments. In each trial a total of 960 day-of-hatch broiler chicks belonging to two fast growing genetic lines (Ross308® and Cobb500®) from two hatcheries (to increase genetic variability) were allocated at 24 pens upon arrival. The name of genetic lines has been blinded and each one is described along the manuscript under the letter X or Y. In order to avoid the possible influence of the parent stock, birds were distributed in such a way that each replicate received the same number of broilers from each hatchery tray. Each pen had a total surface of 2.25 m^2 with 40 birds per pen. Each pen was provided with one individual hopper feeder and two nipple drinkers. The barn is windowless and provided with automatic environment control with a gas heating system by screens and

ventilation by depression. The room also has programmable lighting, provided by TL tubes evenly distributed. The temperature program was adjusted according to the standard program used in the farm: from 0 to 2 days the temperature increased from 32 to 34ºC; from 3 to 7 days the temperature was reduced to 29-31ºC and continued decreasing for 3ºC per week afterwards until reaching 21ºC. The lighting program was 24 hours of light the first two days, 18 hours of light until 7 days, and 14 hours of light per day afterwards. The litter was fresh wood shavings. All birds were vaccinated against Avian Infectious Bronchitis and Gumboro diseases according to the vaccination program usually practised at the hatchery. Moreover, a set of 240 Cobb500 animals from one of the hatcheries in experiment 1 were vaccinated against Marek disease.

The study had a randomised complete block design with a factorial 2 x 2 x 3 arrangement according to broiler line (X or Y), sex (male or female) and dietary treatment (basal diet (BD), probiotic (PR) or phytobiotic (PH)). The three trials were carried out in different seasons of the year: spring (trial A), summer (trial B) and autumn (trial C). Treatments were randomly assigned to one pen of each block (2 blocks per experiment), so that each treatment had 2 replicates (pens) with 40 animals per pen (20 animals from each origin). The six animals to be slaughtered at days 7, 21, and 35 within each pen were randomly selected and marked at day 0 to avoid observer biases in subsequent samplings.

Diets, additives and feeding

Aimed at maximising the effects of the feed additives, the basal diet (BD) was designed as a pro-inflammatory diet, using wheat (more than 50% inclusion) and soybean meal as main ingredients without the addition of enzymes, antibiotics, or coccidiostats. Diets were formulated according to birds' requirements and commercial practices in three different periods: starter (0-9 d), grower (10-23 d) and finisher (24-37 d). Feeds were presented in crumble form for the starter period and in 3 mm pellets later on. Composition of diets and estimated nutrient contents are presented in Table S1. The probiotic treatment (PR) consisted of BD with a mixture of three strains, namely *Bacillus subtilis* DSM 32324, *B. subtilis* DSM 32325 and *B. amyloliquefaciens* DSM 25840, which has been recently authorised as a zootechnical product for poultry species (EFSA Panel on Additives and Products or Substances used in Animal Feed (FEEDAP) et al., 2020). The phytobiotic treatment (PH) consisted of BD with a phytobiotic additive obtained from white grapes and containing procyanidins and polyphenols as active ingredients.

Batch feed samples were taken from each production for proximate analysis (AOAC, 2000, moisture -dry matter- by oven drying –method 2-, nitrogen -crude protein- by combustion -Dumas method-, ether extract on a Soxtec system -method 3B- and ash after muffle furnace incineration -method 12) and to quantify concentrations of probiotic and phytobiotic additives. Data of analytical composition of diets is shown in Table S1. In addition, water and litter samples were collected in each pen at days 0, 7, 21, and 35 for microbiological and litter quality analyses. Moreover, three samples of each feed per diet period were taken from three different bags for microbiological assessment.

Animal monitoring

Animals were counted and weighed by hatchery tray upon arrival and by replicate at days 7, 21, and 35. Growth performance was monitored to replicate the same days. Pen level analyses included average body weight (BW), daily gain (ADG), daily feed intake (ADFI), feed conversion ratio (FCR) and European production efficiency factor (EPEF). The incidence and severity of footpad dermatitis per pen was subjectively evaluated by trained personnel at days 7, 21 and 35 according the 5-point scale described by Butterworth (Welfare Quality® Consortium, Lelystad, the Netherlands, 2009): 0) no evidence of footpad dermatitis; 1 & 2) minimal evidence of footpad dermatitis; 3 & 4) evidence of footpad dermatitis. Dead animals were weighted, and the most probable cause of death recorded. Animals (laggards) excluded from the trial during the first week were not considered.

Animal sampling

At day 0, eight unused animals (leftovers) were euthanised and sampled. At days 7-8, 21-22 and 35-37 (multiple days were necessary due to workload), 6 animals per pen (144 animals per experiment) were randomly selected, euthanised and sampled. Animals were euthanised according to RD 53/2013 (Spain), following the ethical requirements established. After the euthanasia, animals were individually weighed, and footpad dermatitis was assessed. Sections of ileum and cecum, intestinal content from ileum and cecum, feathers, and blood were obtained through the coordinated action of a group of 16-18 researchers. Different aliquots were distributed and properly stored for downstream analyses. Liver, pancreas, thymus, bursa of Fabricius, brain, and spleen samples were also obtained for future analyses. From these animals, individual pathogen detection, corticosterone in feathers, acute phase proteins and lipopolysaccharide concentration in plasma were measured. Complete list of samples is displayed in Table 1.

In addition, at day 37 (commercial slaughtering age), 6 extra animals per pen were randomly selected, euthanised and subjected to evaluation of meat quality traits (carcass, abdominal fat, breast, and leg yield). The oxidative stability of the thigh muscle was determined from randomly selected three animals over a period of 7 days.

Analytical procedures

Pathogen detection

Salmonella, *Campylobacter* and *Clostridium* spp. were detected from caecal content through Real-time PCR. *Salmonella* spp. was detected using the primers 5’-GTGAAATTATCGCCACGTTCGGGCAA-3’ and 5’-TCATCGCACCGTCAAAGGAACC-3’, which are specific for the InvA gene of *Salmonella*. For the detection of *Clostridium perfringens*, the primers 5’-AAGATTTGTAAGGCGCTT-3’ and 5’-ATTTCCTGAAATCCACTC-3’ specific for alpha-toxin gene present in all strains of this bacteria species were used. The presence of *Campylobacter* spp. was detected using the primers 5’-TTGGAAACGACTGCTAATACTCTA-3’ and 5’-AGCCATTAGATTTCACAAGAGACT-3’, which amplify a specific segment of 16S ribosomal RNA gene specific of *Campylobacter* spp.

Corticosterone in feathers

Corticosterone (COR) was determined in accordance with the method described by Bortolotti et al. (2008) (Bortolotti et al., 2008).

Acute phase proteins in plasma

Chicken haptoglobin-like protein (PIT54) was quantified using the haptoglobin ELISA Kit (Ref. ABIN1563052, antibodies-online.com) and C-reactive protein (CRP) was measured using the CRP ELISA Kit (Ref. ABIN4947413, antibodies-online.com) in accordance with manufacturer instructions.

Lipopolysaccharide in plasma

Lipopolysaccharide concentration in plasma was determined using Pierce LAL Chromogenic Endotoxin Quantification Kit (Ref. 88282; Thermofisher, USA) in accordance with manufacturer instructions.

pH of litter

Three samples of 50 g of litter per pen were collected for pH determination at days 7, 21 and 35 avoiding the areas near and below the feeders and drinkers. The three samples collected from the same pen and day were pooled and homogenised and the moisture was determined in a sub-sample of 100 g according to the AOAC method (AOAC, 2000; method 925.09). For pH analysis, a subsample of 10 g was placed in a beaker with 100 ml of distilled water, shaked with a glass rod and allowed to stand for 30 minutes. The pH value was obtained using a pH metre (Crison, L'Hospitalet de Llobregat, Spain).

Oxidative stability of meat

Samples (5 g of muscle from the thigh) were homogenised with an aqueous 7.5% trichloroacetic acid solution, filtered and brought to 20 ml. To proceed, 5 ml of the extraction solution and 5 ml of 0.02 M thiobarbituric acid were mixed and boiled for 15 minutes and then cooled in cold water. Absorbance of the peak was measured at 525 nm as malondialdehyde production in an ultraviolet-visible spectrophotometer (Shimadzu, Japan) using the third derivative of the spectrum between 425 and 650 to correct the baseline. The 1,1,3,3-tetraethoxypropane was used as standard (Botsoglou et al., 1994; Ruiz et al., 1999).

Statistical analyses

Data was explored to discard any possible outlier according to the Kolmogorov-Smirnov test (Massey, 1951). As no outliers were considered, the statistical analysis included all data. The GLIMMIX procedure of SAS software (SAS/STAT 14.1; SAS Institute Inc., Cary, NC, USA) was used to perform the analysis of the different variables. In the case of ELISA determinations, when the limit of detection was not reached, the missing values were replaced by the limit of detection (L)/√2 (Hornung & Reed, 1990). The statistical model used is shown below:

$$y_{ijklm} = \mu + T_i + B_k + S_l + E_m + \gamma_j + e_{ijklm}$$

Where y_{ijklm} is the response variable, T_i is the dietary treatment effect, B_k is the broiler line effect, S_l is the biological sex effect, E_m is the experiment effect, γ_j is the random block effect, and e_{ij} is the error of the experimental unit. The experimental unit was the pen. Results in Tables S1-12 are expressed as least square means ± standard error. Differences were considered significant at $P < 0.05$, while those at $P < 0.1$ are reported as tendencies.

Hierarchical decomposition of the variance was carried out using ANOVA-estimation of variance components as implemented in the fitVCA function of the R package VCA (Schuetzenmeister et al., 2017), using individual iBW as response variable, and pen and experiment as explanatory variables.

Acknowledgements

This research was funded by the European Union's Horizon Research and Innovation Programme under grant agreement No. 817729 (HoloFood, Holistic solution to improve animal food production through deconstructing the biomolecular interactions between feed, gut microorganisms and animals in relation to performance parameters). We would like to thank the researchers, farm workers and volunteers who participated in the design, maintenance and execution of the animal trials, as well as in the collection and processing of the samples.

References

Aggrey, S. E. (2002). Comparison of three nonlinear and spline regression models for describing chicken growth curves. *Poultry Science*, *81*(12), 1782–1788. https://doi.org/10.1093/ps/81.12.1782

Agovino, M., Casaccia, M., Ciommi, M., Ferrara, M., & Marchesano, K. (2019). Agriculture, climate change and sustainability: The case of EU-28. *Ecological Indicators*, *105*, 525–543. https://doi.org/10.1016/j.ecolind.2018.04.064

Ahn, J., Woodfint, R. M., Lee, J., Wu, H., Ma, J., Suh, Y., Hwang, S., Cressman, M., & Lee, K. (2019). Comparative identification, nutritional, and physiological regulation of chicken liver-enriched genes. *Poultry Science*, *98*(7), 3007–3013. https://doi.org/10.3382/ps/pez057

Alberdi, A., Andersen, S. B., Limborg, M. T., Dunn, R., & Gilbert, M. T. P. (2021). Disentangling host-microbiota complexity through hologenomics. *Nature Reviews. Genetics*.

Almeida, A., Mitchell, A. L., Boland, M., Forster, S. C., Gloor, G. B., Tarkowska, A., Lawley, T. D., & Finn, R. D. (2019). A new genomic blueprint of the human gut microbiota. *Nature*, *568*(7753), 499–504. https://doi.org/10.1038/s41586-019-0965-1

Andersen, T. O., Kunath, B. J., Hagen, L. H., Arntzen, M. Ø., & Pope, P. B. (2021). Rumen metaproteomics: Closer to linking rumen microbial function to animal productivity traits. *Methods* , *186*, 42–51. https://doi.org/10.1016/j.ymeth.2020.07.011

Ansari, I., Raddatz, G., Gutekunst, J., Ridnik, M., Cohen, D., Abu-Remaileh, M., Tuganbaev, T., Shapiro, H., Pikarsky, E., Elinav, E., Lyko, F., & Bergman, Y. (2020). The microbiota programs DNA methylation to control intestinal homeostasis and inflammation. *Nature Microbiology*, *5*(4), 610–619. https://doi.org/10.1038/s41564-019-0659-3

Aviagen. (2019). *Ross 308: Performance Objectives*.

Awad, W. A., Hess, C., & Hess, M. (2018). Re-thinking the chicken-Campylobacter jejuni interaction: a review. *Avian Pathology: Journal of the W.V.P.A*, *47*(4), 352–363. https://doi.org/10.1080/03079457.2018.1475724

Awad, W. A., Molnár, A., Aschenbach, J. R., Ghareeb, K., Khayal, B., Hess, C., Liebhart, D., Dublecz, K., & Hess, M. (2015). Campylobacter infection in chickens modulates the intestinal epithelial barrier function. *Innate Immunity*, *21*(2), 151–160. https://doi.org/10.1177/1753425914521648

Baker, M. (2012). De novo genome assembly: what every biologist should know. *Nature Methods*, *9*(4), 333–337. https://doi.org/10.1038/nmeth.1935

Ballou, A. L., Ali, R. A., Mendoza, M. A., Ellis, J. C., Hassan, H. M., Croom, W. J., & Koci, M. D. (2016). Development of the Chick Microbiome: How Early Exposure Influences Future Microbial Diversity. *Frontiers in Veterinary Science*, *3*, 2. https://doi.org/10.3389/fvets.2016.00002

Bortolotti, G. R., Marchant, T. A., Blas, J., & German, T. (2008). Corticosterone in feathers is a long-term, integrated measure of avian stress physiology. *Functional Ecology*, *22*(3), 494–500. https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2435.2008.01387.x

Bortoluzzi, C., Serpa Vieira, B., de Paula Dorigam, J. C., Menconi, A., Sokale, A., Doranalli, K., & Applegate, T. J. (2019). Bacillus subtilis DSM 32315 Supplementation Attenuates the Effects of Clostridium perfringens Challenge on the Growth Performance and Intestinal Microbiota of Broiler Chickens. *Microorganisms*, *7*(3). https://doi.org/10.3390/microorganisms7030071

Botsoglou, N. A., Fletouris, D. J., Papageorgiou, G. E., Vassilopoulos, V. N., Mantis, A. J., & Trakatellis, A. G. (1994). Rapid, Sensitive, and Specific Thiobarbituric Acid Method for Measuring Lipid Peroxidation in Animal Tissue, Food, and Feedstuff Samples. *Journal of Agricultural and Food Chemistry*, *42*(9), 1931–1937. https://doi.org/10.1021/jf00045a019

Broom, L. J., & Kogut, M. H. (2018). The role of the gut microbiome in shaping the immune system of chickens. *Veterinary Immunology and Immunopathology*, *204*, 44–51. https://doi.org/10.1016/j.vetimm.2018.10.002

Carbajal, A., Tallo-Parra, O., Sabes-Alsina, M., Mular, I., & Lopez-Bejar, M. (2014). Feather corticosterone evaluated by ELISA in broilers: a potential tool to evaluate broiler welfare. *Poultry Science*, *93*(11), 2884–2886. https://doi.org/10.3382/ps.2014-04092

Cardoso Dal Pont, G., Farnell, M., Farnell, Y., & Kogut, M. H. (2020). Dietary Factors as Triggers of Low-Grade Chronic Intestinal Inflammation in Poultry. *Microorganisms*, *8*(1). https://doi.org/10.3390/microorganisms8010139

Chaloner, G., Wigley, P., & Humphrey, S. (2014). Dynamics of dual infection with Campylobacter jejuni strains in chickens reveals distinct strain-to-strain variation in infection ecology. *Applied and*. https://journals.asm.org/doi/abs/10.1128/AEM.01901-14

Chamorro, S., Romero, C., Brenes, A., Sánchez-Patán, F., Bartolomé, B., Viveros, A., & Arija, I. (2019). Impact of a sustained consumption of grape extract on digestion, gut microbial metabolism and intestinal barrier in broiler chickens. *Food & Function*, *10*(3), 1444–1454. https://doi.org/10.1039/c8fo02465k

Cobb-Vantress. (2018). *Cobb 500: Broiler Performance & Nutrition Supplement*.

Costantini, D., & Møller, A. P. (2009). Does immune response cause oxidative stress in birds? A meta-analysis. *Comparative Biochemistry and Physiology. Part A, Molecular & Integrative Physiology*, *153*(3), 339–344. https://doi.org/10.1016/j.cbpa.2009.03.010

Cryan, J. F., O'Riordan, K. J., Cowan, C. S. M., Sandhu, K. V., Bastiaanssen, T. F. S., Boehme, M., Codagnone, M. G., Cussotto, S., Fulling, C., Golubeva, A. V., Guzzetta, K. E., Jaggar, M., Long-Smith, C. M., Lyte, J. M., Martin, J. A., Molinero-Perez, A., Moloney, G., Morelli, E., Morillas, E., … Dinan, T. G. (2019). The Microbiota-Gut-Brain Axis. *Physiological Reviews*, *99*(4), 1877–2013. https://doi.org/10.1152/physrev.00018.2018

Dadousis, C., Somavilla, A., Ilska, J. J., Johnsson, M., Batista, L., Mellanby, R. J., Headon, D., Gottardo, P., Whalen, A., Wilson, D., Dunn, I. C., Gorjanc, G., Kranis, A., & Hickey, J. M. (2021). A genome-wide association analysis for body weight at 35 days measured on 137,343 broiler chickens. *Genetics, Selection, Evolution: GSE*, *53*(1), 70. https://doi.org/10.1186/s12711-021-00663-w

Debray, R., Herbert, R. A., Jaffe, A. L., Crits-Christoph, A., Power, M. E., & Koskella, B. (2021). Priority effects in microbiome assembly. *Nature Reviews. Microbiology*.

https://doi.org/10.1038/s41579-021-00604-w

de Zoete, M. R., Keestra, A. M., Roszczenko, P., & van Putten, J. P. M. (2010). Activation of human and chicken toll-like receptors by Campylobacter spp. *Infection and Immunity*, *78*(3), 1229–1238. https://doi.org/10.1128/IAI.00897-09

Diaz Carrasco, J. M., Casanova, N. A., & Fernández Miyakawa, M. E. (2019). Microbiota, Gut Health and Chicken Productivity: What Is the Connection? *Microorganisms*, *7*(10). https://doi.org/10.3390/microorganisms7100374

Duangnumsawang, Y., Zentek, J., & Goodarzi Boroojeni, F. (2021). Development and Functional Properties of Intestinal Mucus Layer in Poultry. *Frontiers in Immunology*, *12*, 3924. https://doi.org/10.3389/fimmu.2021.745849

EFSA Panel on Additives and Products or Substances used in Animal Feed (FEEDAP), Bampidis, V., Azimonti, G., de Lourdes Bastos, M., Christensen, H., Dusemund, B., Kouba, M., Kos Durjava, M., López-Alonso, M., López Puente, S., Marcon, F., Mayo, B., Pechová, A., Petkova, M., Ramos, F., Sanz, Y., Villa, R. E., Woutersen, R., Cocconcelli, P. S., … Pettenati, E. (2020). Safety and efficacy of GalliPro® Fit (Bacillus subtilis DSM 32324, Bacillus subtilis DSM 32325 and Bacillus amyloliquefaciens DSM 25840) for all poultry species for fattening or reared for laying/breeding. *EFSA Journal*, *18*(4), e06094. https://doi.org/10.2903/j.efsa.2020.6094

European Food Safety Authority, & European Centre for Disease Prevention and Control. (2021). The European union one health 2019 zoonoses report. *EFSA Journal*, *19*(2), e06406. https://doi.org/10.2903/j.efsa.2021.6406

Farahat, M., Ibrahim, D., Kishawy, A. T. Y., Abdallah, H. M., Hernandez-Santana, A., & Attia, G. (2021). Effect of cereal type and plant extract addition on the growth performance, intestinal morphology, caecal microflora, and gut barriers gene expression of broiler chickens. *Animal: An International Journal of Animal Bioscience*, *15*(3), 100056. https://doi.org/10.1016/j.animal.2020.100056

Fraune, S., & Bosch, T. C. G. (2010). Why bacteria matter in animal development and evolution. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, *32*(7), 571–580. https://doi.org/10.1002/bies.200900192

Gilroy, R., Ravi, A., Getino, M., Pursley, I., & Horton, D. L. (2021). Extensive microbial diversity within the chicken gut microbiome revealed by metagenomics and culture. *PeerJ*. https://www.ncbi.nlm.nih.gov/pmc/articles/pmc8035907/

Glendinning, L., Stewart, R. D., Pallen, M. J., Watson, K. A., & Watson, M. (2020). Assembly of hundreds of novel bacterial genomes from the chicken caecum. *Genome Biology*, *21*(1), 34. https://doi.org/10.1186/s13059-020-1947-1

Gong, K., Chen, L., Li, X., Sun, L., & Liu, K. (2018). Effects of germination combined with extrusion on the nutritional composition, functional properties and polyphenol profile and related in vitro hypoglycemic effect of whole grain corn. *Journal of Cereal Science*, *83*, 1–8. https://www.sciencedirect.com/science/article/pii/S0733521017309657?casa_token=n7UHKQcCOCsAAAAA:RePWz1dmQN1L4A6-4ACPZYPo-v9sdtIqATfHPYnEq0vexeuPr4HakQOCmRDIJ-0Aele9MANEHoc

Goodarzi Boroojeni, F., Vahjen, W., Männer, K., Blanch, A., Sandvang, D., & Zentek, J. (2018). Bacillus subtilis in broiler diets with different levels of energy and protein. *Poultry Science*, *97*(11), 3967–3976. https://doi.org/10.3382/ps/pey265

Gopi, M., Dutta, N., Pattanaik, A. K., Jadhav, S. E., Madhupriya, V., Tyagi, P. K., & Mohan, J. (2020). Effect of polyphenol extract on performance, serum biochemistry, skin pigmentation and carcass characteristics in broiler chickens fed with different cereal sources under hot-humid conditions. *Saudi Journal of Biological Sciences*, *27*(10), 2719–2726. https://www.sciencedirect.com/science/article/pii/S1319562X20302618

Gous, R. M. (2018). Nutritional and environmental effects on broiler uniformity. *World's Poultry Science Journal*, *74*(1), 21–34. https://doi.org/10.1017/S0043933917001039

Graw, S., Chappell, K., Washam, C. L., Gies, A., Bird, J., Robeson, M. S., 2nd, & Byrum, S. D. (2021). Multi-omics data integration considerations and study design for biological systems and disease. *Molecular Omics*, *17*(2), 170–185. https://doi.org/10.1039/d0mo00041h

Gutiérrez, A. M., Montes, A., Gutiérrez-Panizo, C., Fuentes, P., & De La Cruz-Sánchez, E. (2018). Gender influence on the salivary protein profile of finishing pigs. *Journal of Proteomics*, *178*, 107–113. https://doi.org/10.1016/j.jprot.2017.11.023

Hausman, G. J., Barb, C. R., Fairchild, B. D., Gamble, J., & Lee-Rutherford, L. (2014). Gene expression profiling in adipose tissue from growing broiler chickens. *Adipocyte*, *3*(4), 297–303. https://doi.org/10.4161/adip.29252

HoloFood Consortium. (2019, January 1). *Holistic solution to improve animal food production through deconstructing the biomolecular interactions between feed, gut microorganisms and animals in relation to performance parameters*. CORDIS. https://cordis.europa.eu/project/id/817729

Hornung, R. W., & Reed, L. D. (1990). Estimation of Average Concentration in the Presence of Nondetectable Values. *Applied Occupational and Environmental Hygiene*, *5*(1), 46–51. https://doi.org/10.1080/1047322X.1990.10389587

Iqbal, J., Mukhtar, N., Rehman, Z. U., Khan, S. H., Ahmad, T., Anjum, M. S., Pasha, R. H., & Umar, S. (2017). Effects of egg weight on the egg quality, chick quality, and broiler performance at the later stages of production (week 60) in broiler breeders. *The Journal of Applied Poultry Research*, *26*(2), 183–191. https://doi.org/10.3382/japr/pfw061

Jimeno, B., Hau, M., & Verhulst, S. (2018). Corticosterone levels reflect variation in metabolic rate, independent of "stress." *Scientific Reports*, *8*(1), 13020. https://doi.org/10.1038/s41598-018-31258-z

Johnson, K. V.-A., & Foster, K. R. (2018). Why does the microbiome affect behaviour? *Nature Reviews. Microbiology*, *16*(10), 647–655. https://doi.org/10.1038/s41579-018-0014-3

Jurburg, S. D., Brouwer, M. S. M., Ceccarelli, D., van der Goot, J., Jansman, A. J. M., & Bossers, A. (2019). Patterns of community assembly in the developing chicken microbiome reveal rapid primary succession. *MicrobiologyOpen*, *8*(9), e00821. https://doi.org/10.1002/mbo3.821

Kaab, H., Bain, M. M., & Eckersall, P. D. (2018). Acute phase proteins and stress markers in the immediate response to a combined vaccination against Newcastle disease and infectious bronchitis viruses in specific pathogen free (SPF) layer chicks. *Poultry Science*, *97*(2), 463–469. https://www.sciencedirect.com/science/article/pii/S003257911930896X

Kanehisa, M., Sato, Y., & Kawashima, M. (2021). KEGG mapping tools for uncovering hidden features in biological data. *Protein Science: A Publication of the Protein Society*. https://doi.org/10.1002/pro.4172

Khan, A. A., Yurkovetskiy, L., O'Grady, K., Pickard, J. M., de Pooter, R., Antonopoulos, D. A., Golovkina, T., & Chervonsky, A. (2019). Polymorphic Immune Mechanisms Regulate Commensal Repertoire. *Cell Reports*, *29*(3), 541–550.e4. https://doi.org/10.1016/j.celrep.2019.09.010

Kollarcikova, M., Kubasova, T., Karasova, D., Crhanova, M., Cejkova, D., Sisak, F., & Rychlik, I. (2019). Use of 16S rRNA gene sequencing for prediction of new opportunistic pathogens in chicken ileal and cecal microbiota. *Poultry Science*, *98*(6), 2347–2353. https://doi.org/10.3382/ps/pey594

Lee, K.-C., Kil, D. Y., & Sul, W. J. (2017). Cecal microbiome divergence of broiler chickens by sex and body weight. *Journal of Microbiology* , *55*(12), 939–945. https://doi.org/10.1007/s12275-017-7202-0

Li, C.-L., Wang, J., Zhang, H.-J., Wu, S.-G., Hui, Q.-R., Yang, C.-B., Fang, R.-J., & Qi, G.-H. (2018). Intestinal Morphologic and Microbiota Responses to Dietary Bacillus spp. in a Broiler Chicken Model. *Frontiers in Physiology*, *9*, 1968. https://doi.org/10.3389/fphys.2018.01968

Lieberman-Aiden, E., van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B. R., Sabo, P. J., Dorschner, M. O., Sandstrom, R., Bernstein, B., Bender, M. A., Groudine, M., Gnirke, A., Stamatoyannopoulos, J., Mirny, L. A., Lander, E. S., & Dekker, J. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, *326*(5950), 289–293. https://doi.org/10.1126/science.1181369

Li, R., Li, Y., Fang, X., Yang, H., Wang, J., Kristiansen, K., & Wang, J. (2009). SNP detection for massively parallel whole-genome resequencing. *Genome Research*, *19*(6), 1124–1132. https://doi.org/10.1101/gr.088013.108

Liu, J., Gu, Z., Song, F., Zhang, H., Zhao, J., & Chen, W. (2019). Lactobacillus plantarum ZS2058 and Lactobacillus rhamnosus GG Use Different Mechanisms to Prevent Salmonella Infection in vivo. *Frontiers in Microbiology*, *10*, 299. https://doi.org/10.3389/fmicb.2019.00299

Livingston, M. L., Cowieson, A. J., Crespo, R., Hoang, V., Nogal, B., Browning, M., & Livingston, K. A. (2020). Effect of broiler genetics, age, and gender on performance and blood chemistry. *Heliyon*,

6(7), e04400. https://doi.org/10.1016/j.heliyon.2020.e04400

Lundberg, R., Scharch, C., & Sandvang, D. (2021). The link between broiler flock heterogeneity and cecal microbiome composition. *Animal Microbiome*, *3*(1), 54. https://doi.org/10.1186/s42523-021-00110-7

Ma, D., Bou-Sleiman, M., Joncour, P., Indelicato, C.-E., Frochaux, M., Braman, V., Litovchenko, M., Storelli, G., Deplancke, B., & Leulier, F. (2019). Commensal Gut Bacteria Buffer the Impact of Host Genetic Variants on Drosophila Developmental Traits under Nutritional Stress. *iScience*, *19*, 436–447. https://doi.org/10.1016/j.isci.2019.07.048

Massey, F. J. (1951). The Kolmogorov-Smirnov Test for Goodness of Fit. *Journal of the American Statistical Association*, *46*(253), 68–78. https://doi.org/10.1080/01621459.1951.10500769

McFall-Ngai, M., Hadfield, M. G., Bosch, T. C. G., Carey, H. V., Domazet-Lošo, T., Douglas, A. E., Dubilier, N., Eberl, G., Fukami, T., Gilbert, S. F., Hentschel, U., King, N., Kjelleberg, S., Knoll, A. H., Kremer, N., Mazmanian, S. K., Metcalf, J. L., Nealson, K., Pierce, N. E., … Wernegreen, J. J. (2013). Animals in a bacterial world, a new imperative for the life sciences. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(9), 3229–3236. https://doi.org/10.1073/pnas.1218525110

Messerli, P., Murniningtyas, E., Eloundou-Enyegue, P., Foli, E. G., Furman, E., Glassman, A., Hernández Licona, G., Kim, E. M., Lutz, W., Moatti, J.-P., & Others. (2019). *Global sustainable development report 2019: the future is now--science for achieving sustainable development*. http://pure.iiasa.ac.at/id/eprint/16067/1/24797GSDR_report_2019.pdf

Michalak, L., Gaby, J. C., Lagos, L., La Rosa, S. L., Hvidsten, T. R., Tétard-Jones, C., Willats, W. G. T., Terrapon, N., Lombard, V., Henrissat, B., Dröge, J., Arntzen, M. Ø., Hagen, L. H., Øverland, M., Pope, P. B., & Westereng, B. (2020). Microbiota-directed fibre activates both targeted and secondary metabolic shifts in the distal gut. *Nature Communications*, *11*(1), 5773. https://doi.org/10.1038/s41467-020-19585-0

Mohd Shaufi, M. A., Sieo, C. C., Chong, C. W., Gan, H. M., & Ho, Y. W. (2015). Deciphering chicken gut microbial dynamics based on high-throughput 16S rRNA metagenomics analyses. *Gut Pathogens*, *7*, 4. https://doi.org/10.1186/s13099-015-0051-7

Molnár, A. K., Podmaniczky, B., Kürti, P., Tenk, I., Glávits, R., Virág, G., & Szabó, Z. (2011). Effect of different concentrations of Bacillus subtilis on growth performance, carcase quality, gut microflora and immune response of broiler chickens. *British Poultry Science*, *52*(6), 658–665. https://doi.org/10.1080/00071668.2011.636029

Nordquist, R. E., Zeinstra, E. C., Dougherty, A., & Riber, A. B. (2020). Effects of Dark Brooder Rearing and Age on Hypothalamic Vasotocin and Feather Corticosterone Levels in Laying Hens. *Frontiers in Veterinary Science*, *7*, 19. https://doi.org/10.3389/fvets.2020.00019

Nyholm, L., Koziol, A., Marcos, S., Botnen, A. B., Aizpurua, O., Gopalakrishnan, S., Limborg, M. T., Gilbert, M. T. P., & Alberdi, A. (2020). Holo-Omics: Integrated Host-Microbiota Multi-omics for Basic and Applied Biological Research. *iScience*, *23*(8), 101414. https://doi.org/10.1016/j.isci.2020.101414

Ocejo, M., Oporto, B., & Hurtado, A. (2019). 16S rRNA amplicon sequencing characterization of caecal microbiome composition of broilers and free-range slow-growing chickens throughout their productive lifespan. *Scientific Reports*, *9*(1), 2506. https://doi.org/10.1038/s41598-019-39323-x

O'reilly, E. L., & Eckersall, P. D. (2014). Acute phase proteins: a review of their function, behaviour and measurement in chickens. *World's Poultry Science Journal*, *70*(1), 27–44. https://doi.org/10.1017/S0043933914000038

Pasolli, E., Asnicar, F., Manara, S., Zolfo, M., Karcher, N., Armanini, F., Beghini, F., Manghi, P., Tett, A., Ghensi, P., Collado, M. C., Rice, B. L., DuLong, C., Morgan, X. C., Golden, C. D., Quince, C., Huttenhower, C., & Segata, N. (2019). Extensive Unexplored Human Microbiome Diversity Revealed by Over 150,000 Genomes from Metagenomes Spanning Age, Geography, and Lifestyle. *Cell*, *176*(3), 649–662.e20. https://doi.org/10.1016/j.cell.2019.01.001

Raza, A., Bashir, S., & Tabassum, R. (2019). An update on carbohydrases: growth performance and intestinal health of poultry. *Heliyon*, *5*(4), e01437. https://doi.org/10.1016/j.heliyon.2019.e01437

Reid, W. V., Chen, D., Goldfarb, L., Hackmann, H., Lee, Y. T., Mokhele, K., Ostrom, E., Raivio, K., Rockström, J., Schellnhuber, H. J., & Whyte, A. (2010). Earth System Science for Global Sustainability: Grand Challenges. *Science*.

https://science.sciencemag.org/content/330/6006/916.summary

Ruiz, J. A., Pérez-Vendrell, A. M., & Esteve-García, E. (1999). Effect of β-Carotene and Vitamin E on Oxidative Stability in Leg Meat of Broilers Fed Different Supplemental Fats. *Journal of Agricultural and Food Chemistry*, *47*(2), 448–454. https://doi.org/10.1021/jf980825g

Scanes, C. G. (2016). Biology of stress in poultry with emphasis on glucocorticoids and the heterophil to lymphocyte ratio. *Poultry Science*, *95*(9), 2208–2215. https://doi.org/10.3382/ps/pew137

Schuetzenmeister, A., Dufey, F., Schuetzenmeister, M. A., & Suggests, V. F. P. (2017). Package "VCA." *R Package*, 1–112. https://cran.microsoft.com/snapshot/2018-07-03/web/packages/VCA/VCA.pdf

Shaffer, M., Borton, M. A., McGivern, B. B., Zayed, A. A., La Rosa, S. L., Solden, L. M., Liu, P., Narrowe, A. B., Rodríguez-Ramos, J., Bolduc, B., Gazitúa, M. C., Daly, R. A., Smith, G. J., Vik, D. R., Pope, P. B., Sullivan, M. B., Roux, S., & Wrighton, K. C. (2020). DRAM for distilling microbial metabolism to automate the curation of microbiome function. *Nucleic Acids Research*, *48*(16), 8883–8900. https://doi.org/10.1093/nar/gkaa621

Shah, T. M., Patel, J. G., Gohil, T. P., Blake, D. P., & Joshi, C. G. (2019). Host transcriptome and microbiome interaction modulates physiology of full-sibs broilers with divergent feed conversion ratio. *NPJ Biofilms and Microbiomes*, *5*, 24. https://doi.org/10.1038/s41522-019-0096-3

Slawinska, A., Dunislawska, A., Plowiec, A., & Radomska, M. (2019). Modulation of microbial communities and mucosal gene expression in chicken intestines after galactooligosaccharides delivery In Ovo. *PLoS*. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0212318

Sprockett, D., Fukami, T., & Relman, D. A. (2018). Role of priority effects in the early-life assembly of the gut microbiota. *Nature Reviews. Gastroenterology & Hepatology*, *15*(4), 197–205. https://doi.org/10.1038/nrgastro.2017.173

Tarradas, J., Tous, N., Esteve-Garcia, E., & Brufau, A. J. (2020). The Control of Intestinal Inflammation: A Major Objective in the Research of Probiotic Strains as Alternatives to Antibiotic Growth Promoters in Poultry. *Microorganisms*, *8*(2). https://doi.org/10.3390/microorganisms8020148

Tarsani, E., Kranis, A., Maniatis, G., Avendano, S., Hager-Theodorides, A. L., & Kominakis, A. (2019). Discovery and characterization of functional modules associated with body weight in broilers. *Scientific Reports*, *9*(1), 9125. https://doi.org/10.1038/s41598-019-45520-5

van der Hee, B., & Wells, J. M. (2021). Microbial Regulation of Host Physiology by Short-chain Fatty Acids. *Trends in Microbiology*, *0*(0). https://doi.org/10.1016/j.tim.2021.02.001

Vasdal, G., Granquist, E. G., Skjerve, E., de Jong, I. C., Berg, C., Michel, V., & Moe, R. O. (2019). Associations between carcass weight uniformity and production measures on farm and at slaughter in commercial broiler flocks. *Poultry Science*, *98*(10), 4261–4268. https://doi.org/10.3382/ps/pez252

Vilà, B., Esteve-Garcia, E., & Brufau, J. (2010). Probiotic micro-organisms: 100 years of innovation and efficacy; modes of action. *World's Poultry Science Journal*, *66*(3), 369–380. https://doi.org/10.1017/S0043933910000474

Volf, J., Polansky, O., Sekelova, Z., Velge, P., Schouler, C., Kaspers, B., & Rychlik, I. (2017). Gene expression in the chicken caecum is dependent on microbiota composition. *Veterinary Research*, *48*(1), 85. https://doi.org/10.1186/s13567-017-0493-7

Wang, Y., Cao, X., Luo, C., Sheng, Z., Zhang, C., Bian, C., Feng, C., Li, J., Gao, F., Zhao, Y., Jiang, Z., Qu, H., Shu, D., Carlborg, Ö., Hu, X., & Li, N. (2020). Multiple ancestral haplotypes harboring regulatory mutations cumulatively contribute to a QTL affecting chicken growth traits. *Communications Biology*, *3*(1), 472. https://doi.org/10.1038/s42003-020-01199-3

Wen, C., Yan, W., Mai, C., Duan, Z., Zheng, J., Sun, C., & Yang, N. (2021). Joint contributions of the gut microbiota and host genetics to feed efficiency in chickens. *Microbiome*, *9*(1), 126. https://doi.org/10.1186/s40168-021-01040-x

Whelan, R. A., Doranalli, K., Rinttilä, T., Vienola, K., Jurgens, G., & Apajalahti, J. (2019). The impact of Bacillus subtilis DSM 32315 on the pathology, performance, and intestinal microbiome of broiler chickens in a necrotic enteritis challenge. *Poultry Science*, *98*(9), 3450–3463. https://doi.org/10.3382/ps/pey500

Wick, R. R., Judd, L. M., Gorrie, C. L., & Holt, K. E. (2017). Unicycler: Resolving bacterial genome

assemblies from short and long sequencing reads. *PLoS Computational Biology*, *13*(6), e1005595. https://doi.org/10.1371/journal.pcbi.1005595

Yan, W., Sun, C., Yuan, J., & Yang, N. (2017). Gut metagenomic analysis reveals prominent roles of Lactobacillus and cecal microbiota in chicken feed efficiency. *Scientific Reports*, *7*, 45308. https://doi.org/10.1038/srep45308

Zhang, B., Li, G., Shahid, M. S., Gan, L., Fan, H., Lv, Z., Yan, S., & Guo, Y. (2020). Dietary l-arginine supplementation ameliorates inflammatory response and alters gut microbiota composition in broiler chickens infected with Salmonella enterica serovar Typhimurium. *Poultry Science*, *99*(4), 1862–1874. https://doi.org/10.1016/j.psj.2019.10.049

Zhong, C.-Y., Sun, W.-W., Ma, Y., Zhu, H., Yang, P., Wei, H., Zeng, B.-H., Zhang, Q., Liu, Y., Li, W.-X., Chen, Y., Yu, L., & Song, Z.-Y. (2015). Microbiota prevents cholesterol loss from the body by regulating host gene expression in mice. *Scientific Reports*, *5*, 10512. https://doi.org/10.1038/srep10512

Zhou, Y., Zhang, M., Liu, Q., & Feng, J. (2021). The alterations of tracheal microbiota and inflammation caused by different levels of ammonia exposure in broiler chickens. *Poultry Science*, *100*(2), 685–696. https://doi.org/10.1016/j.psj.2020.11.026

Supplementary material

Tables from supplementary material can be found in Annex 4.

Reduced metabolic capacity of the gut microbiota associates with host growth in broiler chickens

Chapter 4

Abstract

Understanding the development of functional attributes of host-associated microbial communities is essential for developing novel microbe-based solutions for sustainable animal production. We applied multi-omics to 388 broiler chicken caecal samples to characterise and model the functional dynamics of 822 bacterial strains. Although microbial community diversity metrics increased with chicken age as expected, the overall metabolic capacity and activity of the microbiota exhibited an unexpected decrease. This drop occurred due to the spread of non-culturable clades with small genomes and low metabolic capacities, including RF39, RF32, and UBA1242. The intensity of this decrease was associated with animal growth, whereby chickens with higher abundances of low-capacity bacteria exhibited higher body weights. This previously unreported link between metabolic capacity of microbes and animal body weight suggests a relevant role of non-culturable bacteria with reduced-genomes for host biology, and opens new avenues in the search for microbe-based solutions to improve sustainability of animal production.

Publication

Sofia Marcos[1,2], Iñaki Odriozola[2], Ostaizka Aizpurua[2], Raphael Eisenhofer[2], Sarah Siu Tze Mak[2], Garazi Martin[2], Varsha Kale[3], Germana Baldi[3], Robert D Finn[3], Joan Tarradas[4], Andone Estonba[1], M Thomas P Gilbert[2,5], Antton Alberdi[2]

[1]Applied Genomics and Bioinformatics, University of the Basque Country (UPV/EHU), Leioa, Bilbao, Spain
[2]Center for Evolutionary Hologenomics, Globe Institute, University of Copenhagen (UCPH), Copenhagen, Denmark
[3]European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Hinxton, Cambridge, UK
[4]Animal Nutrition, Institute of Agrifood Research and Technology (IRTA), Constantí, Spain
[5]University Museum, NTNU, Trondheim, Norway

Publication status: Under review
Link to preprint: https://www.researchsquare.com/article/rs-2885808/v1

Introduction

Optimising food production is one of the global priorities to ameliorate the ongoing sustainability crisis (Godfray et al., 2010). Gut microbial communities are recognised as relevant assets in the endeavour of finding more sustainable animal farming practices (D'Hondt et al., 2021), especially in intensive production systems that face increasing restrictions over historically employed growth additives, such as antibiotics (Castanon, 2007; Laxminarayan et al., 2015). Hence, the animal farming sector is seeking sustainable microbe-based solutions, such as probiotics and synbiotics, to maximise productivity while minimising environmental impact (Anee et al., 2021).

However, the task of finding optimal microbe-based solutions is challenged by the extreme complexity of animal gut microbiota (Alberdi et al., 2022), as they comprise hundreds of bacteria that vary throughout the development of the animal. Although some microbial taxa have been associated with animal growth performance (Wen et al., 2021; Yan et al., 2017), very few findings have been reproduced in independent trials. In consequence, the beneficial effects of existing microbe-based solutions are modest, and success rates vary across environments (Barba-Vidal et al., 2019; Zommiti et al., 2020). A key roadblock is the limited capacity of conventional methods to uncover how functional features of microbial communities develop alongside animals. While culture-based approaches only enable studying a limited number of functions in a fraction of the relevant bacteria outside their native gut environments (Barberán et al., 2017), targeted sequencing-based approaches do not enable direct functional inferences (Antony-Babu et al., 2017).

In contrast, multi-omic technologies provide greater functional breadth and resolution to understand how gut microbial communities develop and associate with animal growth (Nyholm et al., 2020). For instance, distilling the wealth of functional annotations retrieved from nearly complete bacterial genomes into quantitative traits (Escalas et al., 2019) enables the temporal dynamics of microbial functional capacities to be modelled, while metatranscriptomic gene expression data allows quantifying the realisation of such capabilities. Therefore, the combination of these two approaches can provide an unparalleled level of resolution into what microbial communities both could, and *actually,* do (Alberdi et al., 2022). Such data can be used to address hitherto unaddressable questions, including how functional capacities of gut microbial communities change over time, the degree to which those capacities are realised, and whether this functional variation is associated with animal growth performance.

One of the animal production systems that could benefit from such a deep understanding of microbiota development is poultry, namely the main source of animal meat consumed by humans worldwide (FAO, 2012). Although the gut microbiota of broiler chickens grown for meat production has been intensively studied using targeted (Rychlik, 2020) and shotgun sequencing technologies (Feng et al., 2021; Gilroy et al., 2021; Glendinning et al., 2020; Segura-Wang et al., 2021; Y. Zhang et al., 2022), multi-omic approaches that study whether and how microbial functions change throughout chicken development, and their association

with animal body weight are almost nonexistent (Jing et al., 2021). In light of this, we combined genome-resolved metagenomics with metatranscriptomics to study the complexity and dynamics of the entire functional landscape of the caecum microbiota in two replicate trials conducted in the HoloFood (Tous et al., 2022) project. We first aimed at understanding the relationship between the development of functional features and diversity metrics of the microbiota, for then unveiling their link with animal growth performance.

Results

Distillation of gene annotations reveals the functional complexity of the chicken gut microbiota

We studied the functional temporal dynamics of the broiler chicken gut microbiota using a novel approximation that leveraged the caecum genome catalogue of 822 bacterial strains generated in the HoloFood project (Rogers, 2023). After annotating all genomes against Pfam, KEGG, UniProt, CAZY and MEROPS databases, we distilled the resulting >1.6 million gene annotations into Genome-Inferred Functional Traits (GIFTs) using distillR (Koziol et al., 2023). GIFTs are quantitative metrics that estimate functional capabilities of bacteria from genomic information, which are calculated for each function in each genome as the proportion of biochemical reactions enabled by the genes present in a genome to accomplish a metabolic function (Eisenhofer et al., 2023). These GIFTs relate to synthesis and degradation of biomolecular compounds known to be relevant for animal and microbial biology; including polysaccharides, proteins, aromatic compounds, xenobiotics and antibiotics, among others. We measured 170 GIFTs per genome (complete detailed list can be found in Table S1), whose values were then averaged to obtain a genome-level overall metabolic capacity metric, hereafter referred to as Metabolic Capacity Index (MCI).

The distillation of metabolic attributes of bacterial genomes highlighted the importance of strain-level metabolic characterisation to understand functional dynamics of microbial communities. This is because our analyses yielded a complex functional landscape (Fig. S1) in which some taxa formed either a single (e.g., Cyanobacteria, Verrucomicrobiota) or a few functional clusters (e.g., Bacteroidota), while others displayed much wider spectra of functional profiles (e.g., Firmicutes, Proteobacteria) (Fig. 1a). The functional ordination showed a clear gradient of increasing metabolic capacity from top to bottom, but with a very different arrangement of genomes within and between phyla (Fig. 1b). For instance, Lachnospiraceae (Firmicutes A) genomes layed on the bottom among the genomes with the highest metabolic capacities, whereas other members of the same phylum exhibited mid- (e.g., TANB77) or low-level (e.g., UBA1242) metabolic capacities. Similarly, Proteobacteria contained both high-capacity (e.g., Enterobacteriaceae) and low-capacity (e.g., RF32) genomes. In contrast, Bacilli (Firmicutes) only contained mid- (e.g., Lactobacillales) and low-capacity (e.g., RF39) genomes. All these clades with low metabolic capacity (UBA1242, RF32, RF39) comprise unculturable strains discovered through metagenomic techniques, which have been recently detected in multiple animal gut environments (Crossfield et al., 2022; Gilroy et al., 2021, 2022; Pérez-Brocal et al., 2013; X.-X. Zhang et al., 2022), including

humans (Bowerman et al., 2020; Humbel et al., 2020). Despite their distant phylogenetic relationships, they all are characterised by reduced genome sizes and gene catalogues, as well as numerous auxotrophies (Nayfach et al., 2019; Wang et al., 2020). Nevertheless, these largely unknown bacteria embed within traditional taxonomic groups (e.g., Bacilli, Clostridia) or are closely related to them. An extreme example can be found between *Escherichia* (>5000 protein-coding genes) and RF32 (<2000 genes) Proteobacteria strains, which lay next to each other in the phylogeny yet hold highly different functional repertoires (Fig. 2a). Such a variety of functional attributes within major bacterial clades flags the risk of oversimplifying bacterial functional complexity when analysing gut microbiota features in terms of phylum-level ratios, as commonly done with the Firmicutes/Bacteroidota ratio (Elokil et al., 2022; Magne et al., 2020).

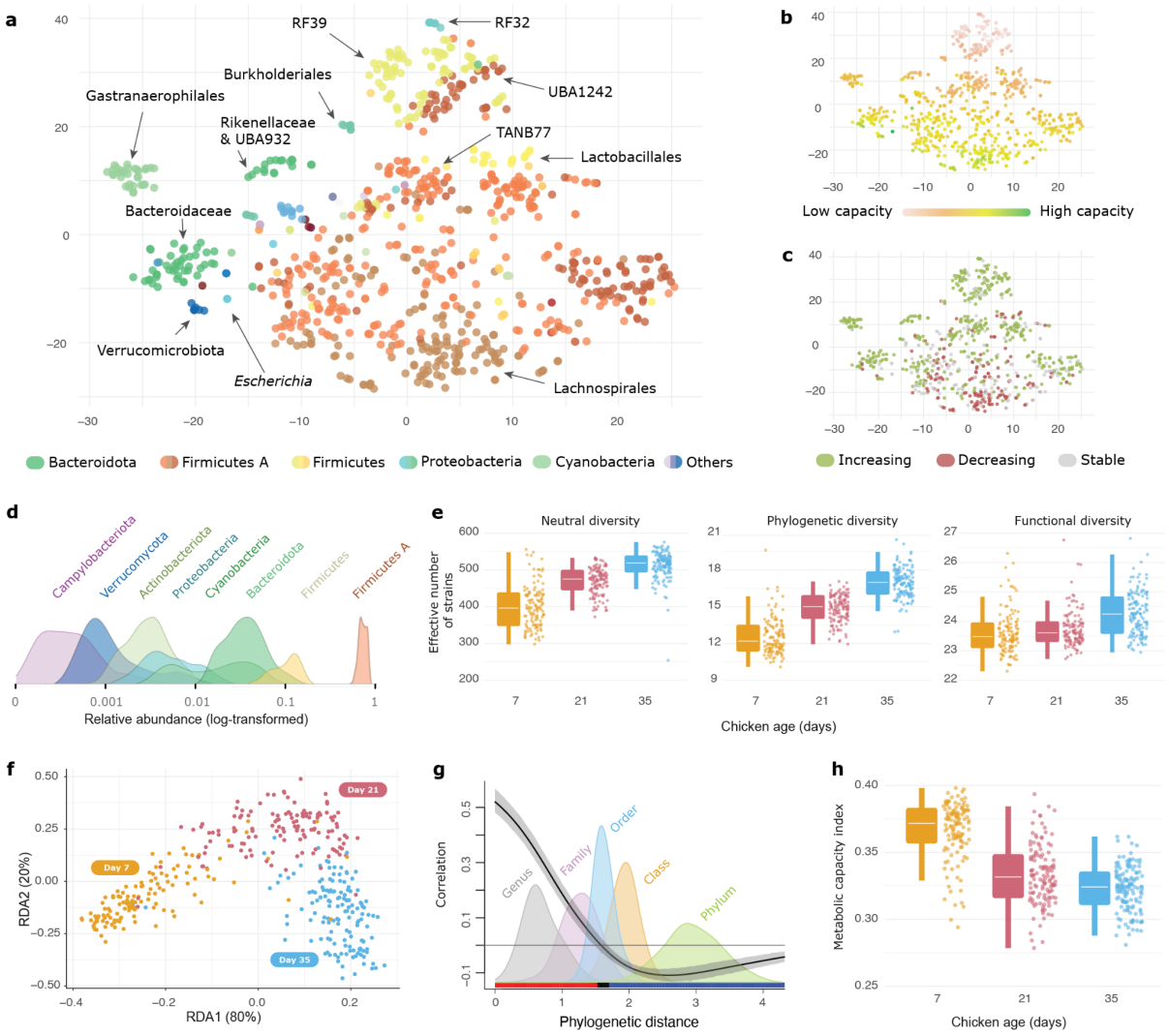


Figure 1. Functional ordination, diversity and temporal dynamics of the chicken caecum microbiota. **a)** t-SNE ordination of bacterial genomes based on their genome-inferred functional traits (GIFTs), coloured by taxonomic order. **b)** Identical ordination coloured by the overall metabolic capacity of each genome, in which the gradient of increasing metabolic capacity can be observed from top to bottom. **c)** Same ordination coloured by the temporal trend displayed by each bacteria in the study. **d)** Density curves of

the overall relative abundances of the main bacterial phyla. Note the x-scale is log-transformed to better visualise the least abundant taxa. **e)** Neutral, phylogenetic and functional diversity metrics across time. **f)** Distance-based redundancy analysis (dbRDA) plot showing the temporal dynamics of the compositional microbiota variation. **g)** Phylogenetic correlogram showing the correlation of temporal dynamics across different phylogenetic distances, expressed over density curves of the phylogenetic distances between the main taxonomic levels. **h)** Community-level metabolic capacity index values across time.

Microbiota diversity increases while functional capacity decreases with age

We then analysed microbial communities of 7, 21 and 35-day old chickens from two independent trials, by mapping 388 metagenomic and 61 metatranscriptomic datasets generated from caecum content samples to the bacterial genome catalogue. Although the conventional microbiota community and diversity analyses confirmed previously described trends and patterns, our detailed functional analyses yielded unexpected results. Surprisingly, we found that the functional features of the microbiota were highly contrasting to the community diversity metrics.

The taxonomic profile of the reconstructed microbiota matched the typical community previously described in broiler chickens (Rychlik, 2020; Segura-Wang et al., 2021), overwhelmingly dominated by Firmicutes A (Clostridia), followed by Firmicutes (Bacilli), Bacteroidota and Proteobacteria (Fig. 1d). All three alpha diversity metrics (neutral, phylogenetic and functional) increased with chicken age (Fig. 1e), and beta diversity metrics, as well as hierarchical models of species communities (Hmsc) (Tikhonov et al., 2020), showed that most of the microbiota variation occurred between time points, with animals from different trials and experimental groups exhibiting similar trajectories (Fig. 1f, Tables S2,S3, S4).

The highest turnover occurred between days 7 and 21, followed by a more modest variation in the following fortnight (Fig. S2). Bacterial dynamics during this entire period showed a very strong phylogenetic signal, both for metagenomics (ρ = 0.94 with 90% CI [0.93, 0.94]) and metatranscriptomics (ρ = 0.96 with 90% CI [0.95, 0.97]) (Fig. S3), with strains within taxonomic families exhibiting highly similar temporal dynamics compared to more distantly related ones (Fig. 1g). Beta diversity analyses, however, showed that the variation of functional attributes ($D^{\beta}_{d7\text{-}35}$ = 0.23±0.06) exceeded that of the phylogenetic change ($D^{\beta}_{d7\text{-}35}$ = 0.08±0.03). This mismatch is highly relevant, because phylogenetic distance is often used as a proxy for functional dissimilarity when direct information of functional attributes is not available (e.g., amplicon sequencing) (Alberdi & Gilbert, 2019). In contrast, our results showed that the turnover of phylogenetically similar bacteria implicated related strains yet with distinct functional attributes, which highlights the value of the direct functional inference capacity provided by genome-resolved metagenomics over approaches based on indirect evidence.

The bacteria that lost and gained representation between days 7 and 35, hereafter referred to as “decreasers” and “increasers”, respectively, were functionally structured (Fig. 1c). Most of the bacteria that became rarer as chickens grew belonged to Lachnospiraceae (55%), a clade characterised for its high metabolic capacity. In contrast, groups of bacteria with mid (e.g., TANB77, Cyanobacteria, Bacteroidota) to low metabolic capacities (e.g., UBA1242, RF39, RF32) tended to increase abundance. The mean metabolic capacity of bacteria whose abundance decreased over time (0.32±0.05) was higher than that of the bacteria that increased (0.24±0.08). As a consequence, and despite the observed increase in diversity, the community-weighted metabolic capacities of the entire microbiota decreased over time (Fig. 1h). This contrasting phenomenon is expected when a community initially dominated by bacteria with high metabolic capacities (yet functionally very similar to each other) transitions into a more diverse microbiota in which bacteria with low metabolic capacities (and thus functionally different to the previous ones) are also recruited. Hence, our results indicate that a more functionally diverse microbiota does not necessarily entail a more functionally capable community with a larger capacity to generate metabolic by-products that could benefit host health and growth.

The dominance of bacteria with high metabolic capacity during the early stages of microbiota development aligns with the concept of metabolic independence (Watson et al., 2023), because bacteria with higher metabolic capacity are less dependent on metabolic by-products produced by other microorganisms. However, as the microbiome matures, microorganisms with lower metabolic capacities, such as RF39, RF32, UBA1242, and TANB77, begin to gain representation at the expense of bacteria with larger metabolic repertoires, such as Lachnospiraceae (Fig. 2b). This process is probably exacerbated by the highly homogeneous dietary regime that industrially produced chickens experience (Futuyma & Moreno, 1988), as low-capacity resource specialists are more likely to outcompete generalist high-capacity bacteria under stable conditions (Watson et al., 2023). Given the remarkable ability of broiler chickens to efficiently absorb nutrients from their diet (Tallentire et al., 2016), low-capacity bacteria might also thrive better in an environment where the animal host absorbs most of the dietary energy, due to their lower energy requirements compared to high-capacity bacteria (Ranea et al., 2005). While the overall microbiota response to time measured through metagenomics and metatranscriptomics (considering all microbial genes) was highly correlated (Fig. 2d), the targeted functional pathways exhibited a different pattern. The microbial functions that tended to lose representation with time were compensated with increased expression, thus in part buffering the loss of capacity with increased activity (Fig. 2e).

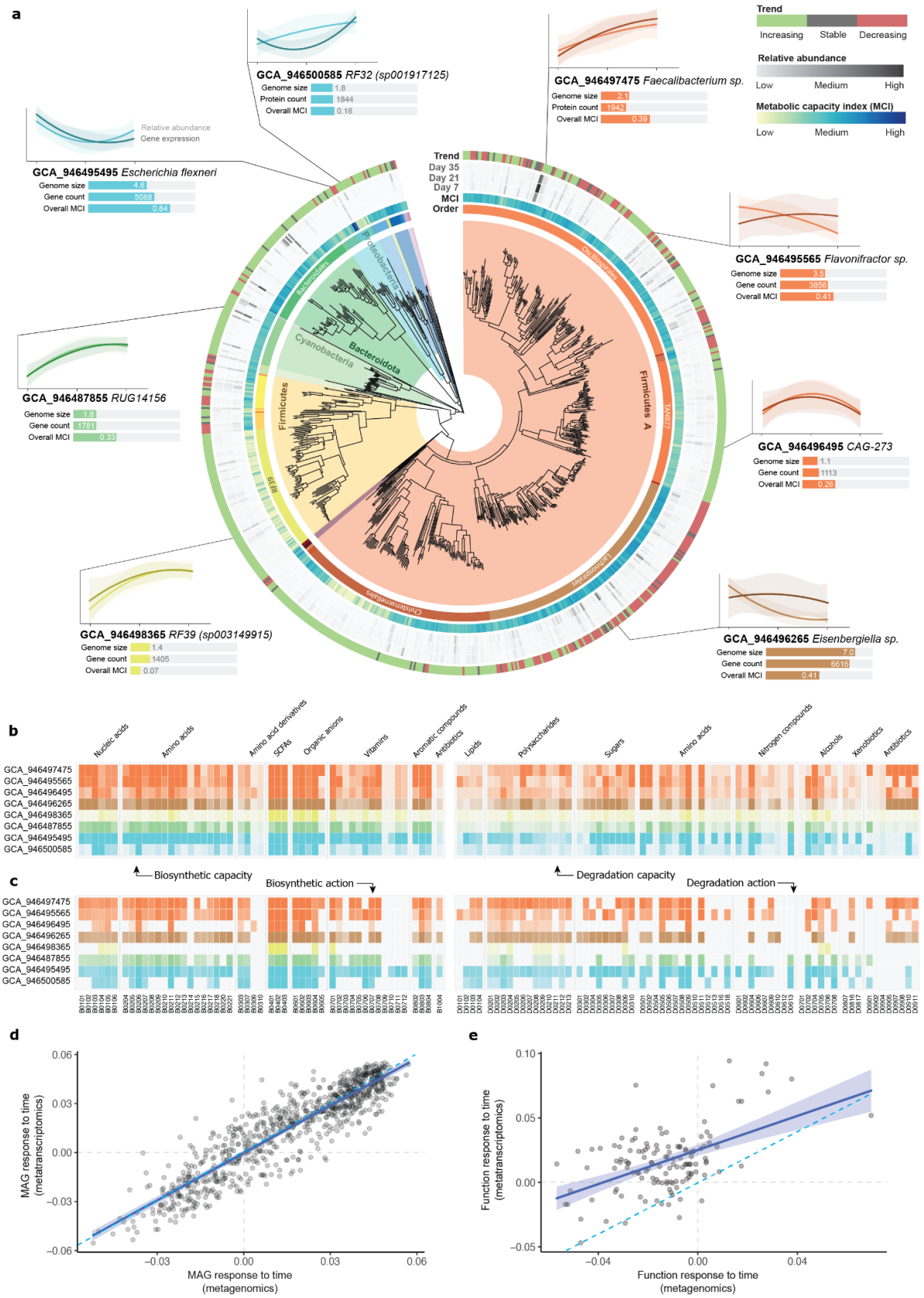


Figure 2. Temporal overview of metabolic capacities and gene expression of caecum bacteria. a) Phylogenetic tree, functional capacity, and temporal trends of the caecum microbiota. Strain boxes display the temporal trend of the relative abundance and gene expression for each strain along with overall genome features. **b)** Metabolic capacity indices

(MCI) to synthesise and degrade metabolic compounds of eight of the most relevant strains selected for illustrative purposes. **c)** Averaged and normalised expression of genes belonging to the pathways required to synthesise and degrade metabolic compounds in the eight highlighted strains. **d)** Correlation between the MAG response to time based on metagenomic (x-axis) and metatranscriptomic (y-axis) data. Each dot represents a MAG, and positive values indicate increasing abundance and gene expression trends. **e)** Correlation between the community-level response of each metabolic function to time based on metagenomic (metabolic capacity) and metatranscriptomic (metabolic activity) data. The dashed blue lines in figures **d** and **e** indicate the 1:1 regression lines.

One of the most distinctive features shared by all these bacteria with low metabolic capacities is that they lack genes involved in *de novo* synthesis of nucleic acids, which have raised speculations about their parasitic or obligate symbiotic relationship with vertebrates (Chklovski et al., 2022). Our metatranscriptomic analysis confirmed that many of the *increasing* bacteria did not synthesise their own nucleic acids (Fig. 2c). However, they all heavily expressed genes involved in SCFA production, particularly acetate, which are linked with beneficial outcomes for animal biology in general, and chickens in particular (Peng et al., 2021). These observations could therefore support the symbiosis hypothesis, whereby bacteria would receive nucleic acids and other essential molecular compounds from the host in return to metabolic benefits (Chklovski et al., 2022).

Reduction of functionally capable bacteria is associated with animal body weight

We finally explored the link between functional dynamics of the microbiota and animal growth performance, which for the first time revealed a surprising association between animal body weights and the proliferation of microorganisms with low metabolic capacity. The microbial turnover from a gut bacterial community with high metabolic capacities to a more mature one with lower capacities did not happen at the same rate in all studied chickens. Some animals retained a microbiota with more similar functional attributes to those exhibited in the beginning, while others underwent a more severe change (Fig. S2). This variation was found to be associated with animal body weight at day 35 (Generalised Linear Mixed Model, slope = 323.87, t-value = 2.97, *p*-value = 0.004) (Fig. 3), which could explain the large interindividual body weight variability we previously reported across animals with identical genetic line, sex and treatment properties (Tous et al., 2022). The robustness of the association was validated by additional analyses that revealed that the positive association held for chicken of both sexes and genetic lines (Fig. S4). Moreover, comparing the 5% chicken with lowest abundance of increaser bacteria with the 5% with highest content, showed a difference of 291 grams, which corresponds with 13% of the average body weight at day 35 (Fig. S5). The presence of individuals from every trial, sex, genetic line and treatment in both extreme groups further ensured that the association was not confounded by other chicken characteristics. The taxonomic drivers of the association were the low metabolic capacity clade RF39 (Bacilli, Firmicutes) and mid metabolic capacity clade TANB77 (Clostridia, Firmicutes A), whose high representation at day 35 was positively

correlated with body weight. In contrast, the abundance of high-level metabolic capacity clade Lachnospiraceae, correlated negatively with chicken body weight (Fig. S6).

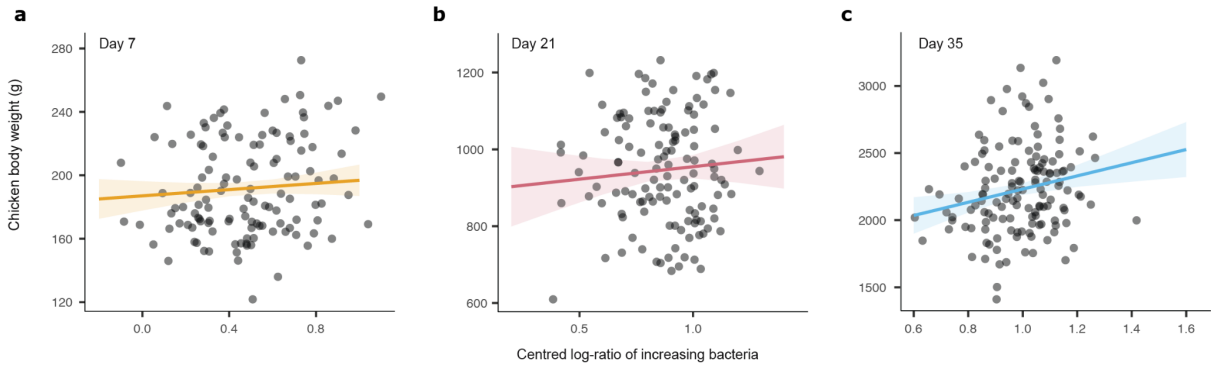


Figure 3. Correlation between representation-increasing bacteria in the caecum microbiota and chicken body weight. Linear regressions at days 7 (a), 21 (b) and 35 (c) between chicken body weight and centred log-ratio of the relative abundance of bacteria that increased their relative representation through time.

While the negative association between body weight and microbial functional capacity may seem counterintuitive, the observation is logical in the context of intensive animal production. Broiler chickens grown for meat production spend their short life (35-42 days) in controlled environments. Feed formulations have been optimised to enable rapid and efficient absorption of nutrients by chickens (Musigwa et al., 2021), while animals with the highest capacity to absorb and metabolise nutrients into skeletal tissue have been selected throughout hundreds of generations (Qanbari et al., 2019). Thus under these conditions, production yield might not be maximised by an intestinal microbiota with high metabolic capacity and activity, but rather one that minimises capture of dietary energy from chickens.

Biological (low microbial biomass) (Marcos et al., 2022) and technical (no enrichment of bacterial DNA) limitations prevented us from replicating the caecum analyses in the ileum, which together with the jejunum are the main sections for nutrient absorption in the gastrointestinal tract. However, the strength of the observed pattern in the caecum still indicates a biologically relevant association, since the caecum is the main fermentation chamber where SCFAs are produced by anaerobic bacteria, and absorbed by the host in the epithelial surface of the proximal caecum and in the colon (Svihus & Choct, 2013). In line with previous observations (Wang et al., 2020), our results indicate that SCFA production is one of the few metabolic functions that bacteria with low capabilities can conduct (Fig. 2e), which enhances their candidacy as beneficial microbes for animal production. Furthermore, microbiomes with average lower microbial capacities, which are more widespread than previously thought, seem to remain more stable over time (Starke et al., 2023), which could be an advantage against possible microbiota disruption or pathogen colonisation.

The association between functional capacities and animal body weight aligns with the mechanistic explanations behind the success of antibiotics as growth promoters. The beneficial effect of antimicrobial compounds in chicken production has been partly attributed to the depletion of the gut microbial community and the consequent maximisation of nutrients

and energy intake from food, which otherwise would be partly used by microorganisms for their own growth (Dibner & Richards, 2005). Although antibiotics have been broadly used in animal production to increase productivity since the 1950's (Castanon, 2007), overwhelming evidence points towards detrimental environmental outcomes related to the spread of antibiotic resistance (Rohr et al., 2019; Silbergeld et al., 2008). In consequence, increasing restrictions are being applied to the use of antibiotics as growth promoters. Although the correlative nature of our findings does not enable causal relationships to be established, we believe this previously unreported link between metabolic capacity of microbes and animal body weight deserves to be further explored, both to proof causality and to test whether microbiota manipulation to favour communities with low functional capabilities is a feasible strategy to be used as an alternative to antibiotics.

Discussion

Our study showcases the power of multi-omic techniques to surface complex patterns that associate production parameters with functional characteristics of the microbial communities. The diverse range of metabolic capacities we describe within each phylum, the different temporal dynamics we report and the mismatch between community diversity and metabolic capacity illustrates the complexity of gut microbiota. Our results highlight the risk of inferring functional outcomes from indirect information, such as 16S amplicon data or phylogenetic relationships between bacteria, and clearly demonstrate that reducing the complexity of the microbiota to phylum-level ratios (e.g., Firmicutes/Bacteroidota ratio) entails a massive oversimplification. In addition, the importance of uncultured taxa (e.g., UBA1242, RF39, RF32) surfaced by our analyses, along with their pervasiveness across animal gut environments, render them worthy of deeper studies to unravel their parallel evolutionary processes of genome shrinkage in light of interactions with hosts and other microorganisms.

As poultry is the main source of animal meat consumed by humans worldwide (FAO, 2012), and the system that is forecasted to undergo the largest production increase in the near future (OECD/FAO, 2021), the industry is immersed in this search for sustainable feed additives. However, we still lack a proper understanding of the functional development of chickens' gut microbiota. Our study provides important insights into the complex development of functional attributes, but also highlights the need for further analyses to understand the underlying reasons of the observed turnover of species and drop of metabolic capacities. The detected link between the growth performance and dynamics of strains with contrasting metabolic capacities opens new avenues in the search for microbe-based solutions to improve animal production. Overall, our study highlights the value of harnessing the power of omic technologies to surface complex patterns, which can inform academic and industrial actors to design and produce optimal feed additives to improve sustainability of animal production while ensuring animal welfare.

Methods

Animal experiments

Molecular and complementary data were generated from 388 animals sampled in the first two experiments performed in 2019 within the H2020 project HoloFood (HoloFood Consortium, 2019). The experiments were designed with a randomised pen design in which broilers from two genetic lines (Cobb 500 and Ross 308) and both sexes were grown for 35–37 days simulating intensive farming conditions. Each trial comprised 12 experimental groups (3 treatments x 2 genetic lines x 2 sex) and two replicates, for a total of 24 pens, each containing 40 animals. Details about the experimental design, procedures and performance results are provided in Tous et al. 2022 (Tous et al., 2022). A total of 6 chickens from each pen were euthanized, weighed, and sampled at days 7–8, 21–22 and 35–37 (multiple days were necessary due to workload, and these differences have been accounted for in the statistical analyses), hereafter simplified to three time points (days 7, 21 and 35). In short, one of the caecum pouches was isolated and longitudinally opened to collect two types of samples. The gut content was gently flushed out and ca. 100 mg of digesta collected for metagenomic and metatranscriptomic analyses. Both types of samples were preserved in DNA/RNA Shield buffer (Zymo Research, USA) and stored at -20 ºC until nucleic acid extraction. In total, 388 metagenomic and 61 metatranscriptomic samples were sequenced and analysed.

Data generation

DNA and RNA extraction

DNA and RNA extractions were conducted using a custom purification method (Bozzi et al., 2021) optimised for samples preserved on DNA/RNA Shield buffer. In short, this protocol included a physical bead-beating step for tissue disruption, followed by digestion, nucleic acid separation (DNA and RNA) and purification steps. Laboratory processing was conducted in batches of 90 samples, along with 6 negative controls, including two extraction, two library preparation and two library indexing blanks. Samples within each batch were randomised using a custom script, but gut content and mucosa samples were not mixed to minimise the risk of cross-contamination due to DNA concentration differences.

Library preparation of metagenomic DNA

Extracted nucleic acids were fragmented to an average length of 400 nucleotides using a Covaris LE220 ultrasonication device. The standard amount of DNA imputed to the library preparation was 200 ng. The BEST (Carøe et al., 2018) ligation-based library preparation protocol was employed for preparing sequencing libraries. The success of the library preparation process was assessed by qPCR assays, through which the optimal number of cycles to achieve the desired DNA molarity while reducing clonality was estimated. When the required number of cycles exceeded 12 cycles, the library preparation was repeated for considering that library-preparation inefficiency could yield technically biassed results. Libraries were subsequently indexed using unique dual tags and the required number of PCR cycles, and bead-purified before the final quality-check using a Fragment Analyser (Agilent)

with NGS fragment assay. Libraries with expected fragment-size distributions and molarities were equimolarly pooled for sequencing. Libraries with too low molarities were re-indexed and pooled to achieve the desired molarity. Libraries exhibiting unusual fragment distributions and large adaptor dimers were re-built to ensure maximum quality of generated data. Sequencing was performed at BGI (Shenzen, China) in multiple MGISeq-2000 runs with 150bp paired-end chemistry. Sequencing effort per sample typically varied between 8GB and 16GB (26 and 52 million reads).

Library preparation of metatranscriptomic RNA

For microbial metatranscriptomic analyses, rRNA depletion was performed using TIANSeq rRNA Depletion Kit (Cat.No. NR101-T1), after which cDNA conversion was carried out with random hexamers and Illumina short-read sequencing libraries were prepared using Novogene NGS RNA Library Prep Set (PT042). Library molarities were checked with Qubit and real-time PCR for quantification and bioanalyzer for size distribution detection. Quantified libraries were pooled and sequenced on an Illumina NovaSeq 6000 platform with 150bp paired-end chemistry, aiming for 5GB of protein-coding gene data.

Bioinformatic data processing

Generation of the MAG catalogue

A dedicated publication that will address the procedures employed for generating the MAG catalogue used in this study is under preparation, and the employed code is available at Workflowhub (https://workflowhub.eu/programmes/28). In short, data from 261 chicken gut metagenomic samples sequenced with MGISeq-2000 were used to generate the caecal MAG catalogue. *De novo* metagenomic assemblies were generated using the MGnify assembly pipeline (Richardson et al., 2023). The assembly tool MetaSPAdes (Nurk et al., 2017) was used preferentially for single-run assemblies, with MEGAHIT (D. Li et al., 2015) being used for co-assemblies where the memory requirements for MetaSPAdes were too high. Groups of samples prioritised for co-assembly were selected by hierarchical clustering based on Jaccard distance between low-quality bins generated by single assembly. Contigs shorter than 500 base pairs were excluded, and further host, human and PhiX decontamination was performed post-assembly with blastn (Y. Chen et al., 2015). Contig binning was performed with metaWRAP's binning and bin_refinement modules. Genome quality was assessed with checkM (Parks et al., 2015) to retain those genomes with completeness >50%, contamination <5%, and QS >50 (where QS = completeness - 5*contamination). Genomes were de-replicated using an Average Nucleotide Identity (ANI) of 95%, and 30% alignment fraction to generate species-level clusters using dRep (Olm et al., 2017). Similarly, a 99% ANI threshold was adopted for strain-level de-replication. GUNC (Orakov et al., 2021) was used to identify potentially chimeric genomes for removal, with the parameters clade separation score >0.45, contamination >0.05, and reference representation score >0.5. Taxonomic assignment was performed with GTDB-tk.

Functional annotation and distillation of MAG catalogue

Taxonomy annotation and phylogenetic tree preparation of the MAG catalogue was performed using GTDB-Tk (Chaumeil et al., 2019). The three archaea present in the MAG catalogue were excluded from the tree and derived analyses due to their near-negligible representation and technical limitations. The MAGs were functionally annotated by the ensemble approach implemented in DRAM (Shaffer et al., 2020), which includes Pfam (Mistry et al., 2021), KEGG (Kanehisa & Goto, 2000), UniProt (UniProt Consortium, 2019), CAZY (Cantarel et al., 2009) and MEROPS (Rawlings et al., 2010) databases. We then used the R package distillR (https://github.com/anttonalberdi/distillR) to transform raw annotations into quantitative genome-inferred functional traits (GIFTs). DistillR contains a set of >300 metabolic curated pathways and modules derived from KEGG (Kanehisa & Goto, 2000) and MetaCyc (Karp et al., 2002) databases, which are used to obtain quantitative estimates of the metabolic capacities of microorganisms through quantifying the relative representation of genes required for accomplishing a metabolic task. GIFTs range between 0-1, the zero indicating none of the genes defined in the pathway are present in the genome and one indicating that all genes are present. When a step within a pathway requires the presence of two Identifiers, the step is considered full if both Identifiers are present, half full if one is present and empty if none is present. We measured 170 GIFTs per genome (complete detailed list can be found in Table S1), whose values were first corrected by MAG genome completeness to reduce functional biases (Eisenhofer et al., 2023), and then averaged to obtain a genome-level overall metabolic capacity metric, hereafter referred to as Metabolic Capacity Index (MCI). Finally, to explore the landscape of the functional capabilities of the chicken caecum microbiome, the bacterial MAGs were ordinated based on their GIFTs through a t-SNE analysis using the R package Rtsne (Krijthe et al., 2017).

Metagenomic data processing and read mapping

Sequencing adapters and exact duplicates were removed using AdapterRemoval 2.2.4 (Schubert et al., 2016) and seqkit 0.7.1 (Shen et al., 2016). Sequences were mapped to the latest chicken reference genome (galGal6, NCBI Assembly accession GCF_000002315.6) using bwa (H. Li & Durbin, 2009) increasing the minimum seed length to 25 in order to reduce the number of incorrectly aligned read pairs from the metagenomic fraction. For assessing the quality of the alignment, mapping statistics including depth and breadth of coverage, and percentage of mapped reads were calculated using SAMtools 1.11 (H. Li & Durbin, 2009). Aligned reads were sorted and the metagenomic fraction was filtered out using SAMtools. Last, metagenomic reads were mapped to the MAG catalogue using bwa at 90% identity and 60% coverage threshold and further summarised with samtools. The mapping success of the caecum metagenomic reads against the bacterial catalogue was around 60-80% (Fig. S7), indicating that we obtained a nearly complete representation of the caecal bacterial community. Read-mapping counts resulting in <30% genome coverage per sample were removed from further analysis. Retained read-mapping counts were divided by the total number of paired-reads per sample, and multiplied by 100 to give the percentage of reads mapped to the MAG catalogue for each sample. Relative abundance was estimated by adapting the RPKG (Reads Per Kilobase per Genome equivalent) formula provided by Nayfach and Pollard. It is referred to as RPMM (Reads Per Million bases of genome, per

Million mapped reads), as reads mapped to MAGs were normalised both by genome length (divided by 1M) and by read length (divided by 1M).

Metatranscriptomic data processing

A custom snakemake (Mölder et al., 2021) pipeline was used for the metatranscriptomic data processing (https://github.com/anttonalberdi/holoflow/tree/EisenRa/workflows/metatranscriptomics). Reads were trimmed and quality controlled using fastp (S. Chen et al., 2018), keeping reads >60 bp and with Phred scores >20. Processed reads were then mapped against the host genome (galGal6) using STAR (Dobin et al., 2013). The unmapped reads were then mapped to a combined database containing the SILVA 16S rRNA SSU and LSU NR 99 (Quast et al., 2013), and the 5SRNAdb (Szymanski et al., 2016) using Bowtie2 (Szymanski et al., 2016) with default parameters. Unmapped reads were then mapped to the MAG catalogue genes (outputted from DRAM; genes.fna.gz) using Bowtie2 with default parameters. Finally, the gene read counts were calculated using CoverM (https://github.com/wwood/CoverM), requiring both pairs of reads to hit the gene (--proper-pairs-only flag).

Data analysis

Metagenomic data analysis

Metagenomic counts were standardised by MAG length and by sequencing depth. Alpha diversity measurements were calculated using neutral, phylogenetic and functional Hill numbers at the order of magnitude q=1, thus weighing MAGs according to their relative abundances, using the functions ‘hill_taxa’, ‘hill_phylo’ and ‘hill_func’, respectively, in the R package hillR (D. Li, 2018). Beta diversity (dissimilarity) metrics were generated for the same order of magnitude using the functions ‘hill_taxa_parti_pairwise’ (neutral), ‘hill_phylo_parti_pairwise’ (phylogenetic) and ‘hill_func_parti_pairwise’ (functional) of the same package. The phylogenetic tree employed to compute phylogenetic metrics was derived from the GTDB (Parks et al., 2022) tree constructed by GTDB-tk (Chaumeil et al., 2019) for taxonomic annotation, after pruning tips of reference genomes using ‘keep.tips’ function included in the R package ape (Paradis et al., 2004). The functional diversity analyses were based on a MAG trait matrix including pathway fullness values of 350 KEGG modules generated with DRAM (Shaffer et al., 2020). To visualise the composition of microbial communities across time, distance-based redundancy analysis (db-RDA) was performed using ‘vegdist’ and ‘rda’ commands from R package vegan (Oksanen et al., 2022).

To assess the temporal development of different components of microbial alpha and beta diversities we used linear mixed effect models (LMM) and PERMANOVAs, respectively. LMMs were fitted through the R package nlme *(Pinheiro et al., 2021)*, using the components of alpha diversity as response variables and trial (categorical variable with two levels), chicken age (numeric variable), sex (categorical variable with two levels), genetic line (categorical variable with two levels) and treatment (categorical variable with three levels) as fixed explanatory variables. A pen-level random intercept was included in the models to specify that chicken individuals were nested within pens. Log-transformed sequencing depth was also included as explanatory variable in all alpha diversity models to account for the

varying sequencing effort across samples. PERMANOVAs were fitted through the 'adonis2' function of R package vegan (Oksanen et al., 2022). Neutral, phylogenetic and functional dissimilarity matrices were included as responses in the PERMANOVAs and trial, chicken age, sex, genetic line and treatment were included as explanatory variables. Permutations were constrained to within pens to account for the nested nature of the data. Very weak effects of the categorical variables sex, genetic line and treatment were observed in the PERMANOVA and LMM analyses, and none of these variables showed significant interactions with chicken age. Therefore, they were included in the rest of the analyses of this study to account for their possible confounding effect, but their effect was not interpreted.

To identify the MAGs that increased and decreased with chicken age, MAG counts were analysed using the hierarchical modelling of species communities (Hmsc) framework (Tikhonov et al., 2020), which belongs to the class of joint species distribution modelling (JSDM) approach and builds a multivariate generalised linear mixed effect model using Bayesian inference. Raw counts of bacterial MAGs (weighted by the size of their genomes) were used as response variables in the model and trial, chicken age, sex, genetic line and treatment were included as fixed explanatory variables. Additionally, log-transformed library size was included as an extra explanatory variable to account for the compositionality of the data. To account for the nested study design a pen-level random effect was included in the model. The response variables were scaled to mean zero and unit variance and the log-normal model was applied. To assess whether the MAGs increased, decreased or remained stable over time the slope parameter linking the MAG abundance and chicken age was used with a posterior support of 0.95 as significance threshold. If >95% of the posterior distribution was positive the MAG was classified as an increaser, if >95% of the posterior distribution was negative the MAG was classified as a decreaser, and the rest of the MAGs were classified as stable. We fitted the models assuming the default priors and sampled the posterior distribution running four Markov Chain Monte Carlo (MCMC) chains, each of which was run for 3,750 iterations, of which 1,250 were discarded as burn-in. We thinned by 10 to obtain a total of 250 posterior samples per chain and 1000 posterior samples in total. We ensured MCMC convergence by measuring the potential scale reduction factor (Tikhonov et al., 2020) for the beta parameters (measuring the response of the MAGs to the fixed effects).

To examine whether the responses of the MAGs showed a phylogenetic signal to time, we included in the analysis a phylogenetic correlation matrix C among the MAGs computed from the above-mentioned phylogenetic tree. In Hmsc, the phylogenetic signal is measured using the parameter ρ, which takes values from 0 to 1, a value of 0 meaning no phylogenetic signal in the response to fixed effects, and a value of 1 meaning a completely phylogenetically structured response. Then, to measure the phylogenetic scale at which the response of the bacterial MAGs to chicken age was structured, we built a phylogenetic correlogram linking the MAGs' associations with time and the phylogenetic distance between MAGs using the R package phylosignal *(Keck et al., 2016)*. Finally, to compute the predictive power of the model we calculated the R^2 using two-fold cross-validation in two alternative ways. In the first case, $CV_{standard}$ hereafter, the samples were divided into the training and testing set randomly, hence the samples from both replicate trials A and B were used to train the model when

making the predictions. In the second case, CV_{trial} hereafter, the samples from experimental trials A and B were assigned separately to the training and testing sets, thus to make the predictions for trial A the model was only trained with samples from trial B, and, vice versa, to make the predictions for trial B the model was only trained with samples from trial A. Comparing R^2 from $CV_{standard}$ with R^2 from CV_{trial} allowed us evaluating whether the trends observed from our experimental trials were similar and thus generalizable between them. Both CV types yielded very similar R^2 values ($CV_{standard}$ = 0.54; CV_{trial} = 0.51) indicating that our results were consistent and generalizable between the two replicate trials. Cross-validation of the Hmsc model for metatranscriptomics yielded lower yet consistent R^2 values ($CV_{standard}$ = 0.15; CV_{trial} = 0.17), confirming the generalisability of our results between trials.

To assess the temporal development of community-level MCIs derived from metagenomics we fitted Generalised Linear Mixed Models (GLMM) using 'glmmTMB' function of glmmTMB package in R (Brooks et al., 2017). The proportion of biochemical reactions present in the genome to perform each specific function were used as response variables, while trial, chicken age, sex, genetic line and treatment were used as fixed explanatory variables. A pen-level random intercept was included in the models to specify that chicken individuals were nested within pens. Since the response variable was proportional (taking values between 0 and 1) the binomial distribution with logit link-function was used, and a sampling unit-level random intercept was included to account for under- or overdispersion and obtain robust standard errors (Harrison, 2015; Papke & Wooldridge, 1996).

Metatranscriptomic data analysis

Quantitative GIFTs were calculated with 'distillq' function using the R package distillR. To identify the expression of MAGs that increased and decreased with chicken age, MAG counts were analysed also using the Hmsc framework. Counts of MAG genes (weighted by the size of the genes) were used as response variables in the model while trial, chicken age, sex, genetic line and treatment were included as fixed explanatory variables. Then, the parameter estimation of the model, the evaluation of convergence and the evaluation of the model fit through cross-validations was conducted as explained in the Hmsc analysis of metagenomic data.

To assess the temporal development of community-level MCIs derived from metatranscriptomics we fitted LMMs using the 'lme' function of R package nlme (Pinheiro et al., 2021). We used as response variables the centred log-ratio (CLR) transformed community-level MCI expression values using the 'clr' function of R package compositions (van den Boogaart & Tolosana-Delgado, 2008). Trial, chicken age, sex, genetic line and treatment were used as fixed explanatory variables and a pen-level random intercept was included in the models to specify that chicken individuals were nested within pens.

Performance association

Associations between MAG abundances and chicken body weight were assessed using LMMs of R package nlme (Pinheiro et al., 2021). We wanted to test the null hypothesis of no

effect of the summed relative abundance of MAGs that increased and decreased over time in explaining body weight variability at each sampling day. Microbial temporal trends calculated with Hmsc grouped MAGs as increasers and decreasers. CLR of the relative abundance of increaser bacteria was calculated using the 'clr' function of R package compositions (van den Boogaart & Tolosana-Delgado, 2008). Then, LMMs were fitted using body weight as a response variable. Trial, age, genetic line, sex, treatment and the sum of the relative abundances of increasing or decreasing MAGs were added as fixed explanatory variables. A pen-level random intercept was included in the models to specify that chicken individuals were nested within pens.

Data availability

All raw DNA and RNA sequences, and the MAG catalogues are available under HoloFood's umbrella project on ENA (Project ID: PRJEB33223) and displayed in the HoloFood Data Portal (www.holofooddata.org). Bioinformatic resources including ENA accession numbers, scripts, data matrixes and files have been archived in Zenodo with the DOI: 10.5281/zenodo.8335509, as a release of the following Github repository: https://github.com/holochicken/func_dynamics.

Acknowledgments

This research was funded by the European Union's Horizon Research and Innovation Programme under grant agreement No. 817729 (HoloFood, Holistic solution to improve animal food production through deconstruct- ing the biomolecular interactions between feed, gut microorganisms, and animals in relation to performance parameters). The work of S.M. was supported by the Basque Government doctoral fellowship.

References

Alberdi, A., Andersen, S. B., Limborg, M. T., Dunn, R. R., & Gilbert, M. T. P. (2022). Disentangling host-microbiota complexity through hologenomics. *Nature Reviews. Genetics*, *23*(5), 281–297. https://doi.org/10.1038/s41576-021-00421-0

Alberdi, A., & Gilbert, M. T. P. (2019). A guide to the application of Hill numbers to DNA based diversity analyses. *Molecular Ecology Resources*, *19*(4), 804–817. https://doi.org/10.1111/1755-0998.13014

Anee, I. J., Alam, S., Begum, R. A., Shahjahan, R., & Khandaker, A. M. (2021). The role of probiotics on animal health and nutrition. *The Journal of Basic and Applied Zoology*, *82*(1), 52. https://doi.org/10.1186/s41936-021-00250-x

Antony-Babu, S., Stien, D., Eparvier, V., Parrot, D., Tomasi, S., & Suzuki, M. T. (2017). Multiple Streptomyces species with distinct secondary metabolomes have identical 16S rRNA gene sequences. *Scientific Reports*, *7*(1), 11089. https://doi.org/10.1038/s41598-017-11363-1

Barba-Vidal, E., Martín-Orúe, S. M., & Castillejos, L. (2019). Practical aspects of the use of probiotics in pig production: A review. *Livestock Science*, *223*, 84–96. https://doi.org/10.1016/j.livsci.2019.02.017

Barberán, A., Caceres Velazquez, H., Jones, S., & Fierer, N. (2017). Hiding in Plain Sight: Mining Bacterial Species Records for Phenotypic Trait Information. *mSphere*, *2*(4). https://doi.org/10.1128/mSphere.00237-17

Bowerman, K. L., Rehman, S. F., Vaughan, A., Lachner, N., Budden, K. F., Kim, R. Y., Wood, D. L. A., Gellatly, S. L., Shukla, S. D., Wood, L. G., Yang, I. A., Wark, P. A., Hugenholtz, P., & Hansbro, P.

M. (2020). Disease-associated gut microbiome and metabolome changes in patients with chronic obstructive pulmonary disease. *Nature Communications*, *11*(1), 5886. https://doi.org/10.1038/s41467-020-19701-0

Bozzi, D., Rasmussen, J. A., Carøe, C., Sveier, H., Nordøy, K., Gilbert, M. T. P., & Limborg, M. T. (2021). Salmon gut microbiota correlates with disease infection status: potential for monitoring health in farmed animals. *Animal Microbiome*, *3*(1), 30. https://doi.org/10.1186/s42523-021-00096-2

Brooks, M. E., Kristensen, K., Van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., Skaug, H. J., Machler, M., & Bolker, B. M. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R Journal*, *9*(2), 378–400. https://www.research-collection.ethz.ch/handle/20.500.11850/239870

Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., & Henrissat, B. (2009). The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Research*, *37*(Database issue), D233–D238. https://doi.org/10.1093/nar/gkn663

Carøe, C., Gopalakrishnan, S., Vinner, L., Mak, S. S. T., Sinding, M. H. S., Samaniego, J. A., Wales, N., Sicheritz-Pontén, T., & Gilbert, M. T. P. (2018). Single-tube library preparation for degraded DNA. *Methods in Ecology and Evolution / British Ecological Society*, *9*(2), 410–419. https://doi.org/10.1111/2041-210X.12871

Castanon, J. I. R. (2007). History of the use of antibiotic as growth promoters in European poultry feeds. *Poultry Science*, *86*(11), 2466–2471. https://doi.org/10.3382/ps.2007-00249

Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P., & Parks, D. H. (2019). GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* . https://doi.org/10.1093/bioinformatics/btz848

Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* , *34*(17), i884–i890. https://doi.org/10.1093/bioinformatics/bty560

Chen, Y., Ye, W., Zhang, Y., & Xu, Y. (2015). High speed BLASTN: an accelerated MegaBLAST search tool. *Nucleic Acids Research*, *43*(16), 7762–7768. https://doi.org/10.1093/nar/gkv784

Chklovski, A., Parks, D. H., Woodcroft, B. J., & Tyson, G. W. (2022). CheckM2: a rapid, scalable and accurate tool for assessing microbial genome quality using machine learning. In *bioRxiv* (p. 2022.07.11.499243). https://doi.org/10.1101/2022.07.11.499243

Crossfield, M., Gilroy, R., Ravi, A., Baker, D., La Ragione, R. M., & Pallen, M. J. (2022). Archaeal and Bacterial Metagenome-Assembled Genome Sequences Derived from Pig Feces. *Microbiology Resource Announcements*, *11*(1), e0114221. https://doi.org/10.1128/mra.01142-21

D'Hondt, K., Kostic, T., McDowell, R., Eudes, F., Singh, B. K., Sarkar, S., Markakis, M., Schelkle, B., Maguin, E., & Sessitsch, A. (2021). Microbiome innovations for a sustainable future. *Nature Microbiology*, *6*(2), 138–142. https://doi.org/10.1038/s41564-020-00857-w

Dibner, J. J., & Richards, J. D. (2005). Antibiotic growth promoters in agriculture: history and mode of action. *Poultry Science*, *84*(4), 634–643. https://doi.org/10.1093/ps/84.4.634

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* , *29*(1), 15–21. https://doi.org/10.1093/bioinformatics/bts635

Eisenhofer, R., Odriozola, I., & Alberdi, A. (2023). Impact of microbial genome completeness on metagenomic functional inference. *ISME Communications*, *3*(1), 12. https://doi.org/10.1038/s43705-023-00221-z

Elokil, A. A., Chen, W., Mahrose, P., Khalid, M., & Elattrouny, M. (2022). Early life microbiota transplantation from highly feed-efficient broiler improved weight gain by reshaping the gut microbiota in laying chicken. *Frontiers in Microbiology*.

Escalas, A., Hale, L., Voordeckers, J. W., Yang, Y., Firestone, M. K., Alvarez-Cohen, L., & Zhou, J. (2019). Microbial functional diversity: From concepts to applications. *Ecology and Evolution*, *9*(20), 12000–12016. https://doi.org/10.1002/ece3.5670

FAO. (2012). *FAOSTAT Database*. FAOSTAT Database. http://faostat.fao.org

Feng, Y., Wang, Y., Zhu, B., Gao, G. F., Guo, Y., & Hu, Y. (2021). Metagenome-assembled genomes and gene catalog from the chicken gut microbiome aid in deciphering antibiotic resistomes. *Communications Biology*, *4*(1), 1305. https://doi.org/10.1038/s42003-021-02827-2

Futuyma, D. J., & Moreno, G. (1988). The evolution of ecological specialization. *Annual Review of Ecology and Systematics*, *19*(1), 207–233. https://doi.org/10.1146/annurev.es.19.110188.001231

Gilroy, R., Leng, J., Ravi, A., Adriaenssens, E. M., Oren, A., Baker, D., La Ragione, R. M., Proudman, C., & Pallen, M. J. (2022). Metagenomic investigation of the equine faecal microbiome reveals extensive taxonomic diversity. *PeerJ*, *10*, e13084. https://doi.org/10.7717/peerj.13084

Gilroy, R., Ravi, A., Getino, M., Pursley, I., & Horton, D. L. (2021). Extensive microbial diversity within the chicken gut microbiome revealed by metagenomics and culture. *PeerJ*. https://www.ncbi.nlm.nih.gov/pmc/articles/pmc8035907/

Glendinning, L., Stewart, R. D., Pallen, M. J., Watson, K. A., & Watson, M. (2020). Assembly of hundreds of novel bacterial genomes from the chicken caecum. *Genome Biology*, *21*(1), 34. https://doi.org/10.1186/s13059-020-1947-1

Godfray, H. C. J., Beddington, J. R., Crute, I. R., Haddad, L., Lawrence, D., Muir, J. F., Pretty, J., Robinson, S., Thomas, S. M., & Toulmin, C. (2010). Food security: the challenge of feeding 9 billion people. *Science*, *327*(5967), 812–818. https://doi.org/10.1126/science.1185383

Harrison, X. A. (2015). A comparison of observation-level random effect and Beta-Binomial models for modelling overdispersion in Binomial data in ecology & evolution. *PeerJ*, *3*, e1114. https://doi.org/10.7717/peerj.1114

HoloFood Consortium. (2019, January 1). *Holistic solution to improve animal food production through deconstructing the biomolecular interactions between feed, gut microorganisms and animals in relation to performance parameters*. CORDIS. https://cordis.europa.eu/project/id/817729

Humbel, F., Rieder, J. H., Franc, Y., Juillerat, P., Scharl, M., Misselwitz, B., Schreiner, P., Begré, S., Rogler, G., von Känel, R., Yilmaz, B., Biedermann, L., & Swiss IBD Cohort Study Group. (2020). Association of Alterations in Intestinal Microbiota With Impaired Psychological Function in Patients With Inflammatory Bowel Diseases in Remission. *Clinical Gastroenterology and Hepatology: The Official Clinical Practice Journal of the American Gastroenterological Association*, *18*(9), 2019–2029.e11. https://doi.org/10.1016/j.cgh.2019.09.022

Jing, Y., Yuan, Y., Monson, M., Wang, P., Mu, F., Zhang, Q., Na, W., Zhang, K., Wang, Y., Leng, L., Li, Y., Luan, P., Wang, N., Guo, R., Lamont, S. J., Li, H., & Yuan, H. (2021). Multi-Omics Association Reveals the Effects of Intestinal Microbiome-Host Interactions on Fat Deposition in Broilers. *Frontiers in Microbiology*, *12*, 815538. https://doi.org/10.3389/fmicb.2021.815538

Kanehisa, M., & Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, *28*(1), 27–30. https://doi.org/10.1093/nar/28.1.27

Karp, P. D., Riley, M., Paley, S. M., & Pellegrini-Toole, A. (2002). The MetaCyc Database. *Nucleic Acids Research*, *30*(1), 59–61. https://doi.org/10.1093/nar/30.1.59

Keck, F., Rimet, F., Bouchez, A., & Franc, A. (2016). phylosignal: an R package to measure, test, and explore the phylogenetic signal. *Ecology and Evolution*, *6*(9), 2774–2780. https://doi.org/10.1002/ece3.2051

Koziol, A., Odriozola, I., Leonard, A., Eisenhofer, R., San Jose, C., Aizpurua, O., & Alberdi, A. (2023). Mammals show distinct functional gut microbiome dynamics to identical series of environmental stressors. *mBio*. https://doi.org/10.1128/mbio.01606-23

Krijthe, J., van der Maaten, L., & Krijthe, M. J. (2017). Package “Rtsne.” *R Package Version 0. 13*. http://www.vps.fmvz.usp.br/CRAN/web/packages/Rtsne/Rtsne.pdf

Laxminarayan, R., Van Boeckel, T., & Teillant, A. (2015). *The economic costs of withdrawing antimicrobial growth promoters from the livestock sector* (OECD Food, Agriculture and Fisheries Papers). Organisation for Economic Co-Operation and Development (OECD). https://doi.org/10.1787/5js64kst5wvl-en

Li, D. (2018). hillR: taxonomic, functional, and phylogenetic diversity and similarity through Hill Numbers. *Journal of Open Source Software*, *3*(31), 1041. https://doi.org/10.21105/joss.01041

Li, D., Liu, C.-M., Luo, R., Sadakane, K., & Lam, T.-W. (2015). MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* , *31*(10), 1674–1676. https://doi.org/10.1093/bioinformatics/btv033

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* , *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Magne, F., Gotteland, M., Gauthier, L., Zazueta, A., Pesoa, S., Navarrete, P., & Balamurugan, R.

(2020). The Firmicutes/Bacteroidetes Ratio: A Relevant Marker of Gut Dysbiosis in Obese Patients? *Nutrients*, *12*(5). https://doi.org/10.3390/nu12051474

Marcos, S., Parejo, M., Estonba, A., & Alberdi, A. (2022). Recovering high‐quality host genomes from gut metagenomic data through genotype imputation. *Advanced Genetics*, *3*(3), 2100065. https://doi.org/10.1002/ggn2.202100065

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L. L., Tosatto, S. C. E., Paladin, L., Raj, S., Richardson, L. J., Finn, R. D., & Bateman, A. (2021). Pfam: The protein families database in 2021. *Nucleic Acids Research*, *49*(D1), D412–D419. https://doi.org/10.1093/nar/gkaa913

Mölder, F., Jablonski, K. P., Letcher, B., Hall, M. B., Tomkins-Tinch, C. H., Sochat, V., Forster, J., Lee, S., Twardziok, S. O., Kanitz, A., Wilm, A., Holtgrewe, M., Rahmann, S., Nahnsen, S., & Köster, J. (2021). Sustainable data analysis with Snakemake. *F1000Research*, *10*, 33. https://doi.org/10.12688/f1000research.29032.2

Musigwa, S., Morgan, N., Swick, R., Cozannet, P., & Wu, S.-B. (2021). Optimisation of dietary energy utilisation for poultry – a literature review. *World's Poultry Science Journal*, *77*(1), 5–27. https://doi.org/10.1080/00439339.2020.1865117

Nayfach, S., Shi, Z. J., Seshadri, R., Pollard, K. S., & Kyrpides, N. C. (2019). New insights from uncultivated genomes of the global human gut microbiome. *Nature*, *568*(7753), 505–510. https://doi.org/10.1038/s41586-019-1058-x

Nurk, S., Meleshko, D., & Korobeynikov, A. (2017). metaSPAdes: a new versatile metagenomic assembler. *Genome / National Research Council Canada = Genome / Conseil National de Recherches Canada*. https://genome.cshlp.org/content/27/5/824.short

Nyholm, L., Koziol, A., Marcos, S., Botnen, A. B., Aizpurua, O., Gopalakrishnan, S., Limborg, M. T., Gilbert, M. T. P., & Alberdi, A. (2020). Holo-Omics: Integrated Host-Microbiota Multi-omics for Basic and Applied Biological Research. *iScience*, *23*(8), 101414. https://doi.org/10.1016/j.isci.2020.101414

OECD/FAO. (2021). *OECD-FAO Agricultural Outlook 2021-2030*. OECD Publishing.

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., & Others. (2022). *vegan: Community Ecology Package. R package version 2.5--7. 2020*.

Olm, M. R., Brown, C. T., Brooks, B., & Banfield, J. F. (2017). dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *The ISME Journal*, *11*(12), 2864–2868. https://doi.org/10.1038/ismej.2017.126

Orakov, A., Fullam, A., Coelho, L. P., Khedkar, S., Szklarczyk, D., Mende, D. R., Schmidt, T. S. B., & Bork, P. (2021). GUNC: detection of chimerism and contamination in prokaryotic genomes. *Genome Biology*, *22*(1), 178. https://doi.org/10.1186/s13059-021-02393-0

Papke, L. E., & Wooldridge, J. M. (1996). Econometric methods for fractional response variables with an application to 401(k) plan participation rates. *Journal of Applied Economics*, *11*(6), 619–632. https://doi.org/10.1002/(sici)1099-1255(199611)11:6<619::aid-jae418>3.0.co;2-1

Paradis, E., Claude, J., & Strimmer, K. (2004). APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* , *20*(2), 289–290. https://www.ncbi.nlm.nih.gov/pubmed/14734327

Parks, D. H., Chuvochina, M., Rinke, C., Mussig, A. J., Chaumeil, P.-A., & Hugenholtz, P. (2022). GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Research*, *50*(D1), D785–D794. https://doi.org/10.1093/nar/gkab776

Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., & Tyson, G. W. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research*, *25*(7), 1043–1055. https://doi.org/10.1101/gr.186072.114

Peng, L.-Y., Shi, H.-T., Gong, Z.-X., Yi, P.-F., Tang, B., Shen, H.-Q., & Fu, B.-D. (2021). Protective effects of gut microbiota and gut microbiota-derived acetate on chicken colibacillosis induced by avian pathogenic Escherichia coli. *Veterinary Microbiology*, *261*, 109187. https://doi.org/10.1016/j.vetmic.2021.109187

Pérez-Brocal, V., García-López, R., Vázquez-Castellanos, J. F., Nos, P., Beltrán, B., Latorre, A., & Moya, A. (2013). Study of the viral and microbial communities associated with Crohn's disease: a

metagenomic approach. *Clinical and Translational Gastroenterology*, *4*(6), e36. https://doi.org/10.1038/ctg.2013.9

Pinheiro, J., Bates, D., DebRoy, S., & Sarkar, D. (2021). *R Core Team (2021). nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-152*.

Qanbari, S., Rubin, C.-J., Maqbool, K., Weigend, S., Weigend, A., Geibel, J., Kerje, S., Wurmser, C., Peterson, A. T., Brisbin, I. L., Jr, Preisinger, R., Fries, R., Simianer, H., & Andersson, L. (2019). Genetics of adaptation in modern chicken. *PLoS Genetics*, *15*(4), e1007989. https://doi.org/10.1371/journal.pgen.1007989

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., & Glöckner, F. O. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research*, *41*(Database issue), D590–D596. https://doi.org/10.1093/nar/gks1219

Ranea, J. A. G., Grant, A., Thornton, J. M., & Orengo, C. A. (2005). Microeconomic principles explain an optimal genome size in bacteria. *Trends in Genetics: TIG*, *21*(1), 21–25. https://doi.org/10.1016/j.tig.2004.11.014

Rawlings, N. D., Barrett, A. J., & Bateman, A. (2010). MEROPS: the peptidase database. *Nucleic Acids Research*, *38*(Database issue), D227–D233. https://doi.org/10.1093/nar/gkp971

Richardson, L., Allen, B., Baldi, G., Beracochea, M., Bileschi, M. L., Burdett, T., Burgin, J., Caballero-Pérez, J., Cochrane, G., Colwell, L. J., Curtis, T., Escobar-Zepeda, A., Gurbich, T. A., Kale, V., Korobeynikov, A., Raj, S., Rogers, A. B., Sakharova, E., Sanchez, S., … Finn, R. D. (2023). MGnify: the microbiome sequence data analysis resource in 2023. *Nucleic Acids Research*, *51*(D1), D753–D759. https://doi.org/10.1093/nar/gkac1080

Rogers, S. (2023, January 4). *HoloFood Data Portal*. HoloFood Data Portal. www.holofooddata.org

Rohr, J. R., Barrett, C. B., Civitello, D. J., Craft, M. E., Delius, B., DeLeo, G. A., Hudson, P. J., Jouanard, N., Nguyen, K. H., Ostfeld, R. S., Remais, J. V., Riveau, G., Sokolow, S. H., & Tilman, D. (2019). Emerging human infectious diseases and the links to global food production. *Nature Sustainability*, *2*(6), 445–456. https://doi.org/10.1038/s41893-019-0293-3

Rychlik, I. (2020). Composition and Function of Chicken Gut Microbiota. *Animals : An Open Access Journal from MDPI*, *10*(1). https://doi.org/10.3390/ani10010103

Schubert, M., Lindgreen, S., & Orlando, L. (2016). AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Research Notes*, *9*, 88. https://doi.org/10.1186/s13104-016-1900-2

Segura-Wang, M., Grabner, N., Koestelbauer, A., Klose, V., & Ghanbari, M. (2021). Genome-Resolved Metagenomics of the Chicken Gut Microbiome. *Frontiers in Microbiology*, *12*, 726923. https://doi.org/10.3389/fmicb.2021.726923

Shaffer, M., Borton, M. A., McGivern, B. B., Zayed, A. A., La Rosa, S. L., Solden, L. M., Liu, P., Narrowe, A. B., Rodríguez-Ramos, J., Bolduc, B., Gazitúa, M. C., Daly, R. A., Smith, G. J., Vik, D. R., Pope, P. B., Sullivan, M. B., Roux, S., & Wrighton, K. C. (2020). DRAM for distilling microbial metabolism to automate the curation of microbiome function. *Nucleic Acids Research*, *48*(16), 8883–8900. https://doi.org/10.1093/nar/gkaa621

Shen, W., Le, S., Li, Y., & Hu, F. (2016). SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. *PloS One*, *11*(10), e0163962. https://doi.org/10.1371/journal.pone.0163962

Silbergeld, E. K., Graham, J., & Price, L. B. (2008). Industrial food animal production, antimicrobial resistance, and human health. *Annual Review of Public Health*, *29*, 151–169. https://doi.org/10.1146/annurev.publhealth.29.020907.090904

Starke, S., Harris, D. M. M., Zimmermann, J., Schuchardt, S., Oumari, M., Frank, D., Bang, C., Rosenstiel, P., Schreiber, S., Frey, N., Franke, A., Aden, K., & Waschina, S. (2023). Amino acid auxotrophies in human gut bacteria are linked to higher microbiome diversity and long-term stability. In *bioRxiv* (p. 2023.03.23.532984). https://doi.org/10.1101/2023.03.23.532984

Svihus, B., & Choct, M. (2013). Function and nutritional roles of the avian caeca: a review. *World's Poultry Science Journal*, *69*(2), 249–264. https://doi.org/10.1017/S0043933913000287

Szymanski, M., Zielezinski, A., Barciszewski, J., Erdmann, V. A., & Karlowski, W. M. (2016). 5SRNAdb: an information resource for 5S ribosomal RNAs. *Nucleic Acids Research*, *44*(D1), D180–D183. https://doi.org/10.1093/nar/gkv1081

Tallentire, C. W., Leinonen, I., & Kyriazakis, I. (2016). Breeding for efficiency in the broiler chicken: A review. *Agronomy for Sustainable Development*, *36*(4), 66. https://doi.org/10.1007/s13593-016-0398-2

Tikhonov, G., Opedal, Ø. H., Abrego, N., Lehikoinen, A., de Jonge, M. M. J., Oksanen, J., & Ovaskainen, O. (2020). Joint species distribution modelling with the r-package Hmsc. *Methods in Ecology and Evolution / British Ecological Society*, *11*(3), 442–447. https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.13345

Tous, N., Marcos, S., Boroojeni, F., Pérez de Rozas, A., Zentek, J., Estonba, A., Sandvang, D., Gilbert, M. T. P., Esteve-Garcia, E., Finn, R., Alberdi, A., & Tarradas, J. (2022). Novel Strategies to Improve Chicken Performance and Welfare by Unveiling Host-Microbiota Interactions through Hologenomics. *Frontiers in Physiology*. https://doi.org/10.3389/fphys.2022.884925

UniProt Consortium. (2019). UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Research*, *47*(D1), D506–D515. https://doi.org/10.1093/nar/gky1049

van den Boogaart, K. G., & Tolosana-Delgado, R. (2008). “compositions”: A unified R package to analyze compositional data. *Computers & Geosciences*, *34*(4), 320–338. https://doi.org/10.1016/j.cageo.2006.11.017

Wang, Y., Huang, J.-M., Zhou, Y.-L., Almeida, A., Finn, R. D., Danchin, A., & He, L.-S. (2020). Phylogenomics of expanding uncultured environmental Tenericutes provides insights into their pathogenicity and evolutionary relationship with Bacilli. *BMC Genomics*, *21*(1), 408. https://doi.org/10.1186/s12864-020-06807-4

Watson, A. R., Füssel, J., Veseli, I., DeLongchamp, J. Z., Silva, M., Trigodet, F., Lolans, K., Shaiber, A., Fogarty, E., Runde, J. M., Quince, C., Yu, M. K., Söylev, A., Morrison, H. G., Lee, S. T. M., Kao, D., Rubin, D. T., Jabri, B., Louie, T., & Eren, A. M. (2023). Metabolic independence drives gut microbial colonization and resilience in health and disease. *Genome Biology*, *24*(1), 78. https://doi.org/10.1186/s13059-023-02924-x

Wen, C., Yan, W., Mai, C., Duan, Z., Zheng, J., Sun, C., & Yang, N. (2021). Joint contributions of the gut microbiota and host genetics to feed efficiency in chickens. *Microbiome*, *9*(1), 126. https://doi.org/10.1186/s40168-021-01040-x

Yan, W., Sun, C., Yuan, J., & Yang, N. (2017). Gut metagenomic analysis reveals prominent roles of Lactobacillus and cecal microbiota in chicken feed efficiency. *Scientific Reports*, *7*, 45308. https://doi.org/10.1038/srep45308

Zhang, X.-X., Lv, Q.-B., Yan, Q.-L., Zhang, Y., Guo, R.-C., Meng, J.-X., Ma, H., Qin, S.-Y., Zhu, Q.-H., Li, C.-Q., Liu, R., Liu, G., Li, S.-H., Sun, D.-B., & Ni, H.-B. (2022). A Catalog of over 5,000 Metagenome-Assembled Microbial Genomes from the Caprinae Gut Microbiota. *Microbiology Spectrum*, e0221122. https://doi.org/10.1128/spectrum.02211-22

Zhang, Y., Jiang, F., Yang, B., Wang, S., Wang, H., Wang, A., Xu, D., & Fan, W. (2022). Improved microbial genomes and gene catalog of the chicken gut from metagenomic sequencing of high-fidelity long reads. *GigaScience*, *11*. https://doi.org/10.1093/gigascience/giac116

Zommiti, M., Chikindas, M. L., & Ferchichi, M. (2020). Probiotics—Live Biotherapeutics: a Story of Success, Limitations, and Future Prospects—Not Only for Humans. *Probiotics and Antimicrobial Proteins*, *12*(3), 1266–1289. https://doi.org/10.1007/s12602-019-09570-5

Supplementary material

Index

Table S1:
Nomenclature, classification and definitions of metabolic pathways employed in the functional distillation analysis using the R package distillR.

Pathway	Element	Function	Definition
B010101	Inosinic acid (IMP)	Nucleic acid biosynthesis	K00764 (K01945,K11787,K11788,K13713) (K00601,K11175,K08289,K11787,K01492) (K01952,(K23269+K23264+K23265),(K23270+K23265)) (K01933,K11787,(K11788 (K01587,K11808,(K01589 K01588)))) (K01923,K01587,K13713) K01756 (K00602,(K01492,(K06863 K11176)))
B010201	Uridylic acid (UMP)	Nucleic acid biosynthesis	(K11540,((K11541 K01465),((K01954,(K01955+K01956)) ((K00609+K00610),K00608) K01465))) (K00226,K00254,K17828) (K13421,(K00762 K01591))
B010301	UDP/UTP	Nucleic acid biosynthesis	(K13800,K13809,K09903)
B010401	CDP/CTP	Nucleic acid biosynthesis	(K00940,K18533) K01937
B010501	ADP/ATP	Nucleic acid biosynthesis	K01939 K01756 (K00939,K18532,K18533,K00944) K00940
B010601	GDP/GTP	Nucleic acid biosynthesis	K00088 K01951 K00942 (K00940,K18533)
B020401	Serine	Amino acid biosynthesis	K00058 K00831 (K01079,K02203,K22305,K25528)
B020501	Threonine	Amino acid biosynthesis	(K00928,K12524,K12525,K12526) K00133 (K00003,K12524,K12525) (K00872,K02204,K02203) K01733
B020601	Cysteine	Amino acid biosynthesis	(K00640,K23304) (K01738,K13034,K17069)
B020602	Cysteine	Amino acid biosynthesis	(K01697,K10150) K01758
B020603	Cysteine	Amino acid biosynthesis	K00789 K17462 K01243 K07173 K17216 K17217
B020701	Methionine	Amino acid biosynthesis	(K00928,K12524,K12525) K00133 (K00003,K12524,K12525) (K00651,K00641) K01739 (K01760,K14155) (K00548,K24042,K00549)
B020801	Valine	Amino acid biosynthesis	(K01652+(K01653,K11258)) K00053 K01687 K00826
B020901	Isoleucine	Amino acid biosynthesis	(K01703+K01704) K00052
B020902	Isoleucine	Amino acid biosynthesis	(K17989,K01754) (K01652+(K01653,K11258)) K00053 K01687 K00826
B021001	Leucine	Amino acid biosynthesis	K01649 (K01702,(K01703+K01704)) K00052
B021101	Lysine	Amino acid biosynthesis	(K00928,K12524,K12525,K12526) K00133 K01714 K00215 K00674 (K00821,K14267) K01439 K01778 (K01586,K12526)
B021102	Lysine	Amino acid biosynthesis	K00928 K00133 K01714 K00215 K05822 K00841 K05823 K01778 K01586
B021103	Lysine	Amino acid biosynthesis	(K00928,K12524,K12525,K12526) K00133 K01714 K00215 K03340 (K01586,K12526)
B021104	Lysine	Amino acid biosynthesis	(K00928,K12524,K12525,K12526) K00133 K01714 K00215 K10206 K01778 (K01586,K12526)
B021105	Lysine	Amino acid biosynthesis	K01655 ((K17450 K01705),(K16792+K16793)) K05824

B021106	Lysine	Amino acid biosynthesis	K05827 K05828 K05829 K05830 K05831
B021201	Arginine	Amino acid biosynthesis	K00611 K01940 (K01755,K14681)
B021202	Arginine	Amino acid biosynthesis	K22478 K00145 K00821 K09065 K01438 K01940 K01755
B021301	Proline	Amino acid biosynthesis	((K00931 K00147),K12657) K00286
B021401	Glutamate	Amino acid biosynthesis	K00673 K01484 K00840 K06447 K05526
B021402	Glutamate	Amino acid biosynthesis	K01745 K01712 K01468 (K01479,K00603,K13990,(K05603 K01458))
B021501	Histidine	Amino acid biosynthesis	K00765 ((K01523 K01496),K11755,K14152) (K01814,K24017) ((K02501+K02500),K01663) ((K01693 K00817 (K04486,K05602,K18649)),(K01089 K00817)) (K00013,K14152)
B021601	Tryptophan	Amino acid biosynthesis	((((K01657+K01658),K13503,K13501,K01656) K00766),K13497) (((K01817,K24017) (K01656,K01609)),K13498,K13501) ((K01695+(K01696,K06001)),K01694)
B021701	Phenylalanine	Amino acid biosynthesis	(((K01850,K04092,K14187,K04093,K04516,K06208,K06209) (K01713,K04518,K05359)),K14170) (K00832,K00838)
B021801	Tyrosine	Amino acid biosynthesis	(((K01850,K04092,K14170,K04093,K04516,K06208,K06209) (K04517,K00211)),K14187) (K00832,K00838)
B021802	Tyrosine	Amino acid biosynthesis	(K01850,K04092,K14170) (K00832,K15849) (K00220,K24018,K15227)
B021901	GABA	Amino acid biosynthesis	K09470 K09471 K09472 K09473
B022001	Beta-alanine	Amino acid biosynthesis	(K00207,(K17722+K17723)) K01464 (K01431,K06016)
B022002	Beta-alanine	Amino acid biosynthesis	6.2.1.17 1.3.8.1 4.2.1.116 3.1.2.4 1.1.159 2.6.1.18
B022101	Ornithine	Amino acid biosynthesis	(K00618,K00619,K14681,K14682,K00620,K22477,K22478) (((K00930,K22478) K00145),K12659) (K00818,K00821) (K01438,K14677,K00620)
B022102	Ornithine	Amino acid biosynthesis	K19412 K05828 K05829 K05830 K05831
B022103	Ornithine	Amino acid biosynthesis	2.3.1.1 2.7.2.8 1.2.1.38 2.6.1.11 3.5.1.16
B030201	Betaine	Amino acid derivative biosynthesis	1.1.99.1 1.2.1.8
B030202	Betaine	Amino acid derivative biosynthesis	2.1.1.156 2.1.1.157
B030301	Ectoine	Amino acid derivative biosynthesis	K00928 K00133 K00836 K06718 K06720
B030701	Spermidine	Amino acid derivative biosynthesis	(K01583,K01584,K01585,K02626) K01480
B030901	Putrescine	Amino acid derivative	K01476 K01581

		biosynthesis	
B031001	Tryptamine	Amino acid derivative biosynthesis	(4.1.1.28,4.1.1.105)
B040101	Acetate	SCFA biosynthesis	(K00625,K13788,K15024) K00925
B040103	Acetate	SCFA biosynthesis	K01067
B040104	Acetate	SCFA biosynthesis	5.4.3.2 5.4.3.3 1.4.1.11 2.3.1.247 1.3.1.109 2.8.3.9 (2.3.1.9,2.3.1.16) 2.3.1.8 (2.7.2.1,2.7.2.15)
B040105	Acetate	SCFA biosynthesis	1.21.4.2 (2.7.2.1,2.7.2.15)
B040106	Acetate	SCFA biosynthesis	(1.2.7.1,1.2.1.104) 2.3.1.8 (2.7.2.1,2.7.2.15)
B040201	Butyrate	SCFA biosynthesis	1.2.7.1 (2.3.1.9,2.3.1.16) 1.1.1.35 4.2.1.150 1.3.1.109 2.3.1.19 (2.7.2.7,2.7.2.14)
B040202	Butyrate	SCFA biosynthesis	2.3.1.8 (2.7.2.1,2.7.2.15) (2.8.3.1,2.8.3.8)
B040203	Butyrate	SCFA biosynthesis	(2.3.1.9,2.3.1.16) 1.1.1.36 4.2.1.55 1.3.1.109 (2.8.3.1,2.8.3.8)
B040204	Butyrate	SCFA biosynthesis	1.4.1.2 1.1.1.399 2.8.3.12 4.2.1.167 7.2.4.5 1.3.1.109 (2.8.3.1,2.8.3.8)
B040205	Butyrate	SCFA biosynthesis	5.4.3.2 5.4.3.3 1.4.1.11 2.3.1.247 1.3.1.109 2.8.3.9
B040206	Butyrate	SCFA biosynthesis	2.8.3.18 1.2.1.76 1.1.1.61 2.8.3.M6 4.2.1.120 1.3.1.109 (2.8.3.1,2.8.3.8)
B040207	Butyrate	SCFA biosynthesis	1.4.1.2 1.1.1.399 2.8.3.12 4.2.1.167 7.2.4.5 1.3.1.109 (2.8.3.1,2.8.3.8)
B040208	Butyrate	SCFA biosynthesis	1.4.1.2 1.1.1.399 2.8.3.12 4.2.1.167 7.2.4.5 1.3.1.109 (2.8.3.1,2.8.3.8)
B040301	Propionate	SCFA biosynthesis	(4.2.1.28,1.1.1.1) 1.2.1.87 2.3.1.222 (2.7.2.1,2.7.2.7,2.7.2.14,2.7.2.15)
B040302	Propionate	SCFA biosynthesis	4.3.1.19 2.3.1.222 (2.7.2.1,2.7.2.7,2.7.2.14,2.7.2.15)
B040304	Propionate	SCFA biosynthesis	2.1.3.1 1.1.1.37 4.2.1.2 1.3.5.1 2.8.3.27
B040305	Propionate	SCFA biosynthesis	2.8.3.1 4.2.1.54 1.3.1.95 2.8.3.1
B040306	Propionate	SCFA biosynthesis	2.6.1.2 1.1.1.28 2.8.3.1 4.2.1.54 1.3.1.95 2.8.3.1
B050101	Indole-3-acetate	Indolic compound biosynthesis	1.13.12.3 3.5.1.4
B050102	Indole-3-acetate	Indolic compound biosynthesis	4.2.1.84 3.5.1.4
B050103	Indole-3-acetate	Indolic compound biosynthesis	3.5.5.1
B050104	Indole-3-acetate	Indolic compound biosynthesis	(2.6.1.1,2.6.1.27) 4.1.1.74 1.2.3.7
B050105	Indole-3-acetate	Indolic compound biosynthesis	(4.1.1.28,4.1.1.105) 1.4.3.4 1.2.3.7
B060401	indole-3-lactate	Organic anion biosynthesis	1.1.1.27
B060402	L-lactate	Organic anion biosynthesis	1.1.1.22
B060501	D-lactate	Organic anion biosynthesis	1.1.1.28
B060101	Succinate	Organic anion biosynthesis	(K01647,K05942) (K01681,K01682) (K00031,K00030) ((((K00164+K00658),K01616)+K00382),(K00174+K00175)) ((K01902+K01903),(K01899+K01900),K18118) ((K00234+K00235+K00236+(K00237,K25801)),(K00239+K00240+K00241),(K00244+K00245+K00246)) (K01676,K01679,(K01677+K01678)) (K00026,K00025,K00024,K00116)
B060102	Succinate	Organic anion biosynthesis	((((K00164+K00658),K01616)+K00382),K00174) ((K01902+K01903),(K01899+K01900),K18118) ((K00234+K00235+K00236+(K00237,K25801)),(K00239+K00240+K00241),(

			K00244+K00245+K00246)) (K01676,K01679,(K01677+K01678)) (K00026,K00025,K00024,K00116)
B060103	Succinate	Organic anion biosynthesis	K01580 (K13524,K07250,K00823,K16871) (K00135,K00139,K17761)
B060104	Succinate	Organic anion biosynthesis	(K00169+K00170+K00171+K00172) K01007 K01595 K00024 (K01677+K01678) (K00239+K00240) (K01902+K01903) (K15038,K15017) K14465 (K14467,K18861) K14534 K15016 K00626
B060105	Succinate	Organic anion biosynthesis	(K02160+K01961+K01962+K01963) K14468 K14469 K15052 K05606 (K01847,(K01848+K01849)) (K14471+K14472) (K00239+K00240+K00241) K01679 K08691 K14449 K14470 K09709
B060106	Succinate	Organic anion biosynthesis	(K00169+K00170+K00171+K00172) (K01959+K01960) K00024 (K01677+K01678) (K18209+K18210) (K01902+K01903) (K00174+K00175+K00176+K00177)
B060201	Fumarate	Organic anion biosynthesis	(K01647,K05942) (K01681,K01682) (K00031,K00030) ((((K00164+K00658),K01616)+K00382),(K00174+K00175)) ((K01902+K01903),(K01899+K01900),K18118) ((K00234+K00235+K00236+(K00237,K25801)),(K00239+K00240+K00241),(K00244+K00245+K00246)) (K01676,K01679,(K01677+K01678)) (K00026,K00025,K00024,K00116)
B060202	Fumarate	Organic anion biosynthesis	((((K00164+K00658),K01616)+K00382),K00174) ((K01902+K01903),(K01899+K01900),K18118) ((K00234+K00235+K00236+(K00237,K25801)),(K00239+K00240+K00241),(K00244+K00245+K00246)) (K01676,K01679,(K01677+K01678)) (K00026,K00025,K00024,K00116)
B060203	Fumarate	Organic anion biosynthesis	K01948 K00611 K01940 (K01755,K14681) K01476
B060204	Fumarate	Organic anion biosynthesis	(K00815,K00838,K00832,K03334) K00457 K00451 K01800 (K01555,K16171)
B060205	Fumarate	Organic anion biosynthesis	K00241+(K00242,K18859,K18860)+K00239+K00240
B060206	Fumarate	Organic anion biosynthesis	K00244+K00245+K00246+K00247
B060207	Fumarate	Organic anion biosynthesis	(K00169+K00170+K00171+K00172) K01007 K01595 K00024 (K01677+K01678) (K00239+K00240) (K01902+K01903) (K15038,K15017) K14465 (K14467,K18861) K14534 K15016 K00626
B060208	Fumarate	Organic anion biosynthesis	(K02160+K01961+K01962+K01963) K14468 K14469 K15052 K05606 (K01847,(K01848+K01849)) (K14471+K14472) (K00239+K00240+K00241) K01679 K08691 K14449 K14470 K09709
B060209	Fumarate	Organic anion biosynthesis	(K00169+K00170+K00171+K00172) (K01959+K01960) K00024 (K01677+K01678) (K18209+K18210) (K01902+K01903) (K00174+K00175+K00176+K00177)
B060210	Fumarate	Organic anion biosynthesis	(K18029+K18030) K14974 K18028 K15357 K13995 K01799
B060211	Fumarate	Organic anion biosynthesis	K00611 K01940 (K01755,K14681)
B060212	Fumarate	Organic anion biosynthesis	K22478 K00145 K00821 K09065 K01438 K01940 K01755
B060213	Fumarate	Organic anion biosynthesis	2.6.1.1,2.6.1.5,2.6.1.27,2.6.1.57 1.13.11.27 1.13.11.5 5.2.1.2 3.7.1.2
B060301	Citrate	Organic anion biosynthesis	(K01647,K05942) (K01681,K01682) (K00031,K00030) ((((K00164+K00658),K01616)+K00382),(K00174+K00175)) ((K01902+K01903),(K01899+K01900),K18118) ((K00234+K00235+K00236+(K00237,K25801)),(K00239+K00240+K00241),(K00244+K00245+K00246)) (K01676,K01679,(K01677+K01678)) (K00026,K00025,K00024,K00116)
B060302	Citrate	Organic anion biosynthesis	(K01647,K05942) (K01681,K01682) (K00031,K00030)
B060303	Citrate	Organic anion biosynthesis	K01647 (K01681,K01682) K01637 (K01638,K19282) (K00026,K00025,K00024)
B060304	Citrate	Organic anion	K01647 K01681 K00031 K00261 (K19268+K01846) K04835 K19280 K14449 K19281 K19282 K00024

		biosynthesis	
B070101	Thiamine (B1)	Vitamin biosynthesis	(((K03148+K03154) K03151),(K03150 K03149)) K03147 ((K00941 K00788),K14153,K21219) K00946
B070102	Thiamine (B1)	Vitamin biosynthesis	(((K03148+K03154) K03151),(K03153 K03149 K10810)) K03147 K00941 K00788 K00946
B070103	Thiamine (B1)	Vitamin biosynthesis	(K22699,K03147) ((K00941 (K00788,K21220)),K21219) K00946
B070104	Thiamine (B1)	Vitamin biosynthesis	(K00941 K00788),K14153,K21219
B070201	Riboflavin (B2)	Vitamin biosynthesis	(((K01497,K14652) ((K01498 K00082),K11752) (K22912,K20860,K20861,K20862,K21063,K21064)),(K02858,K14652)) K00794 K00793 ((K20884 K22949),K11753)
B070301	Niacin (B3)	Vitamin biosynthesis	3.6.1.22 (3.2.2.6,3.2.2.4) 3.4.1.19
B070401	Pantothenate (B5)	Vitamin biosynthesis	((K00826 K00606 K00077),K01579) (K01918,K13799)
B070402	Pantothenate (B5)	Vitamin biosynthesis	((K00606 K00077),(K13367 K00128)) K01918
B070501	Pyridoxal-P (B6)	Vitamin biosynthesis	K03472 K03473 K00831 K00097 K03474 K00275
B070502	Pyridoxal-P (B6)	Vitamin biosynthesis	K06215 K08681
B070601	Biotin (B7)	Vitamin biosynthesis	K00652 (((K00833,K19563) K01935),K19562) K01012
B070602	Biotin (B7)	Vitamin biosynthesis	K00652 K25570 K01935 K01012
B070603	Biotin (B7)	Vitamin biosynthesis	K16593 K00652 K19563 K01935 K01012
B070604	Biotin (B7)	Vitamin biosynthesis	K01906 K00652 (K00833,K19563) K01935 K01012
B070701	Tetrahydrofolate (B9)	Vitamin biosynthesis	(K01495,K09007,K22391) (K01077,K01113,(K08310,K19965)) ((K13939,((K13940,(K01633 K00950)) K00796)),(K01633 K13941)) (K11754,K20457) (K00287,K13998)
B070702	Tetrahydrofolate (B9)	Vitamin biosynthesis	K14652 K22100 K01633 K13941 K22099 K00287
B070801	Cobalamin (B12)	Vitamin biosynthesis	(K02302,((K02303,K13542) (K02304,K24866))) (K02190,K03795,K22011) K03394 (K05934,K13541,K21479) K05936 (K02189,K13541) K02188 K05895 ((K02191 K03399),K00595) K06042 K02224
B070802	Cobalamin (B12)	Vitamin biosynthesis	(K02303,K13542) (K03394,K13540) K02229 (K05934,K13540,K13541) K05936 K02228 K05895 K00595 K06042 K02224 K02230+K09882+K09883
B070803	Cobalamin (B12)	Vitamin biosynthesis	(K00798,K19221) K02232 (K02225,K02227) K02231 K00768 (K02226,K22316) K02233
B070901	Tocopherol/tocotorienol (E)	Vitamin biosynthesis	K09833 (K12502,K18534) K09834 K05928
B071001	Phylloquinone (K1)	Vitamin biosynthesis	((K02552 K02551 K08680 K02549),K14759) (K01911,K14760) K01661 (K19222,K12073) K23094 K17872 K23095
B071101	Menaquinone (K2)	Vitamin biosynthesis	K02552 K02551 K08680 K02549 K01911 K01661 K19222 K02548 K03183
B071102	Menaquinone (K2)	Vitamin biosynthesis	K11782 K18285 (K18286,K20810) K11783 K11784 K11785

B071103	Menaquinone (K2)	Vitamin biosynthesis	K11782 K18285 K18284 K11784 K11785
B071201	Ubiquinone (Q10)	Vitamin biosynthesis	(K03181,K18240) K03179 (K03182+K03186) K18800 K00568 K03185 K03183 (K03184,K06134) K00568
B071202	Ubiquinone (Q10)	Vitamin biosynthesis	K06125 K06126 K00591 K06127 K06134 K00591
B080101	Salicylate	Aromatic compound biosynthesis	5.4.4.2 4.2.99.21
B080201	Gallate	Aromatic compound biosynthesis	4.2.1.10
B080301	Chorismate	Aromatic compound biosynthesis	4.2.1.10 1.1.1.25 2.7.1.71 2.5.1.19 4.2.3.5
B080302	Chorismate	Aromatic compound biosynthesis	2.5.1.54 3.2.3.4 4.2.1.10 1.1.1.25 2.7.1.71 2.5.1.19 4.2.3.5
B080303	Chorismate	Aromatic compound biosynthesis	2.7.2.4 1.2.1.11 2.2.1.10 1.4.1.24 4.2.1.10 1.1.1.25 2.7.1.71 2.5.1.19 4.2.3.5
B080404	Dipicolinate	Aromatic compound biosynthesis	2.7.2.4 1.2.1.11 4.3.3.7
B090101	Staphyloferrin	Metallophore biosynthesis	K21898 K23446 K23447
B090102	Staphyloferrin	Metallophore biosynthesis	(5.1.1.10,5.1.1.12) 6.3.2.58 6.3.2.57
B090103	Staphyloferrin	Metallophore biosynthesis	K23371 K21949 K21721 K23372 K23373 K23374 K23375
B090104	Staphyloferrin	Metallophore biosynthesis	2.7.1.225 2.5.1.140 1.5.1.51 6.3.2.54 4.1.1.117 6.3.2.55 6.3.2.56
B090201	Aerobactin	Metallophore biosynthesis	K03897 K03896 K03894 K03895
B090202	Aerobactin	Metallophore biosynthesis	1.14.13.59 2.3.1.102 6.3.2.38 6.3.2.39
B090301	Staphylopine	Metallophore biosynthesis	(5.1.1.10,5.1.1.24) 2.5.1.152 1.5.1.52
B100402	Bacilysin	Antibiotic biosynthesis	5.4.99.5 4.1.1.100 5.3.3.19 ((5.3.3.19 1.3.1.aa),1.3.1.aa) 1.1.1.385 6.3.2.49
B100601	Carbapenem-3-carboxylate	Antibiotic biosynthesis	K18317 K18316 K18315
B100801	Clavaminate	Antibiotic biosynthesis	K12673 K12674 K12675 K12676
B100802	Clavaminate	Antibiotic biosynthesis	2.5.1.66 6.3.3.4 1.14.11.21 3.5.3.22 1.14.11.21
B101101	Erythromycin	Antibiotic biosynthesis	2.1.1.254 1.14.13.154

B101102	Erythromycin	Antibiotic biosynthesis	2.3.1.94 1.14.15.35 2.4.1.328 2.4.1.278
B101202	Fosfomycin	Antibiotic biosynthesis	5.4.2.9 4.1.1.82 1.1.1.309 2.7.7.104 2.1.1.308 1.11.1.23
B101401	Kanosamine	Antibiotic biosynthesis	K18652 K18653 K18654
B101402	Kanosamine	Antibiotic biosynthesis	1.1.1.361 2.6.1.104 3.1.3.92
B102101	Novobiocin	Antibiotic biosynthesis	6.3.1.15 2.1.1.284 2.4.1.302 2.1.1.285 2.1.3.12
B102201	Paromamine	Antibiotic biosynthesis	4.2.3.124 2.6.1.100 (1.1.1.329,1.1.99.38) 2.6.1.101 2.4.1.283 3.5.1.112
B102401	Pentalenolactone	Antibiotic biosynthesis	K12250 K15907 K18056 K17747 K18091 K18057 K17476
B102402	Pentalenolactone	Antibiotic biosynthesis	4.2.3.7 1.4.15.32 1.14.11.35 1.1.1.340 1.14.13.170 1.14.11.36 1.14.19.8
B102601	Prodigiosin	Antibiotic biosynthesis	((K21780+K21781) K21782 K21783 K21784 K21785 K21786) (K21428 K21778 K21779) K21787
B102801	Pyocyanin	Antibiotic biosynthesis	K13063 K20261 K06998 K20260 K20262 K21103 K20940
B102802	Pyocyanin	Antibiotic biosynthesis	2.1.1.327 1.14.13.218
B102901	Pyrrolnitrin	Antibiotic biosynthesis	K14266 K19981 K14257 K19982
B104101	Validamycin A	Antibiotic biosynthesis	K19969 K20431 K20432 K20433 K20434 K20435 K20436 K20437 K20438
B104102	Validamycin A	Antibiotic biosynthesis	4.2.3.152 5.1.3.33 2.7.1.214 (2.6.1.M1,2.7.7.91) 2.5.1.135 3.1.3.101 2.4.1.338 1.14.11.52
B104201	Violacein	Antibiotic biosynthesis	K20086 (K20087+K20088) K20089 K20090
B104202	Violacein	Antibiotic biosynthesis	1.4.3.23 1.21.98.2 1.14.13.217 1.14.13.224
D010101	Triglyceride	Lipid degradation	(K01046,K12298,K16816,K13534,K14073,K14074,K14075,K14076,K22283, K14452,K22284,K14674,K14675,K17900) (K01054,K25824)
D010102	Triglyceride	Lipid degradation	(3.1.1.3,3.1.1.34) (3.1.1.34,3.1.1.79,3.1.1.116) (3.1.1.23,3.1.1.79)
D010201	Fatty acid	Lipid degradation	(K01897,K15013) (K00232,K00249,K00255,K06445,K09479) (((K01692,K07511,K13767) (K00022,K07516)),K01825,K01782,K07514,K07515,K10527) (K00632,K07508,K07509,K07513)
D010301	Oleate	Lipid degradation	6.2.1.3 1.3.8.8 4.2.1.17 1.1.1.35 2.3.1.16 1.3.8.8 4.2.1.17 1.1.1.35 2.3.1.16 1.3.8.8 4.2.1.17 (1.1.1.35,1.1.1.211) 2.3.1.16 5.3.3.8 4.2.1.74
D010401	Dicarboxylic acids	Lipid degradation	6.2.1.5 1.3.8.7 4.2.1.17 1.1.1.35 2.3.1.174
D020101	Cellulose	Polysaccharide degradation	(3.2.1.4,3.2.1.176,3.2.1.132,3.2.1.73) (3.2.1.176,3.2.1.4,3.2.1.14) (1.14.99.54,1.14.99.56,1.14.99.53) (1.14.99.54,1.14.99.53) 1.14.99.54 (1.14.99.54,1.14.99.56) (1.14.99.54,1.14.99.56,1.14.99.53) (3.2.1.4,3.2.1.8,3.2.1.21,3.2.1.25,3.2.1.45,3.2.1.58,3.2.1.73,3.2.1.74,3.2.1.75 ,3.2.1.78,3.2.1.91,3.2.1.104,3.2.1.123,3.2.1.132,3.2.1.149,3.2.1.151,3.2.1.16 4,3.2.1.168,3.2.1.73,3.2.1.39,3.2.1.52,3.2.1.132,3.2.1.146) (3.2.1.4,3.2.1.91) (3.2.1.4,3.2.1.176,3.2.1.132,3.2.1.73) (3.2.1.132,3.2.1.4,3.2.1.73,3.2.1.8,3.2.1.156) (3.2.1.4,3.2.1.6,3.2.1.21,3.2.1.73,3.2.1.74,3.2.1.91,3.2.1.151,3.2.1.165) (3.2.1.8,3.2.1.32,3.2.1.4) 3.2.1.4 (3.2.1.4,3.2.1.151,3.2.1.73,2.4.1.207) (3.2.1.4,3.2.1.151) (3.2.1.4,3.2.1.151,3.2.1.78) (3.2.1.4,3.2.1.8,3.2.1.21,3.2.1.25,3.2.1.45,3.2.1.58,3.2.1.73,3.2.1.74,3.2.1.75 ,3.2.1.78,3.2.1.91,3.2.1.104,3.2.1.123,3.2.1.132,3.2.1.149,3.2.1.151,3.2.1.16

			4,3.2.1.168,3.2.1.73,3.2.1.39,3.2.1.52,3.2.1.132,3.2.1.146) (3.2.1.4,3.2.1.91) (3.2.1.4,3.2.1.176,3.2.1.132,3.2.1.73) (3.2.1.132,3.2.1.4,3.2.1.73,3.2.1.8,3.2.1.156) (3.2.1.4,3.2.1.6,3.2.1.21,3.2.1.73,3.2.1.74,3.2.1.91,3.2.1.151,3.2.1.165) (3.2.1.8,3.2.1.32,3.2.1.4) 3.2.1.4 (3.2.1.4,3.2.1.151,3.2.1.73,2.4.1.207) (3.2.1.4,3.2.1.151) (3.2.1.4,3.2.1.151,3.2.1.78)
D020201	Xyloglucan	Polysaccharide degradation	(3.2.1.4,3.2.1.8,3.2.1.21,3.2.1.25,3.2.1.45,3.2.1.58,3.2.1.73,3.2.1.74,3.2.1.75 ,3.2.1.78,3.2.1.91,3.2.1.104,3.2.1.123,3.2.1.132,3.2.1.149,3.2.1.151,3.2.1.16 4,3.2.1.168,3.2.1.73,3.2.1.39,3.2.1.52,3.2.1.132,3.2.1.146) (3.2.1.4,3.2.1.6,3.2.1.21,3.2.1.73,3.2.1.74,3.2.1.91,3.2.1.151,3.2.1.165) (3.2.1.4,3.2.1.151,3.2.1.73,2.4.1.207) (2.4.1.207,3.2.1.103,3.2.1.39,3.2.1.6,3.2.1.73,3.2.1.81,3.2.1.83,3.2.1.151,3.2 .1.181,3.2.1.178,3.2.1.35,3.2.1.181) (3.2.1.4,3.2.1.151) (3.2.1.4,3.2.1.151,3.2.1.78) (3.2.1.176,3.2.1.4,3.2.1.14) (3.2.1.4,3.2.1.150,3.2.1.151) (1.14.99.54,1.14.99.56) (3.2.1.20,3.2.1.22,3.2.1.24,3.2.1.84,3.2.1.48,3.2.1.10,3.2.1.177,4.2.2.13,2.4. 1.161) (3.2.1.37,3.2.1.55,3.2.1.8,3.2.1.99,3.2.1.145,3.2.1.146) (3.2.1.8,3.2.1.32,3.2.1.4) (3.2.1.23,3.2.1.25,3.2.1.31,3.2.1.55,3.2.1.152,3.2.1.165,3.2.1.37,3.2.1.146)
D020301	Starch	Polysaccharide degradation	(3.2.1.1,3.2.1.41,2.4.1.19,3.2.1.54,3.2.1.93,3.2.1.10,3.2.1.133,3.2.1.135,3.2. 1.20,3.2.1.60,3.2.1.68,3.2.1.70,3.2.1.98,3.2.1.116,2.4.1.18,5.4.99.16,2.4.1.2 5,2.4.1.4,2.4.1.7,3.2.1.141,5.4.99.11,5.4.99.15,3.2.1.33,2.4.99.16) 3.2.1.2 (3.2.1.1,3.2.1.22,3.2.1.41,3.2.1.54,2.4.1.18,2.4.1.25) 3.2.1.1 3.2.1.33 (3.2.1.3,3.2.1.70,3.2.1.28,2.4.1.2) (3.2.1.3,3.2.1.20,3.2.1.22)
D020401	Chitin	Polysaccharide degradation	(3.2.1.14,3.2.1.17,3.2.1.96) (3.2.1.14,3.2.1.17) (3.2.1.17,4.2.2.n1,3.2.1.14) (3.2.1.17,3.2.1.96) (3.2.1.52,3.2.1.140) (3.2.1.21,3.2.1.37,3.2.1.45,3.2.1.52,3.2.1.55,3.2.1.58,3.2.1.74,3.2.1.120,3.2. 1.126) (3.2.1.4,3.2.1.8,3.2.1.21,3.2.1.25,3.2.1.45,3.2.1.58,3.2.1.73,3.2.1.74,3.2.1.75 ,3.2.1.78,3.2.1.91,3.2.1.104,3.2.1.123,3.2.1.132,3.2.1.149,3.2.1.151,3.2.1.16 4,3.2.1.168,3.2.1.73,3.2.1.39,3.2.1.52,3.2.1.132,3.2.1.146) (3.2.1.52,3.2.1.35,3.2.1.169) (3.2.1.21,3.2.1.37,3.2.1.45,3.2.1.52) (1.14.99.54,1.14.99.56,1.14.99.53) (1.14.99.54,1.14.99.53) (3.1.1.72,3.5.1.41)
D020501	Pectin	Polysaccharide degradation	(3.2.1.15,3.2.1.40,3.2.1.67,3.2.1.82,3.2.1.171,3.2.1.173) (4.2.2.2,4.2.2.9,4.2.2.10) (4.2.2.2,4.2.2.9) (4.2.2.2,4.2.2.9) 4.2.2.2 (4.2.2.23,4.2.2.24) 4.2.2.6 4.2.2.24 4.2.2.23 (3.2.1.40,3.2.1.174) 3.2.1.173 3.1.1.11 3.2.1.172 (3.2.1.122,3.2.1.20,3.2.1.22,3.2.1.86,3.2.1.139,3.2.1.67) 3.1.1.72 (3.2.1.23,3.2.1.25,3.2.1.31,3.2.1.55,3.2.1.152,3.2.1.165,3.2.1.37,3.2.1.146)
D020601	Alpha galactan	Polysaccharide degradation	3.2.1.49 (3.2.1.22,3.2.1.49,3.2.1.94,3.2.1.88) 3.2.1.22 (3.2.1.122,3.2.1.20,3.2.1.22,3.2.1.86,3.2.1.139,3.2.1.67) (3.2.1.20,3.2.1.22,3.2.1.24,3.2.1.84,3.2.1.48,3.2.1.10,3.2.1.177,4.2.2.13,2.4. 1.161) (3.2.1.22,3.2.1.49,2.4.1.67,2.4.1.82) (3.2.1.3,3.2.1.20,3.2.1.22)
D020701	Beta-galactan	Polysaccharide degradation	3.2.1.89 (3.2.1.23,3.2.1.25,3.2.1.31,3.2.1.55,3.2.1.152,3.2.1.165,3.2.1.37,3.2.1.146) (3.2.1.23,3.2.1.165) 3.2.1.23 (3.2.1.21,3.2.1.23,3.2.1.25,3.2.1.31,3.2.1.37,3.2.1.38,3.2.1.62,3.2.1.74,3.2.1. 85,3.2.1.86,3.2.1.105,3.2.1.108,3.2.1.117,3.2.1.118,3.2.1.119,3.2.1.125,3.2.1 .147,3.2.1.149,3.2.1.161,3.2.1.175,3.2.1.182) (3.2.1.23,3.2.1.46)
D020801	Mixed-Linkage glucans	Polysaccharide degradation	3.2.1.71 (3.2.1.8,3.2.1.31,3.2.1.37,3.2.1.38,3.2.1.45,3.2.1.75,3.2.1.136) (3.2.1.39,3.2.1.58,3.2.1.73,3.2.1.175) (3.2.1.4,3.2.1.176,3.2.1.132,3.2.1.73) (3.2.1.132,3.2.1.4,3.2.1.73,3.2.1.8,3.2.1.156) (3.2.1.4,3.2.1.6,3.2.1.21,3.2.1.73,3.2.1.74,3.2.1.91,3.2.1.151,3.2.1.165) (3.2.1.4,3.2.1.8,3.2.1.21,3.2.1.25,3.2.1.45,3.2.1.58,3.2.1.73,3.2.1.74,3.2.1.75 ,3.2.1.78,3.2.1.91,3.2.1.104,3.2.1.123,3.2.1.132,3.2.1.149,3.2.1.151,3.2.1.16 4,3.2.1.168,3.2.1.73,3.2.1.39,3.2.1.52,3.2.1.132,3.2.1.146) (3.2.1.58,3.2.1.39) 3.2.1.39 (3.2.1.21,3.2.1.37,3.2.1.45,3.2.1.52,3.2.1.55,3.2.1.58,3.2.1.74,3.2.1.120,3.2. 1.126) (2.4.1.207,3.2.1.103,3.2.1.39,3.2.1.6,3.2.1.73,3.2.1.81,3.2.1.83,3.2.1.151,3.2 .1.181,3.2.1.178,3.2.1.35,3.2.1.181) (3.2.1.39,3.2.1.58,3.2.1.73,3.2.1.175) (3.2.1.58,3.2.1.39) (3.2.1.21,3.2.1.23,3.2.1.25,3.2.1.31,3.2.1.37,3.2.1.38,3.2.1.62,3.2.1.74,3.2.1. 85,3.2.1.86,3.2.1.105,3.2.1.108,3.2.1.117,3.2.1.118,3.2.1.119,3.2.1.125,3.2.1 .147,3.2.1.149,3.2.1.161,3.2.1.175,3.2.1.182)
D020901	Xylans	Polysaccharide degradation	(3.2.1.8,3.2.1.32,3.2.1.4) (3.2.1.78,3.2.1.100,3.2.1.32,3.2.1.73) (3.2.1.132,3.2.1.4,3.2.1.73,3.2.1.8,3.2.1.156) (3.2.1.4,3.2.1.8,3.2.1.21,3.2.1.25,3.2.1.45,3.2.1.58,3.2.1.73,3.2.1.74,3.2.1.75 ,3.2.1.78,3.2.1.91,3.2.1.104,3.2.1.123,3.2.1.132,3.2.1.149,3.2.1.151,3.2.1.16 4,3.2.1.168,3.2.1.73,3.2.1.39,3.2.1.52,3.2.1.132,3.2.1.146) (3.2.1.8,3.2.1.32) (3.2.1.8,3.2.1.31,3.2.1.37,3.2.1.38,3.2.1.45,3.2.1.75,3.2.1.136) (3.2.1.102,3.2.1.8,3.2.1.8) (3.2.1.51,3.2.1.8) (3.2.1.76,3.2.1.37) (3.2.1.21,3.2.1.37,3.2.1.45,3.2.1.52,3.2.1.55,3.2.1.58,3.2.1.74,3.2.1.120,3.2. 1.126) (3.2.1.20,3.2.1.22,3.2.1.24,3.2.1.84,3.2.1.48,3.2.1.10,3.2.1.177,4.2.2.13,2.4. 1.161) 3.2.1.37 (3.2.1.139,3.2.1.131)
D021001	Beta-mannan	Polysaccharide degradation	3.2.1.78 (3.2.1.78,3.2.1.100,3.2.1.32,3.2.1.73) (3.2.1.4,3.2.1.8,3.2.1.21,3.2.1.25,3.2.1.45,3.2.1.58,3.2.1.73,3.2.1.74,3.2.1.75 ,3.2.1.78,3.2.1.91,3.2.1.104,3.2.1.123,3.2.1.132,3.2.1.149,3.2.1.151,3.2.1.16

			4,3.2.1.168,3.2.1.73,3.2.1.39,3.2.1.52,3.2.1.132,3.2.1.146) (3.2.1.78,3.2.1.100,3.2.1.32,3.2.1.73) (2.4.1.281,2.4.1.319,2.4.1.320) 3.2.1.78 3.1.1.72 (3.2.1.23,3.2.1.25,3.2.1.31,3.2.1.55,3.2.1.152,3.2.1.165,3.2.1.37,3.2.1.146)
D021101	Alpha-mannan	Polysaccharide degradation	3.2.1.130 (3.2.1.101,3.2.1.20) (3.2.1.101,3.2.1.20) (3.2.1.113,3.2.1.24) (3.2.1.24,3.2.1.113,3.2.1.114,3.2.1.170) (3.2.1.106,3.2.1.84,3.2.1.20,3.2.1.170,3.2.1.208) 3.2.1.113 (2.4.1.281,2.4.1.319,2.4.1.320)
D021201	Arabinan	Polysaccharide degradation	(3.2.1.37,3.2.1.55,3.2.1.8,3.2.1.99,3.2.1.145,3.2.1.146) (3.2.1.11,3.2.1.57,3.2.1.95) (3.2.1.37,3.2.1.55,3.2.1.8,3.2.1.99,3.2.1.145,3.2.1.146) (3.2.1.4,3.2.1.8,3.2.1.37,3.2.1.55,3.2.1.73) (3.2.1.21,3.2.1.37,3.2.1.45,3.2.1.52,3.2.1.55,3.2.1.58,3.2.1.74,3.2.1.120,3.2.1.126) (3.2.1.55,3.2.1.37) 3.2.1.55
D021301	Mucin	Polysaccharide degradation	3.2.1.97 (3.2.1.22,3.2.1.49,3.2.1.94,3.2.1.88) 3.2.1.49 (2.4.1.211,2.4.1.247) (3.2.1.20,3.2.1.22,3.2.1.24,3.2.1.84,3.2.1.48,3.2.1.10,3.2.1.177,4.2.2.13,2.4.1.161) (3.2.1.22,3.2.1.49,2.4.1.67,2.4.1.82)
D030101	Lactose	Sugar degradation	3.2.1.85 5.3.1.26 2.7.1.144 4.1.2.40
D030201	Sucrose	Sugar degradation	2.7.1.211 3.2.1.48 2.7.1.4
D030302	D-Apiose	Sugar degradation	1.1.1.420 3.1.1.115
D030401	D-Arabinose	Sugar degradation	5.3.1.3 2.7.1.47
D030402	D-Arabinose	Sugar degradation	5.3.1.3 2.7.1.51 4.1.2.17 1.2.1.21
D030501	D-Mannose	Sugar degradation	2.7.1.191 5.3.1.8
D030502	D-Mannose	Sugar degradation	2.7.1.7 5.3.1.8
D030601	D-Xylose	Sugar degradation	5.3.1.5 2.7.1.17
D030602	D-Xylose	Sugar degradation	1.1.1.9 2.7.1.17
D030603	D-Xylose	Sugar degradation	1.1.1.424 3.1.1.68 4.2.1.82 4.2.1.141 1.2.1.26
D030604	D-Xylose	Sugar degradation	1.1.1.359 3.1.1.110 4.2.1.82 4.1.2.28 1.1.1.26 2.3.3.9
D030605	D-Xylose	Sugar degradation	1.1.1.175 3.1.1.110 4.2.1.82 4.2.1.1.141 1.2.1.26
D030701	L-Fucose	Sugar degradation	5.1.3.29 5.3.1.25 2.7.1.51 4.1.2.17
D030702	L-Fucose	Sugar degradation	5.1.3.29 4.2.1.68 1.1.1.M68 3.7.1.26
D030801	L-Rhamnose	Sugar degradation	5.1.3.32 5.3.1.14 2.7.1.5 4.1.2.19
D030802	L-Rhamnose	Sugar degradation	(1.1.1.173,1.1.1.378) 3.1.1.65 4.2.1.90 4.1.2.53
D030803	L-Rhamnose	Sugar degradation	(1.1.1.173,1.1.1.378) 3.1.1.65 4.2.1.90 1.1.1.401 3.7.1.26
D030901	Galactose	Sugar degradation	K01785 K00849 K00965 K01784
D031001	NeuAc	Sugar degradation	4.1.3.3 5.1.3.8
D031002	NeuAc	Sugar degradation	4.1.3.3 2.7.1.60 5.1.3.9
D040101	Albumin	Protein degradation	A1A (S8A C13 (S1A (S8B S26B)))
D040201	Actin	Protein degradation	S1A (M10A A1A M35 (M9B S1B S1C C1A (M12A C14A A2A C13 M10C)))
D040301	Collagen	Protein degradation	C1A M10A (M12B (S1A A1A (C13 M35 (M12A S1D S8A A2A M4 S8B))))
D040401	Elastin	Protein degradation	M10A (S1A (M23A S01C C1A))
D040501	Glutelin	Protein degradation	S26B S8A S9A
D040601	Keratin	Protein degradation	S1D M12A S1A
D040701	Tropomyosin	Protein	S1B (C2A C1A S1A (M4 C13 M10A (S1D M12A)))

		degradation	
D040801	Troponin	Protein degradation	C2A (A2A M35 C14A)
D050101	Serine	Amino acid degradation	4.3.1.17,4.3.1.18
D050201	Threonine	Amino acid degradation	4.3.1.19
D050301	Cysteine	Amino acid degradation	4.4.1.1,4.4.1.28
D050302	Cysteine	Amino acid degradation	2.6.1.3 2.8.1.2
D050401	Methionine	Amino acid degradation	4.4.1.11
D050501	Valine	Amino acid degradation	2.6.1.42 1.2.1.25 1.3.8.5 4.2.1.150 3.1.2.4 1.1.1.31 1.2.1.27
D050502	Valine	Amino acid degradation	2.6.1.42 4.1.1.72 1.1.1.1
D050601	Isoleucine	Amino acid degradation	2.6.1.42 1.2.1.25 1.3.8.5 4.2.1.150 1.1.1.178 2.3.1.16
D050602	Isoleucine	Amino acid degradation	2.6.1.42 1.2.7.7
D050701	Leucine	Amino acid degradation	K00826 (((K00166+K00167),K11381)+K09699+K00382) (K00253,K00249) (K01968+K01969) (K05607,K13766) K01640
D050801	Lysine	Amino acid degradation	4.1.1.18 2.6.1.82 1.2.1.19 2.6.1.48 1.2.1.20 1.14.11.64 1.1.5.13
D050802	Lysine	Amino acid degradation	1.13.12.2 3.5.1.30 1.6.1.48 1.2.1.20 2.8.3.13
D050803	Lysine	Amino acid degradation	2.6.1.36 1.2.1.31
D050804	Lysine	Amino acid degradation	4.1.1.18 2.6.1.82 1.2.1.19 2.6.1.48 1.2.1.20 2.8.3.13
D050805	Lysine	Amino acid degradation	K01582 K09251 K00137 K07250 K00135 K15737 K15736
D050806	Lysine	Amino acid degradation	K00468 K01506 (K14268,K07250) K00135 ((K15737 K15736),(K01041 K00252 (K01692,K01825,K01782) (K01825,K01782) K00626))
D050901	Arginine	Amino acid degradation	4.1.1.19 3.5.3.11
D050902	Arginine	Amino acid degradation	4.1.1.19 3.5.3.12 3.5.1.53
D050903	Arginine	Amino acid degradation	1.13.12.1 3.5.1.4 3.5.3.7
D050904	Arginine	Amino acid degradation	K01476 K01581
D050905	Arginine	Amino acid degradation	K00613 K00542 K00933
D050906	Arginine	Amino acid degradation	(K01583,K01584,K01585,K02626) K01480 K01611 K00797

D051001	Proline	Amino acid degradation	5.1.1.4 1.21.4.1
D051101	Glutamate	Amino acid degradation	1.4.1.2,1.4.1.3
D051102	Glutamate	Amino acid degradation	2.6.1.1 4.3.1.1
D051103	Glutamate	Amino acid degradation	1.4.1.2 1.1.1.399 2.8.3.12 4.2.1.167 7.2.4.5
D051104	Glutamate	Amino acid degradation	5.4.99.1 4.3.1.2 4.2.1.34 4.1.3.22
D051105	Glutamate	Amino acid degradation	4.1.1.15
D051201	Histidine	Amino acid degradation	K01745 K01712 K01468 (K01479,K00603,K13990,(K05603 K01458))
D051301	Tryptophan	Amino acid degradation	(K00453,K00463) (K01432,K14263,K07130) K00486 K01556 K00452 K03392 (K10217,K23234)
D051302	Tryptophan	Amino acid degradation	4.1.99.1
D051601	Beta-alanine	Amino acid degradation	2.6.1.18 1.2.1.18
D051602	Beta-alanine	Amino acid degradation	2.6.1.120 1.1.1.298
D051701	Ornithine	Amino acid degradation	4.1.1.17 ((2.6.1.82 1.2.1.19),(6.3.1.11 1.4.3.M3 1.2.1.00 3.5.1.94))
D051801	GABA	Amino acid degradation	2.6.1.19 (1.2.1.24,1.2.1.16)
D051802	GABA	Amino acid degradation	2.6.1.19 1.1.1.61 2.8.3.M6 4.2.1.120 1.3.1.109 (2.8.3.1,2.8.3.8)
D060101	Nitrate	Nitrogen compound degradation	((K00370+K00371+K00374),(K02567+K02568)) ((K00362+K00363),(K03385+K15876))
D060102	Nitrate	Nitrogen compound degradation	1.7.5.1 1.7.2.1 1.7.2.5 1.7.2.4
D060103	Nitrate	Nitrogen compound degradation	1.9.6.1 1.7.2.2
D060105	Nitrate	Nitrogen compound degradation	1.7.7.2 1.7.7.1
D060201	Urea	Nitrogen compound degradation	6.3.4.6 3.5.1.54
D060202	Urea	Nitrogen compound degradation	3.5.1.5
D060301	Urate	Nitrogen compound	1.7.3.3 3.5.2.17 4.1.1.97

		degradation	
D060302	Urate	Nitrogen compound degradation	1.14.13.113 3.5.2.17 4.1.1.97
D060401	GlcNAc	Nitrogen compound degradation	2.7.1.59 3.5.1.25 3.5.99.6
D060402	GlcNAc	Nitrogen compound degradation	3.5.1.25 3.5.99.6
D060601	Allantoin	Nitrogen compound degradation	3.5.2.5 3.5.3.9 3.5.3.26 (1.1.1.350,1.1.1.154) 2.1.3.5
D060602	Allantoin	Nitrogen compound degradation	3.5.2.5 3.5.3.4 4.3.2.3
D060603	Allantoin	Nitrogen compound degradation	3.5.2.5 3.5.3.9 3.5.3.26 4.3.2.3
D060701	Creatinine	Nitrogen compound degradation	3.5.4.21 3.5.2.14 3.5.1.59 1.5.3.1
D060801	Betaine	Nitrogen compound degradation	2.1.1.5 1.5.3.10 (1.5.3.24,1.5.3.1)
D060901	L-carnitine	Nitrogen compound degradation	1.14.13.239 1.2.1.4 1.1.1.38
D061001	Methylamine	Nitrogen compound degradation	1.4.9.1
D061002	Methylamine	Nitrogen compound degradation	6.3.4.12 2.1.1.21 1.5.99.5
D061101	Phenylethylamine	Nitrogen compound degradation	(1.4.3.4,1.4.3.21) 1.2.1.39
D061201	Hypotaurine	Nitrogen compound degradation	2.6.1.77 1.2.1.3
D061301	Taurine	Nitrogen compound degradation	2.6.1.77
D061303	Taurine	Nitrogen compound degradation	2.5.1.55
D070101	2,3-Butanediol	Alcohol degradation	1.1.1.4,1.1.1.76
D070201	Ethanol	Alcohol degradation	1.1.1.1 1.2.1.10

D070202	Ethanol	Alcohol degradation	1.1.1.1 1.2.1.3 6.2.1.1
D070401	Glycerol	Alcohol degradation	2.7.1.30 1.1.5.3
D070402	Glycerol	Alcohol degradation	1.1.1.6 2.7.1.29
D070403	Glycerol	Alcohol degradation	4.2.1.30 1.1.1.202
D070404	Glycerol	Alcohol degradation	1.1.1.6 2.7.1.121
D070501	Propylene glycol	Alcohol degradation	4.2.1.28 (1.1.1.1,(1.2.1.87 2.3.1.222 (2.7.2.1,2.7.2.7,2.7.2.14,2.7.2.15)))
D070601	Ethylene glycol	Alcohol degradation	1.1.1.77 1.2.1.21
D070801	Phytol	Alcohol degradation	1.1.1.1 1.2.1.3 6.2.1.3 1.3.1.38
D070901	Polyvinyl alcohol	Alcohol degradation	1.1.2.6
D080101	Toluene	Xenobiotic degradation	(K15760+K15761+K15763+K15764) K00055 K00141
D080103	Toluene	Xenobiotic degradation	K07540 (K07543+K07544) K07545 K07546 (K07547+K07548) (K07549+K07550)
D080201	Xylene	Xenobiotic degradation	(K15757+K15758) K00055 K00141
D080402	Benzene	Xenobiotic degradation	K16249+K16243+K16244+K16242+K16245+K16246
D080501	Benzoate	Xenobiotic degradation	(K05549+K05550+K05784) K05783
D080502	Benzoate	Xenobiotic degradation	K04116 K04117 K07534 K07535 K07536
D080601	Anthranilate	Xenobiotic degradation	(K05599+K05600+K11311),(K16319+K16320+K18248+K18249)
D080701	Catechol	Xenobiotic degradation	K03381 K01856 K03464 (K01055,K14727)
D080702	Catechol	Xenobiotic degradation	(K00446,K07104) ((K10217 K01821 K01617),K10216) (K18364,K02554) (K18365,K01666) (K18366,K04073)
D080801	Cumate	Xenobiotic degradation	(K10619+K16303+K16304+K18227) K10620 K10621 K10622 K10623
D080901	Biphenyl	Xenobiotic degradation	(K08689+K15750+K18087+K18088) K08690 K00462 K10222
D081001	Carbazole	Xenobiotic degradation	K15751 (K15754+K15755) K15756
D081101	Benzoyl-CoA	Xenobiotic degradation	((K04112+K04113+K04114+K04115),(K19515+K19516)) K07537 K07538 K07539
D081201	Naphthalene	Xenobiotic degradation	(K14579+K14580+K14578+K14581) K14582 K14583 K14584 K14585 K00152
D081301	Salicylate	Xenobiotic	K18242+K18243+K14578+K14581

		degradation	
D081401	Terephthalate	Xenobiotic degradation	(K18074+K18075+K18077) K18076
D081501	Phthalate	Xenobiotic degradation	(K18068+K18069) K18067 K04102
D081601	Phenylacetate	Xenobiotic degradation	K01912 (K02609+K02610+K02611+K02612+K02613) K15866 K02618 K02615 K01692 K00074
D081701	Trans-cinnamate	Xenobiotic degradation	(((K05708+K05709+K05710+K00529) K05711),K05712) K05713 K05714 K02554 K01666 K04073
D081801	Caffeine	Xenobiotic degradation	K21722 K21723 K21724
D081901	Mercury	Xenobiotic degradation	4.99.1.2 1.16.1.1
D090101	Penicillin	Antibiotic degradation	K18698,K18699,K18796,K18767,K18797,K19097,K19317,K18768,K18970, K19316,K22346,K18795,K19218,K19217,K17836,K18766
D090201	Carbapenem	Antibiotic degradation	K17837,K18782,K18781,K18780,K19099,K19216
D090301	Cephalosporin	Antibiotic degradation	K19095,K19096,K19100,K19101,K19214,K19215,K20319,K20320,K01467
D090401	Oxacillin	Antibiotic degradation	K17838,K18790,K18791,K19098,K18792,K19213,K21276,K18793,K18971, K22352,K19209,K18976,K18973,K18794,K18972,K21277,K19210,K19211, K19212,K22335,K19319,K22331,K22351,K19320,K19318,K19321,K19322, K21266,K22334,K22333,K22332
D090501	Streptogramin	Antibiotic degradation	K19349,K19350
D090601	Fosfomycin	Antibiotic degradation	K21252
D090701	Tetracycline	Antibiotic degradation	K08151,K08168,K18214,K18218,K18220,K18221
D090801	Macrolide	Antibiotic degradation	K06979,K08217,K18230,K18231,K21251
D091001	Chloramphenicol	Antibiotic degradation	K00638,K08160,K18552,K18553,K18554,K19271
D091101	Lincosamide	Antibiotic degradation	K18236,K19349,K19350,K19545
D091201	Streptothricin	Antibiotic degradation	K19273,K20816

Table S2:
Anova tables of linear mixed models for different components of alpha diversity.

Neutral alpha diversity:

```
> anova(M_divN)
                numDF denDF  F-value p-value
(Intercept)         1   342 43477.19  <.0001
log_seq_depth       1   342    48.80  <.0001
trial               1    42    17.03  0.0002
sex                 1    42     5.48  0.0241
age                 1   342   430.79  <.0001
treatment           2    42     1.77  0.1835
breed               1    42     2.47  0.1238
sex:age             1   342     0.54  0.4647
age:treatment       2   342     2.96  0.0531
age:breed           1   342     0.02  0.8927
```

Phylogenetic alpha diversity:

```
> anova(M_divP)
                numDF denDF  F-value p-value
(Intercept)         1   342 58777.06  <.0001
log_seq_depth       1   342     1.16  0.2813
trial               1    42     0.00  0.9945
sex                 1    42     2.38  0.1307
age                 1   342   877.25  <.0001
treatment           2    42     3.66  0.0343
breed               1    42    13.45  0.0007
sex:age             1   342     2.92  0.0883
age:treatment       2   342     0.28  0.7543
age:breed           1   342     1.05  0.3073
```

Functional alpha diversity:

```
> anova(M_divP)
                numDF denDF  F-value p-value
(Intercept)         1   342 58777.06  <.0001
log_seq_depth       1   342     1.16  0.2813
trial               1    42     0.00  0.9945
sex                 1    42     2.38  0.1307
age                 1   342   877.25  <.0001
treatment           2    42     3.66  0.0343
breed               1    42    13.45  0.0007
sex:age             1   342     2.92  0.0883
age:treatment       2   342     0.28  0.7543
age:breed           1   342     1.05  0.3073
```

Table S3:
Permanova tables for different components of beta diversity.

Neutral beta diversity:

```
adonis2(formula = betadiv_N_caecum_D ~ trial + sex * age + tr
ns = perm)
              Df SumOfSqs      R2        F Pr(>F)
trial          1   0.2537 0.03715  23.3230  0.001 ***
sex            1   0.0248 0.00363   2.2823  0.001 ***
age            1   2.1485 0.31451 197.4727  0.001 ***
treatment      2   0.0419 0.00614   1.9266  0.001 ***
breed          1   0.0994 0.01455   9.1338  0.001 ***
sex:age        1   0.0207 0.00303   1.9026  0.121
age:treatment  2   0.0267 0.00390   1.2254  0.263
age:breed      1   0.0267 0.00391   2.4569  0.051 .
Residual     385   4.1887 0.61318
Total        395   6.8312 1.00000
```

Phylogenetic beta diversity:

```
adonis2(formula = betadiv_P_caecum_D ~ trial + sex * age + tr
ns = perm)
              Df SumOfSqs      R2        F Pr(>F)
trial          1  0.01815 0.02472  22.3960  0.001 ***
sex            1  0.00212 0.00289   2.6150  0.001 ***
age            1  0.37818 0.51492 466.5788  0.001 ***
treatment      2  0.00633 0.00862   3.9073  0.001 ***
breed          1  0.01267 0.01725  15.6329  0.001 ***
sex:age        1  0.00256 0.00348   3.1563  0.072 .
age:treatment  2  0.00201 0.00273   1.2384  0.312
age:breed      1  0.00036 0.00049   0.4462  0.603
Residual     385  0.31206 0.42489
Total        395  0.73444 1.00000
```

Functional beta diversity:

```
adonis2(formula = betadiv_F_caecum_D ~ trial + sex * age +
ns = perm)
              Df SumOfSqs      R2        F Pr(>F)
trial          1   0.2910 0.04111  25.9092  0.001 ***
sex            1   0.0260 0.00367   2.3137  0.001 ***
age            1   2.2157 0.31306 197.3020  0.001 ***
treatment      2   0.0435 0.00615   1.9378  0.001 ***
breed          1   0.1037 0.01465   9.2321  0.001 ***
sex:age        1   0.0213 0.00301   1.8979  0.111
age:treatment  2   0.0267 0.00377   1.1878  0.270
age:breed      1   0.0262 0.00371   2.3362  0.065 .
Residual     385   4.3235 0.61088
Total        395   7.0775 1.00000
```

Table S4:
Variance partition using hierarchical modelling of species communities (Hmsc). Variance percentages were similar for metagenomic (MG) and metatranscriptomic (MT) samples. Most of the microbiota variance occurred between time points (age).

Variance percentage (%)	**MG**	**MT**
Age	65	49
Pen	26	32
Trial	6	5
Treatment	1.15	8
Genetic line	1.15	3
Sex	0.86	3

Figure S1: Heatmap indicating the Genome-Inferred Functional Trait (GIFT) values for each bacterial genome. These quantitative traits range between 0 and 1, with higher numbers indicating a bigger presence of genes involved in each metabolic pathway.

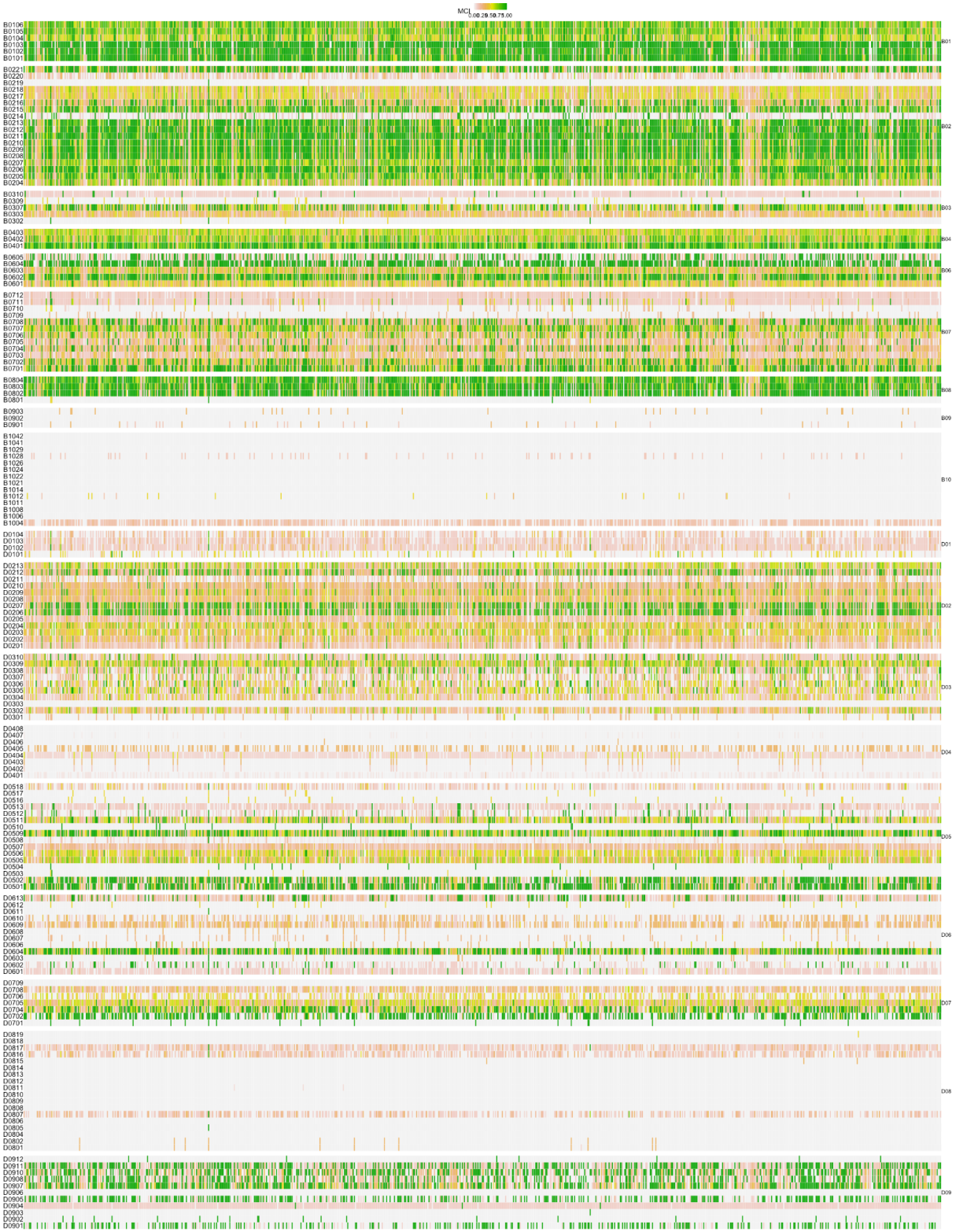


Figure S2:
Temporal differences for neutral, phylogenetic and functional beta diversities comparing animals from the same trial and same pen. The variance of functional attributes ($D^{\beta}_{d7\text{-}35}$ = 0.23±0.06) exceeded that of the phylogenetic change ($D^{\beta}_{d7\text{-}35}$ = 0.08±0.03).

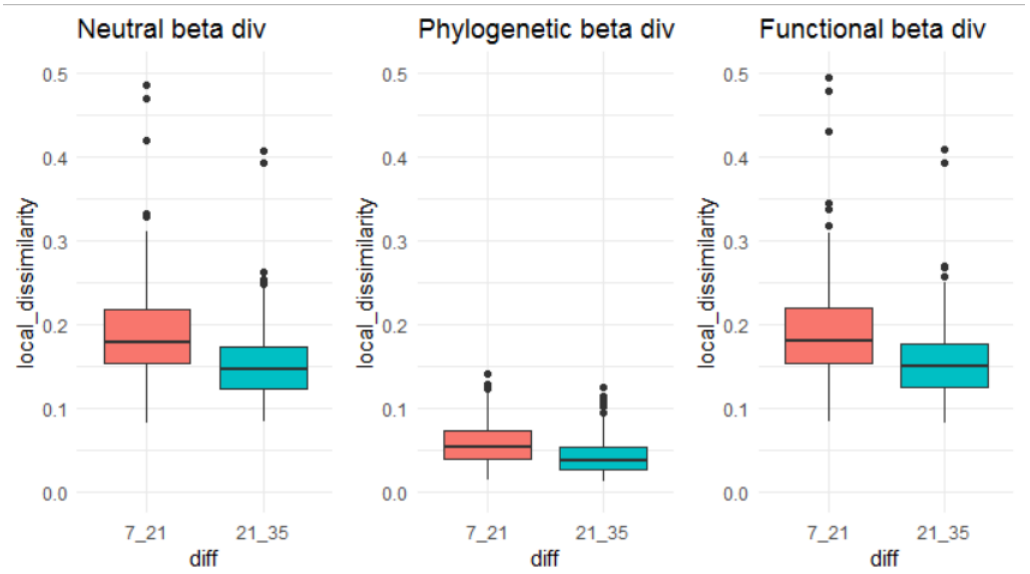


Figure S3:
Phylogenetic correlogram for metatranscriptomic data, which shows a similar trend to the metagenomic data one in Fig. 1g of the main text.

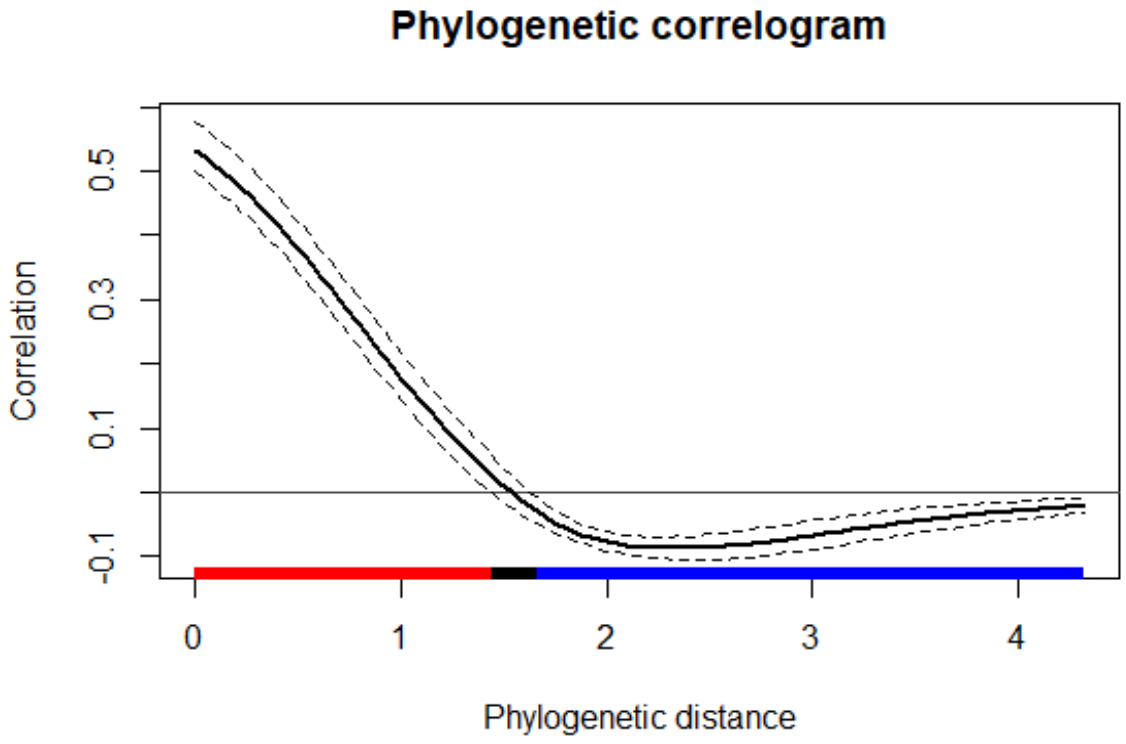


Figure S4:
Linear regressions on day 35 separating chickens by sex and genetic line. The relationship between body weight and increaser bacteria was positive for the two sexes and two chicken genetic lines.

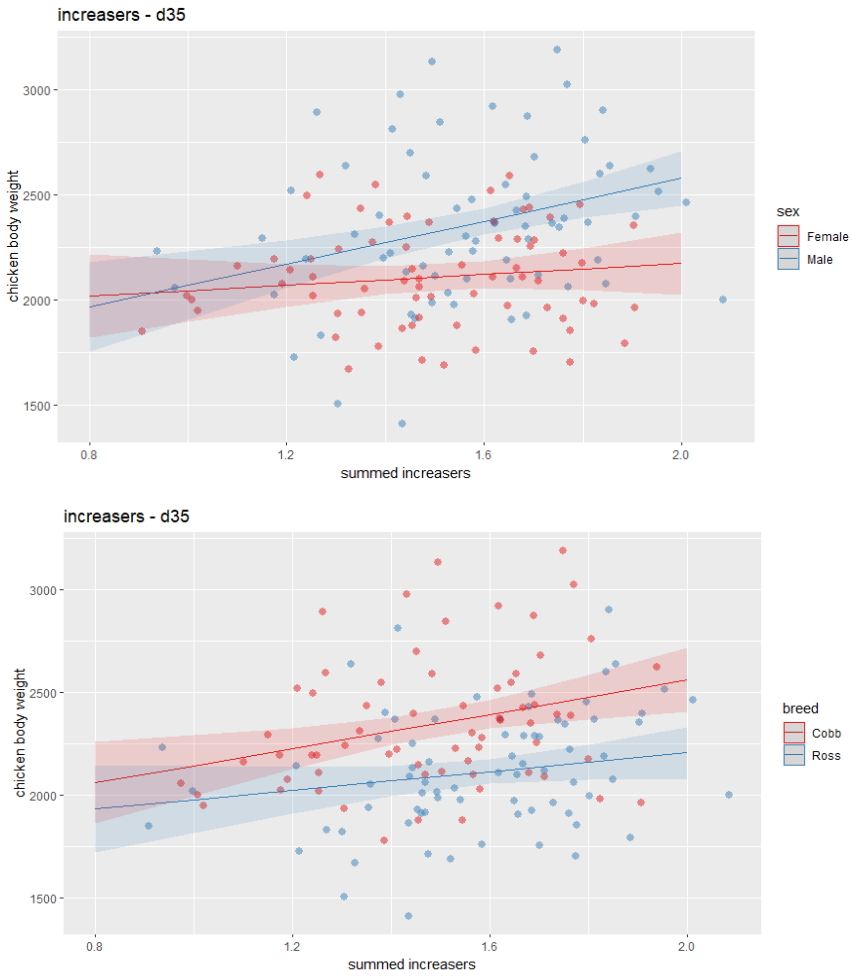


Figure S5:
Mean chicken body weight of 5th and 95th percentile subsets. The difference between both groups is 291 gr. i.e. around 10% difference in total body weight between groups. The null hypothesis of no difference between the two groups is rejected when submitted to a two-way Welch's T-test (t = -2.7073, df = 8.8646, p-value = 0.02442).

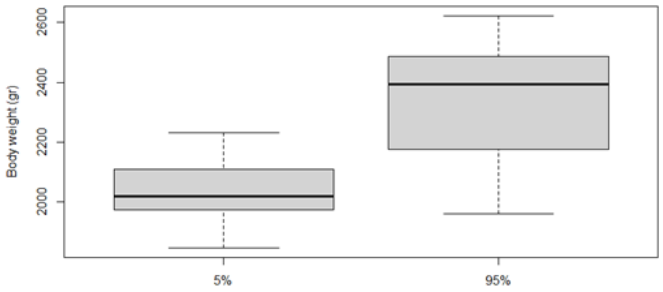


Figure S6:
Linear regressions on days 7 (a), 21 (b) and 35 (c) between chicken body weight and centred log-ratio of the relative abundance of bacteria that decreased their relative representation through time. The linear regressions using decreaser bacteria show the opposite trend that the trend increaser bacteria showed in Fig. 3 of the main text.

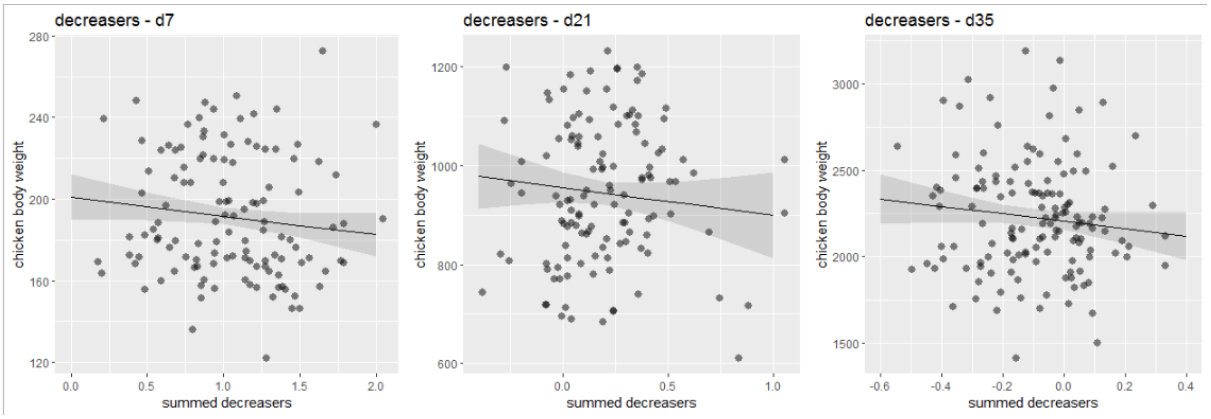
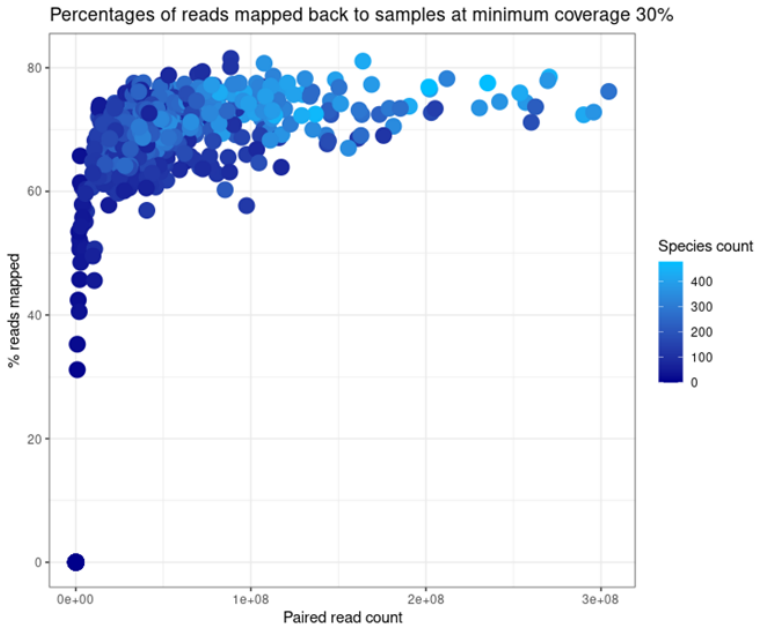


Figure S7:
Percentages of reads mapped back to samples at minimum coverage 30%. Apart from samples with low read depth, the percentage of mapped reads is consistent for all samples, even for those that did not contribute to the generation of the catalogue. A linear correlation between read count and species representation in samples can be observed. The 388 samples used for this study exhibited high and consistent metagenomic read mapping rates of around 60-80%.



Priority effects and microbial cross-feeding shape zoonotic agent spread in broiler chickens

Chapter 5

Abstract

Unravelling the colonisation dynamics and physiological effects of zoonotic bacteria such as *Campylobacter* is imperative to prevent foodborne diseases. We employed a hologenomic approach to jointly analyse metabolic networks and gene expression of the caecal microbiota, with the intestinal gene expression of 613 broiler chickens that did and did not undergo an opportunistic *Campylobacter* colonisation. We report that an early development of a distinct microbial enterotype enriched with *Bacteroides fragilis_A,* changed the community to a functional profile that likely benefited *Campylobacter* through production of key metabolites. The resulting enterotype was not associated with a host immune response, but exhibited an enriched and energetically more demanding functional repertoire compared to the standard enterotype, which could have caused the growth decline observed in *Campylobacter*-colonised animals. We provide unique insights into microbe-microbe and host-microbe interactions, which point to the early-stage microbiota-development as a relevant factor for later *Campylobacter* spread in broiler chickens.

Publication

Sofia Marcos[1,2], Iñaki Odriozola[2], Jorge Langa[1,2], Germana Baldi[3], Eray Sahin[4], Sarah Siu Tze Mak[2], Louisa Pless[2], Joan Tarradas[5], Andone Estonba[1], Antton Alberdi[2]

[1]Applied Genomics and Bioinformatics, University of the Basque Country (UPV/EHU), Leioa, Bilbao, Spain
[2]Center for Evolutionary Hologenomics, The Globe Institute, University of Copenhagen, Copenhagen, Denmark
[3]European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Hinxton, Cambridge, UK
[4]Biostatistics and Bioinformatics PhD Program, Institute of Health Sciences, Acibadem Mehmet Ali Aydinlar University, Istanbul, Turkey
[5]Animal Nutrition, Institute of Agrifood Research and Technology (IRTA), Constantí, Spain

Publication status: Under review
Link to preprint: https://www.researchsquare.com/article/rs-3588367/v1

Introduction

Zoonotic bacteria responsible for foodborne diseases represent a significant global concern due to their public health and economic implications (Abebe et al., 2020). However, the efficacy of prevention strategies is often hindered by a limited understanding of the precise colonisation dynamics and the physiological effects within vector hosts (Abd El-Hack et al., 2021; Sahoo et al., 2022). Fortunately, the emergence of multi-omic technologies has greatly enhanced our ability to bridge this knowledge gap and comprehend the ecological dynamics of zoonotic agents in source animals (Salmon-Divon et al., 2022). By facilitating the study of functional interactions among zoonotic bacteria, the animal host, and its associated microbial community, multi-omics offer invaluable insights crucial for the development of effective strategies to mitigate the impact of these diseases on human populations (Bäumler & Sperandio, 2016).

Campylobacteriosis is the most frequently reported zoonosis in the EU due to the presence of *Campylobacteraceae* strains in broiler chickens (Food Safety Authority, 2021). These bacteria are typically detected in chickens during the third week of their life, a period when the gut microbiota begins to stabilise after an initial rapid development characterised by significant species turnover (Ijaz et al., 2018; Rychlik, 2020). While *Campylobacter* has been considered a causative agent of microbiota rearrangements (Connerton et al., 2018; Thibodeau et al., 2015; Yan et al., 2021), the drivers of *Campylobacter* colonisation remain unclear (Awad et al., 2018). The metabolic auxotrophies of *Campylobacter* render it a potential scavenger of metabolic by-products produced by other bacteria (Garber et al., 2020; Luijkx et al., 2020). Thus, prior alterations in the biochemical conditions during initial microbial succession could drive spread of *Campylobacteraceae* at a later stage (Han et al., 2017; Ijaz et al., 2018; Kaakoush et al., 2014; Thibodeau et al., 2015; Yan et al., 2021). The construction of metabolic networks using metagenome-assembled genomes now allows for the study of such interdependencies (Belcour et al., 2020), enabling investigations into whether different paths of microbiota development can either facilitate or hinder the establishment and transmission of zoonotic strains.

The physiological effects of *Campylobacteraceae* on the host are also a subject of ongoing debate, as the literature presents contradictory findings regarding the relationship between *Campylobacter* and animal growth performance. While certain studies report no significant impact on chicken body weight (Connerton et al., 2018; Munoz et al., 2023), others describe weight loss after *Campylobacter* colonisation (Awad et al., 2015; Kollarcikova et al., 2019). There is also considerable variability in the host immune response to the presence of *Campylobacter (Awad et al., 2014; Connerton et al., 2018; Han et al., 2016; Humphrey et al., 2014)*. Furthermore, it remains unresolved whether these patterns are directly attributable to the action of *Campylobacter* itself or to the microbial community as a whole. The application of host and microbial (meta)transcriptomics enables us to delve into the actual interactions occurring between the two domains during different developmental stages.

In the H2020 project HoloFood (*HoloFood*, n.d.), we conducted three five-week-long experimental replicates to understand the effect of host-microbiota interactions in broiler

chicken production. In the last of the three trials, we detected an opportunistic colonisation of *Campylobacter* spp. in almost all chickens slaughtered after the third week (Tous et al., 2022). While the first two experimental trials allowed us to characterise the functional dynamics of the caecal microbiota development (Marcos et al., 2023), the third trial presented an exceptional opportunity to study the microbial and host alterations associated with *Campylobacter* colonisation with unprecedented resolution. We generated multiple omic data sets that encompassed host and microbial domains of the hologenomic landscape (Nyholm et al., 2020), namely genome-resolved metagenomics, microbial metatranscriptomics and chicken intestinal transcriptomics. By integrating those omic layers we compared the temporal development of the functional capabilities and activity of the caecal microbiota, along with the intestinal gene expression, of broiler chickens that experienced opportunistic *Campylobacter* colonisation and those that did not. We first analysed the alterations preceding *Campylobacter* spread using genome-scale metabolic networks (GSMNs), followed by host and microbial gene expression variation associated with the presence of *Campylobacter* and related bacteria.

Results

A distinct microbiota development precedes the spread of *Campylobacteraceae* bacteria

We analysed the microbial communities of 7-, 21- and 35-day-old chickens from three experimental replicates (trials A, B and C) by mapping 613 metagenomic datasets generated from caecum content samples to a bacterial genome catalogue generated from the same pool of animals (Marcos et al., 2023). Our metagenomic analysis confirmed the initial detection of *Campylobacter* spp. through PCR screening (Tous et al., 2022). All animals in trial C from day 21 onwards got colonised by at least one of the two *Campylobacteraceae* species, namely *C. jejuni* and *C. coli*. In addition, we detected another species belonging to the *Campylobacterales* order, *Helicobacter pullorum* (Fig. 1a). *C. jejuni* and *C. coli* represent prominent zoonotic agents responsible for human diarrheal diseases in industrialised and developing countries (Kaakoush et al., 2015), while *H. pullorum* is an emerging zoonotic pathogen linked to colitis and hepatitis in humans (Javed et al., 2017). Although chickens are predominantly colonised by *C. jejuni*, co-colonisation with *C. coli* and *Helicobacter* strains is often detected, which can lead to either commensalism or competition between them (Kaakoush et al., 2014; Rzeznitzeck et al., 2022).

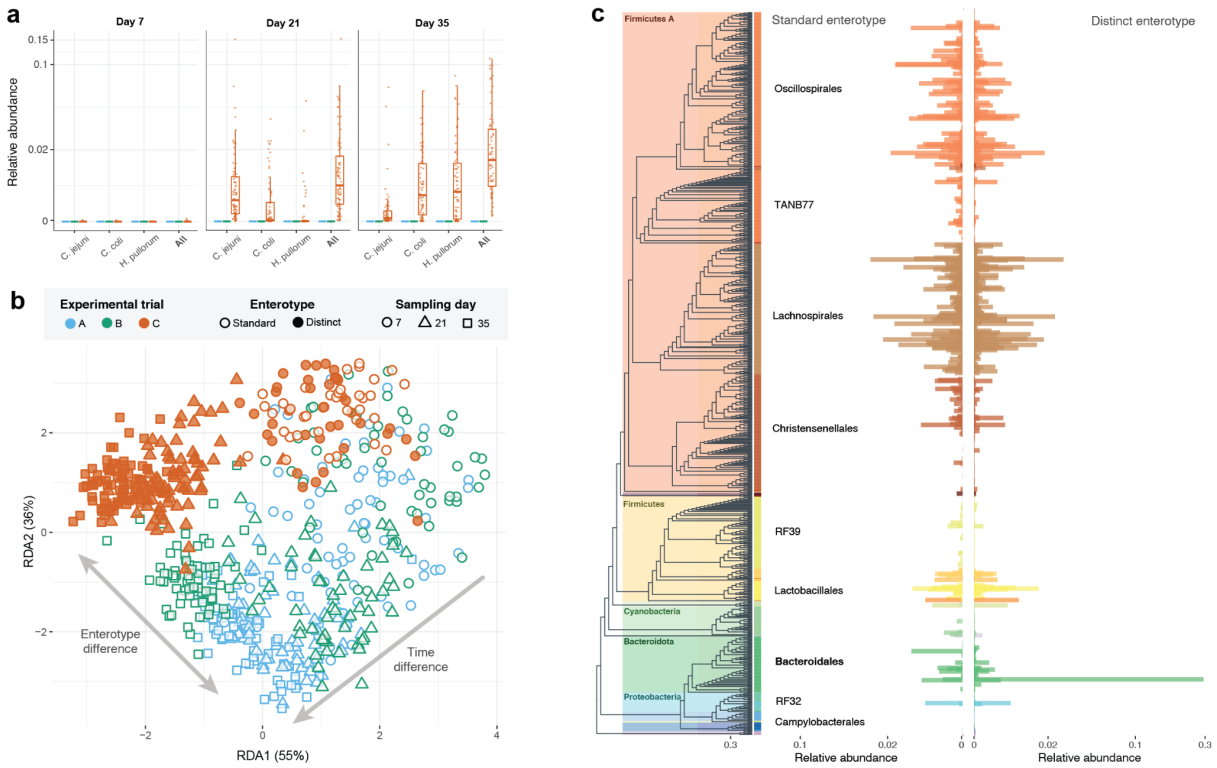


Figure 1. Overview of the two caecum microbiota enterotypes. a) Relative abundances of the three *Campylobacterales* strains across the three experimental trials and sampling points. **b)** Distance-based redundancy analysis of studied microbial communities, in which the two microbial enterotypes defined by Dirichlet Multinomial Models, the standard (empty shapes) and the distinct (solid shapes), are illustrated (R^2_{adj} = 0.34). Colours indicate experimental trials, while shapes indicate sampling times. Arrows denote the directionality of the variables enterotype (standard and distinct) and chicken age (days 7, 21 and 35) with respect to microbial composition. **c)** Cladogram based on the phylogenetic tree of the 822 caecum bacteria included in the study, with their average relative abundances in the two enterotypes at day 7. A detailed phylogenetic tree can be found in a previous publication by Marcos *et al*. (Marcos et al., 2023).

Community-level analyses revealed a robust association between the propagation of *Campylobacterales* strains in trial C and a distinctive microbiota development, contrasting with the trajectories observed in trials A and B (Fig. 1b). The distinct enterotype became evident at an early stage, manifesting noticeable differences in community composition (R^2 = 0.01, *F*-value = 2.02, *p*-value = 0.018) and functional profile (R^2 = 0.14, *F*-value = 49.05, *p*-value = 0.001) by the first sampling point at 7 days of age. This enterotype was characterised by an atypical dominance of *Bacteroidota*, primarily driven by *Bacteroides fragilis_A* according to Dirichlet Multinomial Mixture models (Fig. S1). *B. fragilis_A* was the most abundant bacteria in approximately half of the day-7 animals in trial C (Fig. 1c), in contrast to the standard microbiota development observed in trials A and B (Fig. S2). In those trials, *Bacteroidales* proliferated in the third week of age, after an initial period dominated by *Lactobacillus* and *Lachnospiraceae* clades, as previously reported in the

literature (Rychlik, 2020). *B. fragilis_A* is a non-spore-forming obligate anaerobe that usually colonises the chicken intestine at a later stage, likely due to its lower colonisation ability compared to spore-forming bacteria such as *Lachnospiraceae*, whose spores are widespread in the farm environment (Rychlik, 2020). Our observations are therefore in line with previous studies that reported a link between *Campylobacter* and obligate anaerobes within orders *Clostridiales* and *Bacteroidales* (Thibodeau et al., 2015; Yan et al., 2021)

To delve into the potential causes of the development of a distinct enterotype, we studied how the abundance of *B. fragilis_A* distributed across experimental groups at day 7. We found that *B. fragilis_A* was not randomly distributed, but aggregated in pens, as indicated by the 79% of the variance explained by the random effect of pens. However, the distribution was not associated with pen-specific chicken characteristics such as genetic line (*t*-value = -0.91, *p*-value =0.37) and sex (*t*-value = 0.24, *p*-value = 0.81). These observations suggest that the development of the distinct enterotype in trial C animals was primarily driven by the expedited colonisation of *B. fragilis_A* in a few animals, likely occurring before pen allocation, followed by subsequent transmission to pen-mates.

Bacteroides fragilis as a facilitator of *Campylobacter* colonisation

Despite inherent limitations in establishing causal associations from observational data, our investigation into bacterial metabolic dependencies unveiled several molecular mechanisms that shed light on how the early development of a *Bacteroides*-enriched enterotype might promote the subsequent proliferation of *Campylobacter* strains (Kaakoush et al., 2014; Yan et al., 2021). To quantify these metabolic dependencies, we employed 822 genome-scale metabolic networks (GSMNs) (Burgard et al., 2004) constructed using Pathway Tools (Karp et al., 2021), and grounded on EggNOG annotations (Huerta-Cepas et al., 2019). We categorised each metabolite for each bacterium as a source, transit, or sink metabolite based on its capacity for utilisation or production. *Source* metabolites are those that a bacterium can utilise but not produce, *transit* metabolites can be both produced and consumed, and *sink* metabolites are produced but not used (Fig. 2a).

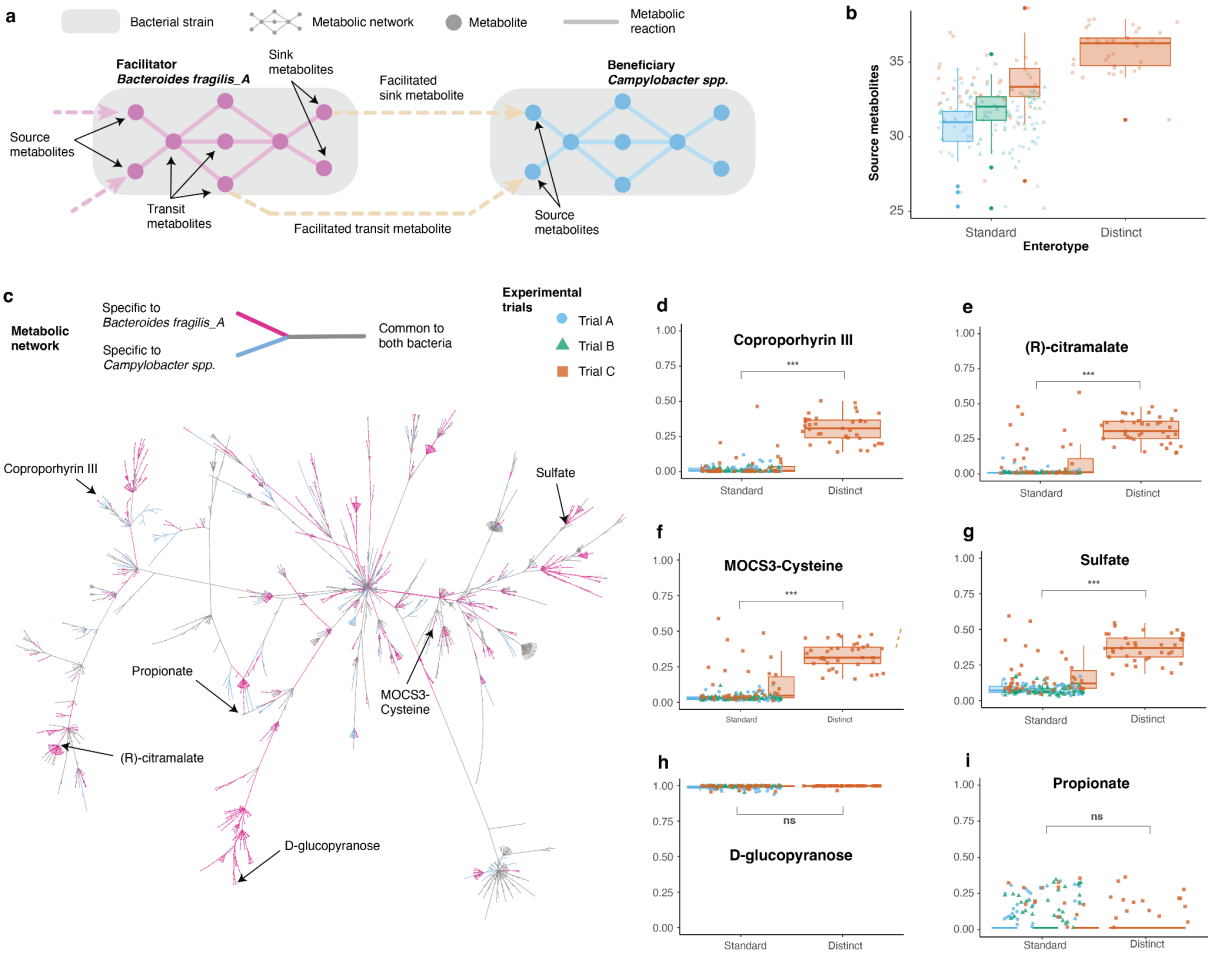


Figure 2. Enterotype contribution to *Campylobacter* metabolism. a) Conceptual representation of the metabolic dependencies between *Bacteroides fragilis_A* and *Campylobacter* spp. **b)** Community-weighted average of the number of *Campylobacter* source metabolites each enterotype is capable of producing. **c)** Spanning tree of the joint genome-scale metabolic network model between *Bacteroides fragilis_A* and *Campylobacter jejuni*. Nodes indicate metabolites, while branches indicate reactions. Grey branches indicate metabolic reactions that are present in both genomes, pink branches are only present in *B. fragilis_A* while blue branches are only present in *C. jejuni*. **d-i)** Cumulative relative abundance of bacteria capable of producing specific metabolites in the standard and distinct enterotypes. Colours indicate experimental trials, and boxplots are specific to each trial and enterotype.

We determined the existence of 385 and 395 source metabolites for *C. jejuni* and *C. coli*, respectively, of which 32.2% and 32.7% could be produced by other members of the community. Both strains shared 95% of their metabolic networks, diverging only in *C. coli*'s superior capability to metabolise some by-products such as glutathione, succinate, and oxaloacetate (Fig. S3). To assess whether the distinct enterotype conferred metabolic advantages to *Campylobacter*, we calculated the weighted capacity of each enterotype to produce those source metabolites at day 7. Our findings revealed that the distinct enterotype exhibited a higher capacity compared to the standard one (LMM, estimate = 2.23, t-value = 5.43, p-value < 0.01) in producing source metabolites that *Campylobacter* cannot synthesise

on its own (Fig. 2b). Broadening our analysis, we observed that 1,043 out of the 4,533 identified metabolites were overrepresented in the distinct enterotype, encompassing 31 and 35 source metabolites for *C. jejuni* and *C. coli*, respectively, which can be potentially produced by the rest of the community (Table S1).

In vitro assays suggested *Bacteroides* as a potential facilitator of *Campylobacter* colonisation via the provision of free sugars and short-chain fatty acids (SCFAs) (Fan et al., 2023; Garber et al., 2020; Luijkx et al., 2020). Our joint GSMN of *B. fragilis_A* and *C. jejuni* highlighted numerous ways in which *Bacteroides* could contribute to *Campylobacter* through the production of relevant metabolic by-products (Fig. 2c). However, we found no evidence of an enhanced genomic capacity of the distinct enterotype for polysaccharide degradation (Fig. 2h) or SCFA production (Fig. 2i). As the dominant taxa in the standard enterotype (e.g., *Lachnospirales* and *Oscillospirales*) also possess these metabolic attributes (Vacca et al., 2020), it is unlikely that *Campylobacter* colonisation was primarily linked to these metabolites.

Nonetheless, we identified other metabolites that were likely to play pivotal roles in the *Bacteroides*-*Campylobacter* interaction. The two most relevant source metabolites for *Campylobacter* were coproporphyrin III and (R)-citramalate, as they stood out due to their pronounced differences between enterotypes (Figs. 2d, 2e) and their classification as sink metabolites for *B. fragilis_A*. This indicates that *B. fragilis_A* likely overproduces these metabolites, which may become available for *Campylobacter*. Coproporphyrin III is an essential component of one of the three haem biosynthesis pathways. Haem is an iron-chelated modified tetrapyrrole and is a key compound for proteins involved in several essential cellular processes (Zamarreño Beas et al., 2022). (R)-citramalate is a metabolic intermediate that participates in the synthesis of tricarboxylic acids (Petushkova et al., 2021) and is known to be a substrate of the alternative threonine-independent isoleucine synthesis pathway (Risso et al., 2008). We quantified gene expression and verified the utilisation of both metabolites by *Campylobacter* strains (EC:4.99.1.9 and EC:4.2.1.35) and the production of coproporphyrin III by *B. fragilis_A* (EC:1.3.3.15).

Two other source metabolites for *Campylobacter*, MOCS3-Cysteine and sulphate, were also disproportionately prevalent in the distinct enterotype (Fig. 2f, 2g). However, unlike the previously mentioned metabolites, *B. fragilis_A* has the capacity to utilise them. MOCS3-Cysteine, a sulphur transferase enzyme crucial for molybdopterin biosynthesis, plays a pivotal role in the formation of redox enzymes (Mendel & Leimkühler, 2015). Sulphate can be reduced to hydrogen sulphide, required for cysteine synthesis (Kredich, 1992). Despite reported auxotrophies related to sulphate assimilation in *Campylobacter* (Man et al., 2020), our strains exhibited gene expression for the enzymes responsible for consuming these metabolites (EC:2.7.7.4 and EC:2.8.1.7, respectively). Our results therefore suggest that priority effects, whereby the order of microbial species colonisation influences longer term microbiome composition, likely play a central role in shaping temporal *Campylobacter* dynamics. Specifically, the *Bacteroides*-dominated enterotype creates a metabolically favourable environment for *Campylobacter* establishment and colonisation, likely facilitated by acquisition of compounds involved in central metabolic processes.

The distinct enterotype correlates with host body weight

Chickens from trial C not only underwent a distinct enterotype development followed by the spread of *Campylobacterales* strains, but also exhibited a significantly reduced growth performance as compared to chickens from trials A and B (Fig. S4). Unlike in humans, *Campylobacterales* strains do not cause disease symptoms in chickens (Wyszyńska & Godlewska, 2021). However, impaired performance has been observed in multiple trials, which has fueled discussions about strain-specific mechanisms by which *Campylobacter* could affect chicken growth (Awad et al., 2018; Wyszyńska & Godlewska, 2021). In light of this, we posed two non-exclusive hypotheses on how microbe-host interactions could have contributed to the reduction of animal growth: i) colonisation by *Campylobacterales* triggers an inflammatory response, which hinders the correct functioning of the intestine and deviates energy from growth to immunity (Humphrey et al., 2014); ii) the distinct enterotype is functionally different, which affects host energy balance (Marcos et al., 2023).

To assess whether the distinct enterotype triggered a persistent pro-inflammatory response in the host's intestine, we performed differential expression analyses between the two enterotypes. The study of 169 host transcriptomic datasets derived from caecal mucus samples collected at three distinct time points, revealed no substantial differences in the expression profiles between animals hosting each enterotype. Although we detected the largest difference at day 35, with 36 differentially expressed genes (Fig. S5, Table S2), no clear Gene Ontology or KEGG pathway enrichment could be observed, thus yielding no evidence of an inflammatory response from the host. The critical window in the immune cell development is identified between days 14 and 28, in which certain bacteria play a key role in their maturation process (Liu et al., 2023). *Campylobacter* is recognised by Toll-like receptors and can induce an inflammatory response by increasing expression of cytokines and immune-associated genes (Awad et al., 2014; Connerton et al., 2018; Humphrey et al., 2014). In fact, enzyme immunoassay conducted in blood did detect a significant peak of C-reactive protein in the distinct enterotype at day 21 (Tous et al., 2022). This time point coincides with the initial detection of *Campylobacter,* which pointed towards a possible response from the host. Nonetheless, neither the rest of inflammation (haptoglobin-like protein) nor stress (corticosterone) biomarkers analysed in the same animals pointed towards a significant inflammatory response (Tous et al., 2022). We therefore deem unlikely that the observed growth deceleration in chickens from trial C was due to an immune response towards *Campylobacteraceae* bacteria.

Instead, we hypothesised that the lower body weights associated with the distinct enterotype could result from a heightened metabolic capacity of the microbial communities. An increased metabolic demand of the caecal microbiota might cause the microbial community to compete for resources with the host, restricting the host’s absorption of nutrients in the proximal part of the caecum (Marcos et al., 2023). Once validated that the distinct enterotype had higher metabolic capacities than the standard in the first weeks of the trials (Fig. S6), we explored whether these capacities were actually realised. For this purpose, we compared the metabolic activity of both enterotypes across the three time points. We distilled microbial gene expression data from 125 microbial metatranscriptomic datasets into 170 quantitative

Genome-Inferred Functional Traits (GIFTs) (Table S3) per genome, by pondering gene expression values according to the weight of each gene in each metabolic pathway. Community level analyses showed that the distinct enterotype tended to overexpress genes involved in organic anion biosynthesis (B06) and, particularly, nitrogen compound degradation (D06) already from day 7, but exacerbated at days 21 and 35 (Figs. 3a, 3b, 3c). *Bacteroidales* and *Campylobacterales* were the main contributors to the B06 and D06 function groups (Fig. 3d), as different strains of the phylum *Bacteroidota* emerged at days 21 and 35 (Fig. S2). In addition, the decline of *Oscillospirales* and *Lachnospirales* clades in the distinct enterotype caused a reduction of amino acid derivative biosynthesis (B03) and lipid degradation (D01). The D06 function group consists mainly in nitrate, urate, taurine and hypotaurine degradation. B06, although comprising biosynthesis pathways, is derived mainly from degradation of lipids, proteins and carbohydrates through Krebs cycle and other processes to produce succinate, fumarate and citrate, which together with D06 points towards a higher catalytic activity of microbes. Our results are in line with previous studies which reported that the reduced body weight gain in *Campylobacter*-colonised chickens could be due to the extensive amino acid utilisation by *Campylobacter,* which caused lower concentrations of amino acids in the ileum, and a reduced expression of peptide and amino acid transporters in the caecum (Awad et al., 2014, 2015). Nevertheless, our data suggests that such heightened energy utilisation should not only be attributed to *Campylobacter*, but also extended to *Bacteroidales*strains, which collectively contributed to the increased metabolic action of the distinct enterotype.

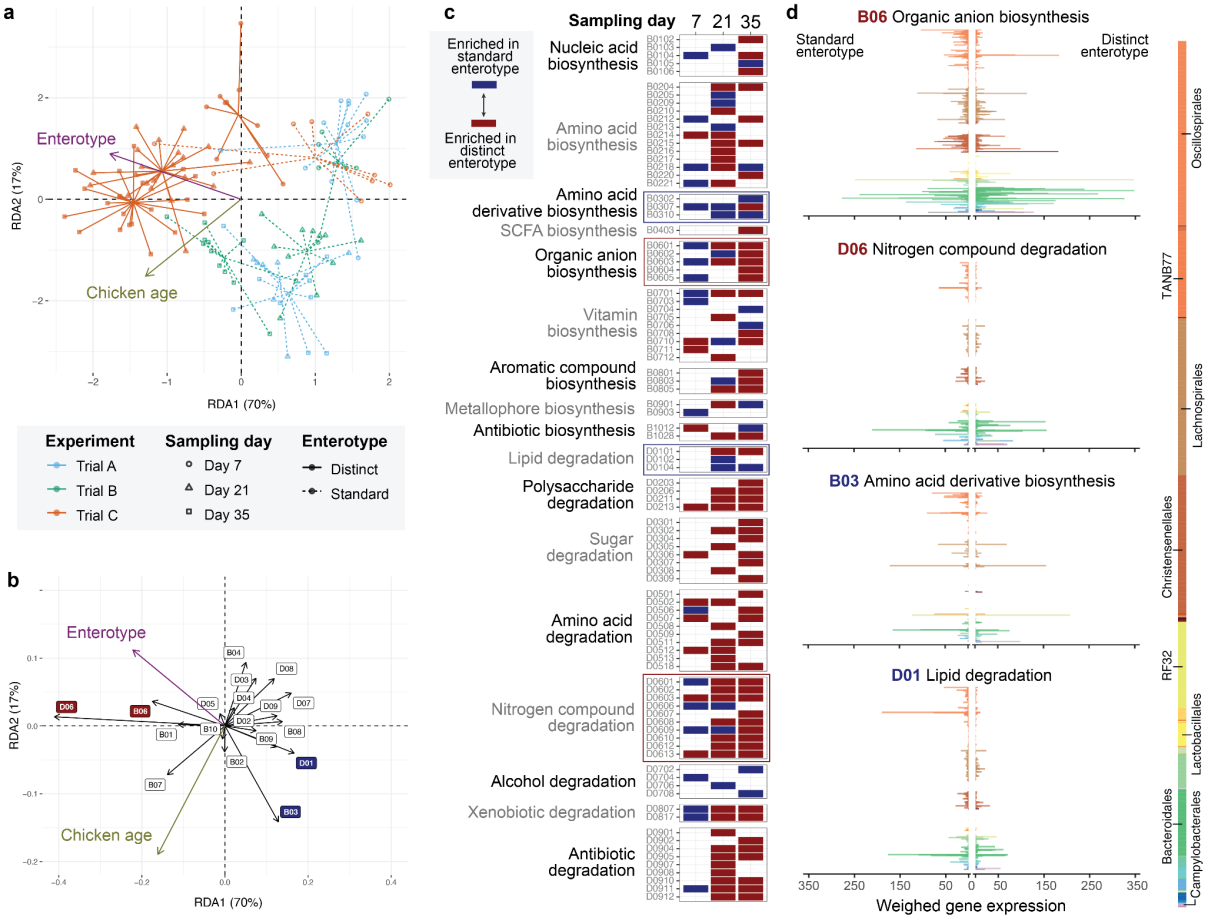


Figure 3. Functional activity of enterotypes. a) Distance-based redundancy analysis of the microbial expression profiles in which two microbial enterotypes defined by Dirichlet Multinomial Models, the standard and the distinct, across the three sampling points are shown (R^2_{adj} = 0.25). **b)** The same distance-based redundancy analysis in which the arrows illustrate the expression of function groups. The length of the arrow shows the strength of the correlation between function groups and microbial expression. The four functions with the highest effect sizes (D06, B06, D01 and B03) are highlighted in red (positive effect towards the distinct enterotype) and blue (negative effect towards the distinct enterotype). Green and purple arrows denote the directionality of the variables enterotype (standard and distinct) and chicken age (days 7, 21 and 35). **c)** Differentially expressed functions across sampling times in distinct communities. Blue boxes indicate functional pathways over-expressed in the standard enterotype, while red boxes indicate pathways over-expressed in the distinct enterotype. The coloured contours of D06, B06, D01 and B03 highlight the four functions showcased in figure d. **d)** Expression profiles of the 822 caecum bacteria in the two microbial enterotypes at day 35. The expression of four relevant function groups were illustrated B06 (organic anion biosynthesis), D06 (nitrogen compound degradation), B03 (Amino acid derivative biosynthesis), and D01 (Lipid degradation).

Discussion

Foodborne zoonotic bacteria not only give rise to infectious diseases in humans and have environmental implications, but also impose significant economic and resource burdens on the meat industry (Smith et al., 2019). Nevertheless, current methods for prevention and

early detection of zoonotic agents often fall short, partly due to our limited knowledge about their interactions with the rest of the gut microorganisms and host animals. Our multi-omic study, focusing on the temporal development of functional capacities and activity within *Campylobacterales* bacteria and the rest of the microbiota, unveiled numerous novel insights into these intricate interactions. We observed that the distinct enterotype preceding the widespread emergence of *Campylobacterales* demonstrated a heightened ability to meet the metabolic demands of *Campylobacter* spp., in contrast to the standard enterotype associated with *Campylobacter*-free animals. While it is worth noting that the enterotypes presented in this study are likely only two of many possible microbiota development trajectories (Kaakoush et al., 2014; Thibodeau et al., 2015; Yan et al., 2021), our findings suggest that metabolic interdependencies and priority effects significantly influence the likelihood of *Campylobacter* colonisation within chicken gut environments. Notably, the emergence of this distinct enterotype primarily led by *Bacteroides fragilis_A* opens the possibility of using the early presence of this bacteria as a biomarker for subsequent *Campylobacter* colonisation. While manipulation of the early-life microbiota followed by experimental infection with zoonotic strains will be necessary to validate our results, these findings pave the way for exploring strategies to manipulate early-life microbiota compositions that minimise metabolic advantages for *Campylobacter*.

The colonisation of *Campylobacter* spp. was associated with a marked reduction of body weight in the studied chickens. However, neither the gene expression analyses conducted on intestinal mucosal samples, nor the examination of numerous complementary markers, revealed clear indications of a systemic immune response to the presence of *Campylobacter* spp. or the distinct enterotype. We nevertheless observed notable disparities in the metabolic capacities and activities of bacteria in both enterotypes, with the distinct enterotype exhibiting enhanced activity across most metabolic domains, with a particular emphasis on nitrogen compound utilisation. These results align with previous observations that indicated a negative impact of increased metabolic activity on animal growth, likely stemming from increased competition for nutritional resources (Marcos et al., 2023). Therefore, observed correlations between the presence of *Campylobacter* spp. and reduced body weight, in the absence of an inflammatory response, may be attributed to a distinct metabolic activity of the associated microbiota.

In summary, our study underscores the importance of studying zoonotic bacteria, their accompanying microbiota and the host organism in combination, all while harnessing the power of multi-omic technologies. Only through the high-resolution functional analysis of the three mentioned elements will we be able to resolve the complex tripartite interactions between them, and in doing so gain knowledge to develop novel sustainable strategies to improve safety and sustainability of poultry production.

Methods

Animal experiments

A total of 613 animals were sampled in the three experimental replicates (trials A, B, and C) performed in 2019 within the H2020 project HoloFood (*HoloFood*, n.d.); 205, 182 and 226 birds from A, B and C trials respectively. Broiler chickens from two genetic lines (Cobb 500 and Ross 308) and both sexes were reared in intensive farm conditions for 35-37 days. Each trial comprised 24 pens (12 groups replicated twice distributed in 3 treatments x 2 genetic lines x 2 sexes), each pen containing 40 animals. More details about the experimental design, diet and performance results are available in Tous et al. (Tous et al., 2022). Chickens were euthanized, weighed, and sampled at days 7–8, 21–22 and 35–37 (multiple days were necessary due to workload, and these differences have been accounted for in the statistical analyses), hereafter simplified to three time points (days 7, 21 and 35). Caecal pathogens detection (*Salmonella* spp., *Campylobacter* spp., and *Clostridium* spp.) procedures using PCR are explained in detail in Tous et al. (Tous et al., 2022). Molecular data was obtained from three different sections of the caecum. In short, the end of one of the caecum bags was isolated and longitudinally opened to gently collect ca. 100 mg of digesta for each metagenomic and metatranscriptomic analysis. After carefully washing the intestinal surface with saline solution, the mucosal layer was scraped and ca. 100 mg of mucosa were collected for host transcriptomic analyses. All types of samples were preserved in DNA/RNA Shield buffer (Zymo) and stored at -20 ºC until nucleic acid extractions.

Data generation

DNA and RNA extraction

A total of 613 metagenomic, 125 metatranscriptomic and 169 host transcriptomic datasets were generated. Both nucleic acids were extracted using a custom purification method optimised for samples preserved on DNA/RNA Shield buffer (Bozzi et al., 2021). The protocol consisted in a bead-beating for tissue disruption, followed by digestion, nucleic acid separation (DNA and RNA) and purification steps. Samples were processed in batches of 90 samples, along with 6 extraction, library preparation, and library indexing blanks (2x2x2). Samples within each batch were randomised using a custom script, but different sample types were not mixed to minimise the risk of cross-contamination due to DNA concentration differences.

Library preparation of metagenomic DNA

The nucleic acids extracted were fragmented to obtain an average length of 400 bp using a Covaris LE220 ultrasonication device. A standard amount of 200 ng of DNA was used for the library preparation. We used the BEST (Carøe et al., 2018) ligation-based library preparation protocol to prepare sequencing libraries. In order to evaluate the success of the libraries, we conducted quality controls using qPCR assays. The optimal number of cycles was estimated to achieve the desired DNA molarity while reducing clonality. Any libraries that exceeded 12 cycles were repeated for library preparation due to potential technical biases. Subsequently,

libraries were indexed using unique dual tags, along with the necessary PCR cycles. Before final quality-checks were performed with a DNA Fragment Analyser (Agilent), bead purification was carried out. Libraries with expected fragment-size distributions and molarities were equimolarly pooled for sequencing. Libraries that showed too low molarities were re-indexed to achieve the desired molarity, and the ones exhibiting unusual fragment distributions and large adaptor dimers were re-built. Sequencing was performed in multiple BGIseq runs with 150bp paired-end chemistry. Sequencing effort per sample typically varied between 8GB and 16GB, equivalent to 26 and 52 million reads.

Library preparation of metatranscriptomic RNA

rRNA depletion was performed using TIANSeq rRNA Depletion Kit (Animal) (Cat.No. NR101-T1), and the remaining RNA was fragmented into 250-300bp, to finally reverse-transcribe into double stranded cDNA with random hexamers. Total cDNA was sent to Novogene for library preparation and sequencing. In short, cDNA libraries were constructed using Novogene NGS RNA Library Prep Set (PT042), which comprises end repair, A-tailing and adapter ligation steps. The libraries were checked with Qubit and real-time PCR for quantification and Bioanalyzer (Agilent) for size distribution detection. Quantified libraries were pooled and sequenced on an Illumina NovaSeq 6000 platform with 150bp paired-end chemistry, aiming for 5GB of protein-coding gene data.

Library preparation of chicken transcriptomic RNA

Total RNA was quantified using Nanodrop (Thermo Scientific) and Bioanalyzer 2100 (Agilent), as well as analysed for integrity (Agilent 2100) and purity (agarose electrophoresis and Nanodrop). Samples were subjected to rRNA removal step by poly-A enrichment, using poly-T oligo-attached magnetic beads. After fragmentation, the first strand cDNA was synthesised using random hexamer primers. During the second strand cDNA synthesis, dUTPs were replaced with dTTPs in the reaction buffer. Total cDNA was sent to Novogene for library preparation and sequencing. The directional libraries were ready after end repair, A-tailing, adapter ligation, size selection, USER enzyme digestion, amplification, and purification. Libraries were checked with Qubit and real-time PCR for quantification, as well as with bioanalyzer for size distribution detection. Quantified libraries were pooled and sequenced on NovaSeq 6000 (Illumina), according to effective library concentration and data amount required.

Bioinformatic data processing

Generation of the MAG catalogue

Details on the procedures employed to create the MAG catalogue used in this study will be published, and code can also be accessed at Workflowhub (https://workflowhub.eu/programmes/28). In short, data from 261 caecal metagenomic samples collected from chickens from the three experiments were used to generate the caecal MAG catalogue. We performed *de novo* metagenomic assemblies using the MGnify assembly pipeline (Richardson et al., 2023). The assembly tool MetaSPAdes (Nurk et al., 2017) was used preferentially for single-run assemblies, while MEGAHIT (D. Li et al., 2015) was used for co-assemblies when memory requirements for MetaSPAdes were too high.

Samples prioritised for co-assembly were selected by hierarchical clustering based on Jaccard distance between low-quality bins generated by single assembly. Contigs shorter than 500 base pairs were excluded, and further host, human and PhiX decontamination was performed post-assembly with blastn (Y. Chen et al., 2015). Contig binning was performed using 'binning' and 'bin_refinement' modules of metaWRAP's. Genome quality assessment was conducted using checkM (Parks et al., 2015), with the criteria of retaining genomes with completeness >50%, contamination <5%, and a quality score (QS) >50 (where QS = completeness - 5*contamination). Genomes were de-replicated using an Average Nucleotide Identity (ANI) of 95%, and 30% alignment fraction to generate species-level clusters using dRep (Olm et al., 2017). Lastly, GUNC (Orakov et al., 2021) was employed to identify potentially chimeric genomes for subsequent removal, utilising specific parameters that included a clade separation score >0.45, contamination >0.05, and reference representation score >0.5.

Functional annotation and distillation of MAG catalogue

Taxonomy annotation and phylogenetic tree construction was carried out using GTDB-Tk (Chaumeil et al., 2019). Functional annotation of the MAGs was performed through an ensemble approach implemented in DRAM (Shaffer et al., 2020). This approach incorporates data from various databases, including Pfam (Mistry et al., 2021), KEGG (Kanehisa & Goto, 2000), UniProt (UniProt Consortium, 2019), CAZY (Cantarel et al., 2009) and MEROPS (Rawlings et al., 2010). To distil these annotations into quantitative genome-inferred functional traits (GIFTs) representing metabolic capacities provided by the microbiota to its host, we used the R package DistillR, which can be found at the following link: (https://github.com/anttonalberdi/distillR). DistillR contains a set of >300 metabolic curated metabolic pathways and modules derived from KEGG and MetaCyc (Karp, Riley, et al., 2002) databases, which are used to obtain quantitative estimates of the metabolic capacities of microorganisms through quantifying the relative representation of genes required for accomplishing a metabolic task. GIFTs range between 0-1, the zero indicating none of the genes defined in the pathway are present in the genome and one indicating that all genes are present. In cases where a step within a pathway requires the presence of two Identifiers, the step is considered full if both Identifiers are present, half full if one is present, and empty if none is present. We quantified 170 GIFTs per genome (complete detailed list can be found in Table S1), whose values were first corrected by MAG genome completeness to reduce functional biases (Eisenhofer et al., 2023), and then averaged to obtain a genome-level overall metabolic capacity metric, hereafter referred to as Metabolic Capacity Index (MCI). We also distilled microbial gene expression data into 170 GIFTs per genome, by weighing gene expression values according to the weight of each gene in each metabolic pathway.

Genome-Scale Metabolic Networks

A genome-scale metabolic network (GSMN) is a comprehensive representation of all the metabolic reactions that occur within an organism. It is constructed based on the genomic information of the organism and integrates biochemical and genetic knowledge to capture the complexity of the organism's metabolism. We employed the software metage2metabo

(Belcour et al., 2020), which in turns relies on Pathway Tools (Karp, Paley, et al., 2002) to reconstruct the GSMNs of every bacterial genome in our study using a custom snakemake (Mölder et al., 2021) pipeline. Shortly, due to software dependencies, bacterial genomes were re-annotated using eggnog-mapper2 (Cantalapiedra et al., 2021) against the eggNOG 5.0 database (Huerta-Cepas et al., 2019). The annotation files were transformed into Genbank annotation files (gbk) using 'emapper2gbk' and SBML files generated using 'm2m recon', as implemented in metage2metabo. SBML files were analysed using the package COBRApy (Ebrahim et al., 2013) and custom python scripts to quantify source, transit and sink metabolites as explained below.

Metagenomic data processing and read mapping

Sequencing adapters and identical duplicates were filtered out using AdapterRemoval 2.2.4 (Schubert et al., 2016) and seqkit 0.7.1 (Shen et al., 2016). Sequences were mapped to the latest chicken reference genome (galGal6, NCBI Assembly accession GCF_000002315.6) using bwa (H. Li & Durbin, 2009) increasing the minimum seed length to 25 to minimise the likelihood of incorrect read pair alignments from the metagenomic fraction. To evaluate the quality of the alignment, mapping statistics including depth and breadth of coverage, and percentage of mapped reads were calculated using SAMtools 1.11 (H. Li et al., 2009). Aligned reads were sorted and the metagenomic fraction was isolated using SAMtools. Metagenomic reads were mapped to the MAG catalogue using bwa at 90% identity and 60% coverage threshold and further summarised with samtools. Read-mapping counts resulting in < 30% genome coverage per sample were removed from further analysis. Retained read mapping counts were divided by the total number of paired-reads per sample, and multiplied by 100 to give the percentage of reads mapped to the MAG catalogue for each sample. Relative abundance was estimated by adapting the RPKG (Reads Per Kilobase per Genome equivalent) formula provided by Nayfach and Pollard. It is referred to as RPMM (Reads Per Million bases of genome, per Million mapped reads), as reads mapped to MAGs were normalised both by genome length (divided by 1M) and by read length (divided by 1M).

Metatranscriptomic data processing and read mapping

We employed a custom snakemake pipeline for preprocessing metatranscriptomic data (https://github.com/anttonalberdi/holoflow/tree/EisenRa/workflows/metatranscriptomics). In short, reads were trimmed and quality controlled using fastp (S. Chen et al., 2018), keeping reads >60 bp and with Phred scores >20. Processed reads were then mapped against the host genome (galGal6) using STAR (Dobin et al., 2013). The unmapped reads were subsequently mapped to a combined database containing SILVA 16S rRNA SSU and LSU NR 99 (Quast et al., 2013), as well as 5SRNAdb (Szymanski et al., 2016) using Bowtie2 (Langmead & Salzberg, 2012) with default parameters. Unmapped reads were then mapped to the MAG catalogue genes (outputted from DRAM; genes.fna.gz) using Bowtie2 with default parameters. Finally, gene read counts were calculated using CoverM (https://github.com/wwood/CoverM), requiring both pairs of reads to hit the gene (--proper-pairs-only flag).

Chicken transcriptomic data processing and read mapping

Raw transcriptomic reads were quality-filtered using fastp, mapped against the host reference genome using STAR, and gene count data extracted using the gene count option. Each sample yielded on average 12.5±3.2 million reads against the 24,131 genes annotated in the chicken reference genome (galGal6).

Data analysis

Metagenomic data analysis

Metagenomic counts were standardised by MAG length and sequencing depth. Dirichlet Multinomial Mixtures (DMM) (Holmes et al., 2012) were utilised to profile and identify enterotypes in chicken microbial communities. Models were run by setting the maximum allowed number of community types to 5. A total of three runs were performed, one for each sampling day. At day 7, two community profiles were defined, where half of individuals from trial C formed an enterotype, and the rest of individuals from trials C grouped together with animals from trials A and B, forming another enterotype. At day 21, the model detected two clearly defined enterotypes, where in one of them all animals from trials A and B clustered together, and in the other enterotype all animals from trial C. At day 35, three enterotypes that were consistent with trials were detected. Thus, we defined microbial enterotypes as the distinct and standard enterotypes. The *distinct* enterotype comprised chickens from trial C that grouped separately from the rest at day 7, and the rest of the chickens from trials C at days 21 and 35. The rest of enterotypes were grouped together under the *standard* enterotype. Top community driver bacteria (i.e. MAGs with highest contribution to discriminate between enterotypes) were identified by selecting the 3% of MAGs with the highest posterior probabilities in DMM analysis.

To assess the temporal development of the composition of microbial communities across time, the MAG sequence count table was transformed using centred log-ratio (CLR) (Lahti et al., 2017) and submitted to the constrained ordination from R package vegan (Oksanen et al., 2022). MAG sequence count table was constrained by the factor trial (categorical variable with three levels: trials A. B and C), sampling time (categorical variable with three levels: days 7, 21 and 35) and their interaction. The significance of the constraining variables was tested using 999 permutations. To test the null hypothesis of no differences between enterotypes in microbial composition and function at day 7, PERMANOVAs were fitted through the ‘adonis2’ function of R package vegan. Euclidean distance matrices of CLR-transformed microbial abundances and functional profiles were included as responses in the PERMANOVAs and trial, enterotype, chicken age, sex, genetic line and treatment were included as explanatory variables. *P*-values were generated with 999 permutations.

A cladogram derived from the GTDB (Parks et al., 2022) tree constructed by GTDB-tk for taxonomic annotation was built with R package ggtree (Yu et al., 2017). Tips of reference genomes were pruned using the ‘keep.tips’ function included in the R package ape (Paradis et al., 2004). Relative abundances of each bacterial genome of the catalogue for both

enterotypes at each sampling time were illustrated with barplots after counts were standardised by MAG length and sequencing depth.The order of the bacteria was based on the tree obtained with GTDB.

To assess the drivers of the relative abundance of *B. fragilis_A* (the main indicator of the distinct enterotype at day 7) on trial C and day 7 we used linear mixed effect models as implemented in the R package lme4 (Bates et al., 2014). *P*-values for the fixed effects were computed with the R package lmerTest (Kuznetsova et al., 2017). CLR-transformed abundance of *B. fragilis_A* was used as response variable and trial, treatment, chicken age, sex and genetic line were included as fixed explanatory variables. To account for the fact that chickens were nested within pens we included a pen-level random intercept (1|pen). Then, we calculated the marginal and conditional R^2 using the R package MuMIn (Barton, 2009): marginal R^2 captures the variance explained by fixed effects whereas the conditional R^2 quantifies the variance explained by fixed and random effects together. Therefore, the variance associated with random effects (between-pen variance in relative abundance of *B. fragilis_A*) was calculated by subtracting the marginal from the conditional R^2.

Genome-Scale Metabolic Network analysis

General statistics of metabolic properties were calculated for each bacterial genome using custom python functions. These included listing and quantifying source, transit and sink metabolites. *Source* metabolites were defined as metabolites that a given bacteria is able to use as reactant in at least one metabolic reaction inferred from the genomic information, but that the bacteria is unable to produce itself. Therefore, source metabolites have to be acquired from elsewhere, and can potentially be provided by other bacteria. *Transit* metabolites were defined as any metabolite that a bacteria can produce and utilise, while *sink* metabolites are a special case of transit metabolites, defined as metabolites that a bacteria can produce but is unable to use itself. Sink metabolites and, potentially transit metabolites, are therefore metabolic by-products that may become available for other bacteria.

Using the GSMNs of all 882 characterised bacteria, we calculated the community-weighted average number of source metabolites for *Campylobacter* that the microbial community associated with each chicken at day 7 could potentially produce. Additionally, we calculated the capacity of each enterotype to produce specific source metabolites for *Campylobacter*. We did so by calculating the cumulative relative abundance of bacteria in each sample collected at day 7, capable of producing each metabolite. First, to test the null hypothesis of no difference between enterotypes to produce source metabolites for *Campylobacter*, we used linear mixed models as implemented in the R package lme4. The capacity of each enterotype to produce source metabolites was used as response variable and enterotype, trial, treatment, chicken age, sex and genetic line were included as fixed explanatory variables. A pen-level random intercept (1|pen) was used to account for the nested design. The assumptions of homoscedasticity and normal distribution of errors were verified with visual inspection of residual plots. Then, the null hypothesis of no difference in capability of producing specific source metabolites for *Campylobacter* between enterotypes was tested

using generalised linear mixed models using the function glmmPQL() of R package MASS (Venables & Ripley, 2002). In this case, the response variables were fractional (i.e. they take values between 0 and 1), thus we used a quasibinomial distribution with logit link function, which allowed us modelling the fractional response variables while accounting for under- or overdispersion and obtaining robust standard errors (Papke & Wooldridge, 1996). The set of fixed explanatory variables and random effects were the same as in above linear mixed models. Since multiple metabolites were tested consecutively, *P*-values were corrected for multiple testing using the Benjamini-Hochberg false discovery rate procedure (Benjamini & Hochberg, 1995).

Metatranscriptomic data analysis

Quantitative GIFTs were calculated with 'distillq' function using the R package distillR. To explore the temporal development of the functional expression profile of the standard and distinct chicken caecum enterotypes, the CLR-transformed community level quantitative GIFT profiles were ordinated with the constrained ordination RDA using 'rda' command from R package vegan. The ordination was constrained with the continuous variable chicken age, the categorical variable enterotype and their interaction. The significance of the factors was assessed using 999 permutations.

To identify specific quantitative GIFTs differentially expressed between standard and distinct chickens at different time points, linear mixed effect models were used with the R package lme4. CLR-transformed community level quantitative GIFTs were used as response variables in linear mixed models. As fixed explanatory variables in the models we used the enterotype, chicken age, sex, genetic line and treatment, and a pen-level random intercept (1|pen) was included to account for the nested design. *P*-values were corrected for multiple testing using the Benjamini-Hochberg false discovery rate procedure.

Lastly, to assess which bacterial strains were contributing the most to the expression of specific functions at different time points and enterotypes, we calculated the average relative expressions (given as gene counts per million) of specific functions for each MAG, at different combinations of enterotype and sampling time.

Chicken transcriptomic data analysis

Gene counts were processed with the tidybulk R package (Mangiola et al., 2021). Briefly, gene counts were imported with the tidyverse metapackage (Wickham et al., 2019). Then, counts were normalised using the TMM method from edgeR (Robinson et al., 2010). Samples were clearly differentiated by sex and age. Thus, sex was a controlled variable for subsequent analyses. No statistical differentiation between breeds could be observed. We set as confounding variables the sex, breed, laboratory (two sample extraction batches) and sequencing batch (three sequencing batches). Then, samples were compared for differential expression between trials A and B versus C at the three sampling times. The method chosen for differential expression was the one implemented in edgeR. p-values were corrected for multiple testing using the Benjamini-Hochberg method (Benjamini & Hochberg, 1995). Analysis of overrepresented Gene Ontologies (Gene Ontology Consortium, 2021) and KEGG pathways was done with clusterProfiler (Wu et al., 2021).

Data availability

All raw DNA and RNA sequences, and the MAG catalogues are available under HoloFood's umbrella project on ENA (Project ID: PRJEB43192) and displayed in the HoloFood Data Portal (www.holofooddata.org). Bioinformatic resources including ENA accession numbers, scripts, data matrixes and files have been archived in Zenodo with the DOI 10.5281/zenodo.8429925, as a release of the following Github repository: https://github.com/holochicken/priority_effects.

Acknowledgments

This research was funded by the European Union's Horizon Research and Innovation Programme under grant agreement No. 817729 (HoloFood, Holistic solution to improve animal food production through deconstructing the biomolecular interactions between feed, gut microorganisms, and animals in relation to performance parameters), as well as the Danish National Research Foundation grant DNRF143. The work of S.M. was supported by the Basque Government doctoral fellowship. The visit of E.S. was funded by TUBITAK 2214-A International Research Scholarship.

References

Abd El-Hack, M. E., El-Saadony, M. T., Shehata, A. M., Arif, M., Paswan, V. K., Batiha, G. E.-S., Khafaga, A. F., & Elbestawy, A. R. (2021). Approaches to prevent and control Campylobacter spp. colonization in broiler chickens: a review. *Environmental Science and Pollution Research International*, *28*(5), 4989–5004. https://doi.org/10.1007/s11356-020-11747-3

Abebe, E., Gugsa, G., & Ahmed, M. (2020). Review on Major Food-Borne Zoonotic Bacterial Pathogens. *Journal of Tropical Medicine*, *2020*, 4674235. https://doi.org/10.1155/2020/4674235

Awad, W. A., Aschenbach, J. R., Ghareeb, K., Khayal, B., Hess, C., & Hess, M. (2014). Campylobacter jejuni influences the expression of nutrient transporter genes in the intestine of chickens. *Veterinary Microbiology*, *172*(1-2), 195–201. https://doi.org/10.1016/j.vetmic.2014.04.001

Awad, W. A., Hess, C., & Hess, M. (2018). Re-thinking the chicken-Campylobacter jejuni interaction: a review. *Avian Pathology: Journal of the W.V.P.A*, *47*(4), 352–363. https://doi.org/10.1080/03079457.2018.1475724

Awad, W. A., Smorodchenko, A., Hess, C., Aschenbach, J. R., Molnár, A., Dublecz, K., Khayal, B., Pohl, E. E., & Hess, M. (2015). Increased intracellular calcium level and impaired nutrient absorption are important pathogenicity traits in the chicken intestinal epithelium during Campylobacter jejuni colonization. *Applied Microbiology and Biotechnology*, *99*(15), 6431–6441. https://doi.org/10.1007/s00253-015-6543-z

Barton, K. (2009). MuMIn : multi-model inference. *Http://r-Forge.r-Project.org/projects/mumin/*. https://cir.nii.ac.jp/crid/1572824499154168192

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting Linear Mixed-Effects Models using lme4. In *arXiv [stat.CO]*. arXiv. http://arxiv.org/abs/1406.5823

Bäumler, A. J., & Sperandio, V. (2016). Interactions between the microbiota and pathogenic bacteria in the gut. *Nature*, *535*(7610), 85–93. https://doi.org/10.1038/nature18849

Belcour, A., Frioux, C., Aite, M., Bretaudeau, A., Hildebrand, F., & Siegel, A. (2020). Metage2Metabo, microbiota-scale metabolic complementarity for the identification of key species. *eLife*, *9*, e61968. https://doi.org/10.7554/eLife.61968

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful

approach to multiple testing. *Journal of the Royal Statistical Society*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Bozzi, D., Rasmussen, J. A., Carøe, C., Sveier, H., Nordøy, K., Gilbert, M. T. P., & Limborg, M. T. (2021). Salmon gut microbiota correlates with disease infection status: potential for monitoring health in farmed animals. *Animal Microbiome*, *3*(1), 30. https://doi.org/10.1186/s42523-021-00096-2

Burgard, A. P., Nikolaev, E. V., Schilling, C. H., & Maranas, C. D. (2004). Flux coupling analysis of genome-scale metabolic network reconstructions. *Genome Research*, *14*(2), 301–312. https://doi.org/10.1101/gr.1926504

Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P., & Huerta-Cepas, J. (2021). eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Molecular Biology and Evolution*, *38*(12), 5825–5829. https://doi.org/10.1093/molbev/msab293

Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., & Henrissat, B. (2009). The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Research*, *37*(Database issue), D233–D238. https://doi.org/10.1093/nar/gkn663

Carøe, C., Gopalakrishnan, S., Vinner, L., Mak, S. S. T., Sinding, M. H. S., Samaniego, J. A., Wales, N., Sicheritz-Pontén, T., & Gilbert, M. T. P. (2018). Single-tube library preparation for degraded DNA. *Methods in Ecology and Evolution / British Ecological Society*, *9*(2), 410–419. https://onlinelibrary.wiley.com/doi/pdf/10.1111/2041-210X.12871

Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P., & Parks, D. H. (2019). GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* . https://doi.org/10.1093/bioinformatics/btz848

Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* , *34*(17), i884–i890. https://doi.org/10.1093/bioinformatics/bty560

Chen, Y., Ye, W., Zhang, Y., & Xu, Y. (2015). High speed BLASTN: an accelerated MegaBLAST search tool. *Nucleic Acids Research*, *43*(16), 7762–7768. https://doi.org/10.1093/nar/gkv784

Connerton, P. L., Richards, P. J., Lafontaine, G. M., O'Kane, P. M., Ghaffar, N., Cummings, N. J., Smith, D. L., Fish, N. M., & Connerton, I. F. (2018). The effect of the timing of exposure to Campylobacter jejuni on the gut microbiome and inflammatory responses of broiler chickens. *Microbiome*, *6*(1), 88. https://doi.org/10.1186/s40168-018-0477-5

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* , *29*(1), 15–21. https://doi.org/10.1093/bioinformatics/bts635

Ebrahim, A., Lerman, J. A., Palsson, B. O., & Hyduke, D. R. (2013). COBRApy: COnstraints-Based Reconstruction and Analysis for Python. *BMC Systems Biology*, *7*, 74. https://doi.org/10.1186/1752-0509-7-74

Eisenhofer, R., Odriozola, I., & Alberdi, A. (2023). Impact of microbial genome completeness on metagenomic functional inference. *ISME Communications*, *3*(1), 12. https://doi.org/10.1038/s43705-023-00221-z

Fan, Y., Ju, T., Bhardwaj, T., Korver, D. R., & Willing, B. P. (2023). Week-Old Chicks with High Bacteroides Abundance Have Increased Short-Chain Fatty Acids and Reduced Markers of Gut Inflammation. *Microbiology Spectrum*, e0361622. https://doi.org/10.1128/spectrum.03616-22

Food Safety Authority, E. (2021). The European Union one health 2020 zoonoses report. *EFSA*. https://efsa.onlinelibrary.wiley.com/doi/abs/10.2903/j.efsa.2021.6971

Garber, J. M., Nothaft, H., Pluvinage, B., Stahl, M., Bian, X., Porfirio, S., Enriquez, A., Butcher, J., Huang, H., Glushka, J., Line, E., Gerlt, J. A., Azadi, P., Stintzi, A., Boraston, A. B., & Szymanski, C. M. (2020). The gastrointestinal pathogen Campylobacter jejuni metabolizes sugars with potential help from commensal Bacteroides vulgatus. *Communications Biology*, *3*(1), 2. https://doi.org/10.1038/s42003-019-0727-5

Gene Ontology Consortium. (2021). The Gene Ontology resource: enriching a GOld mine. *Nucleic Acids Research*, *49*(D1), D325–D334. https://doi.org/10.1093/nar/gkaa1113

Han, Z., Willer, T., Li, L., Pielsticker, C., Rychlik, I., Velge, P., Kaspers, B., & Rautenschlein, S. (2017). Influence of the Gut Microbiota Composition on Campylobacter jejuni Colonization in Chickens.

Infection and Immunity, *85*(11). https://doi.org/10.1128/IAI.00380-17

Han, Z., Willer, T., Pielsticker, C., Gerzova, L., Rychlik, I., & Rautenschlein, S. (2016). Differences in host breed and diet influence colonization by Campylobacter jejuni and induction of local immune responses in chicken. *Gut Pathogens*, *8*, 56. https://doi.org/10.1186/s13099-016-0133-1

Holmes, I., Harris, K., & Quince, C. (2012). Dirichlet multinomial mixtures: generative models for microbial metagenomics. *PloS One*, *7*(2), e30126. https://doi.org/10.1371/journal.pone.0030126

HoloFood. (n.d.). CORDIS. https://cordis.europa.eu/project/rcn/218793/factsheet/en

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S. K., Cook, H., Mende, D. R., Letunic, I., Rattei, T., Jensen, L. J., von Mering, C., & Bork, P. (2019). eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research*, *47*(D1), D309–D314. https://doi.org/10.1093/nar/gky1085

Humphrey, S., Chaloner, G., Kemmett, K., Davidson, N., Williams, N., Kipar, A., Humphrey, T., & Wigley, P. (2014). Campylobacter jejuni is not merely a commensal in commercial broiler chickens and affects bird welfare. *mBio*, *5*(4), e01364–14. https://doi.org/10.1128/mBio.01364-14

Ijaz, U. Z., Sivaloganathan, L., McKenna, A., Richmond, A., Kelly, C., Linton, M., Stratakos, A. C., Lavery, U., Elmi, A., Wren, B. W., Dorrell, N., Corcionivoschi, N., & Gundogdu, O. (2018). Comprehensive Longitudinal Microbiome Analysis of the Chicken Cecum Reveals a Shift From Competitive to Environmental Drivers and a Window of Opportunity for Campylobacter. *Frontiers in Microbiology*, *9*, 2452. https://doi.org/10.3389/fmicb.2018.02452

Javed, S., Gul, F., Javed, K., & Bokhari, H. (2017). Helicobacter pullorum: An Emerging Zoonotic Pathogen. *Frontiers in Microbiology*, *8*, 604. https://doi.org/10.3389/fmicb.2017.00604

Kaakoush, N. O., Castaño-Rodríguez, N., Mitchell, H. M., & Man, S. M. (2015). Global Epidemiology of Campylobacter Infection. *Clinical Microbiology Reviews*, *28*(3), 687–720. https://doi.org/10.1128/CMR.00006-15

Kaakoush, N. O., Sodhi, N., Chenu, J. W., Cox, J. M., Riordan, S. M., & Mitchell, H. M. (2014). The interplay between Campylobacter and Helicobacter species and other gastrointestinal microbiota of commercial broiler chickens. *Gut Pathogens*, *6*, 18. https://doi.org/10.1186/1757-4749-6-18

Kanehisa, M., & Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, *28*(1), 27–30. https://doi.org/10.1093/nar/28.1.27

Karp, P. D., Midford, P. E., Billington, R., Kothari, A., Krummenacker, M., Latendresse, M., Ong, W. K., Subhraveti, P., Caspi, R., Fulcher, C., Keseler, I. M., & Paley, S. M. (2021). Pathway Tools version 23.0 update: software for pathway/genome informatics and systems biology. *Briefings in Bioinformatics*, *22*(1), 109–126. https://doi.org/10.1093/bib/bbz104

Karp, P. D., Paley, S., & Romero, P. (2002). The Pathway Tools software. *Bioinformatics* , *18 Suppl 1*, S225–S232. https://doi.org/10.1093/bioinformatics/18.suppl_1.s225

Karp, P. D., Riley, M., Paley, S. M., & Pellegrini-Toole, A. (2002). The MetaCyc Database. *Nucleic Acids Research*, *30*(1), 59–61. https://doi.org/10.1093/nar/30.1.59

Kollarcikova, M., Kubasova, T., Karasova, D., Crhanova, M., Cejkova, D., Sisak, F., & Rychlik, I. (2019). Use of 16S rRNA gene sequencing for prediction of new opportunistic pathogens in chicken ileal and cecal microbiota. *Poultry Science*, *98*(6), 2347–2353. https://doi.org/10.3382/ps/pey594

Kredich, N. M. (1992). The molecular basis for positive regulation of cys promoters in Salmonella typhimurium and Escherichia coli. *Molecular Microbiology*, *6*(19), 2747–2753. https://doi.org/10.1111/j.1365-2958.1992.tb01453.x

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*, 1–26. https://doi.org/10.18637/jss.v082.i13

Lahti, L., Shetty, S., Blake, T., & Salojarvi, J. (2017). Tools for microbiome analysis in R. *Version*.

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923

Li, D., Liu, C.-M., Luo, R., Sadakane, K., & Lam, T.-W. (2015). MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph.

Bioinformatics , *31*(10), 1674–1676. https://doi.org/10.1093/bioinformatics/btv033

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* , *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* , *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Liu, Y., Feng, Y., Yang, X., Lv, Z., Li, P., Zhang, M., Wei, F., Jin, X., Hu, Y., Guo, Y., & Liu, D. (2023). Mining chicken ileal microbiota for immunomodulatory microorganisms. *The ISME Journal*, *17*(5), 758–774. https://doi.org/10.1038/s41396-023-01387-z

Luijkx, Y. M. C. A., Bleumink, N. M. C., Jiang, J., Overkleeft, H. S., Wösten, M. M. S. M., Strijbis, K., & Wennekes, T. (2020). Bacteroides fragilis fucosidases facilitate growth and invasion of Campylobacter jejuni in the presence of mucins. *Cellular Microbiology*, *22*(12), e13252. https://doi.org/10.1111/cmi.13252

Mangiola, S., Molania, R., Dong, R., Doyle, M. A., & Papenfuss, A. T. (2021). tidybulk: an R tidy framework for modular transcriptomic data analysis. *Genome Biology*, *22*(1), 42. https://doi.org/10.1186/s13059-020-02233-7

Man, L., Dale, A. L., Klare, W. P., Cain, J. A., Sumer-Bayraktar, Z., Niewold, P., Solis, N., & Cordwell, S. J. (2020). Proteomics of Campylobacter jejuni Growth in Deoxycholate Reveals Cj0025c as a Cystine Transport Protein Required for Wild-type Human Infection Phenotypes. *Molecular & Cellular Proteomics: MCP*, *19*(8), 1263–1280. https://doi.org/10.1074/mcp.RA120.002029

Marcos, S., Odriozola, I., Eisenhofer, R., Aizpurua, O., Tarradas, J., Martin, G., Estonba, A., Gilbert, M. T. P., Kale, V., Baldi, G., & Others. (2023). Reduced metabolic capacity of the gut microbiota associates with host growth in broiler chickens. In *ResearchSquare*. https://doi.org/10.21203/rs.3.rs-2885808/v1

Mendel, R. R., & Leimkühler, S. (2015). The biosynthesis of the molybdenum cofactors. *Journal of Biological Inorganic Chemistry: JBIC: A Publication of the Society of Biological Inorganic Chemistry*, *20*(2), 337–347. https://doi.org/10.1007/s00775-014-1173-y

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L. L., Tosatto, S. C. E., Paladin, L., Raj, S., Richardson, L. J., Finn, R. D., & Bateman, A. (2021). Pfam: The protein families database in 2021. *Nucleic Acids Research*, *49*(D1), D412–D419. https://doi.org/10.1093/nar/gkaa913

Mölder, F., Jablonski, K. P., Letcher, B., Hall, M. B., Tomkins-Tinch, C. H., Sochat, V., Forster, J., Lee, S., Twardziok, S. O., Kanitz, A., Wilm, A., Holtgrewe, M., Rahmann, S., Nahnsen, S., & Köster, J. (2021). Sustainable data analysis with Snakemake. *F1000Research*, *10*, 33. https://doi.org/10.12688/f1000research.29032.2

Munoz, L. R., Bailey, M. A., Krehling, J. T., Bourassa, D. V., Hauck, R., Pacheco, W. J., Chaves-Cordoba, B., Chasteen, K. S., Talorico, A. A., Escobar, C., Pietruska, A., & Macklin, K. S. (2023). Effects of dietary yeast cell wall supplementation on growth performance, intestinal Campylobacter jejuni colonization, innate immune response, villus height, crypt depth, and slaughter characteristics of broiler chickens inoculated with Campylobacter jejuni at d 21. *Poultry Science*, *102*(5), 102609. https://doi.org/10.1016/j.psj.2023.102609

Nurk, S., Meleshko, D., Korobeynikov, A., & Pevzner, P. A. (2017). metaSPAdes: a new versatile metagenomic assembler. *Genome Research*, *27*(5), 824–834. https://doi.org/10.1101/gr.213959.116

Nyholm, L., Koziol, A., Marcos, S., Botnen, A. B., Aizpurua, O., Gopalakrishnan, S., Limborg, M. T., Gilbert, M. T. P., & Alberdi, A. (2020). Holo-Omics: Integrated Host-Microbiota Multi-omics for Basic and Applied Biological Research. *iScience*, *23*(8), 101414. https://doi.org/10.1016/j.isci.2020.101414

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., & Others. (2022). *vegan: Community Ecology Package. R package version 2.5--7. 2020*.

Olm, M. R., Brown, C. T., Brooks, B., & Banfield, J. F. (2017). dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *The ISME Journal*, *11*(12), 2864–2868. https://doi.org/10.1038/ismej.2017.126

Orakov, A., Fullam, A., Coelho, L. P., Khedkar, S., Szklarczyk, D., Mende, D. R., Schmidt, T. S. B., & Bork, P. (2021). GUNC: detection of chimerism and contamination in prokaryotic genomes. *Genome Biology*, *22*(1), 178. https://doi.org/10.1186/s13059-021-02393-0

Papke, L. E., & Wooldridge, J. M. (1996). Econometric methods for fractional response variables with an application to 401(k) plan participation rates. *Journal of Applied Economics*, *11*(6), 619–632. https://doi.org/10.1002/(sici)1099-1255(199611)11:6<619::aid-jae418>3.0.co;2-1

Paradis, E., Claude, J., & Strimmer, K. (2004). APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* , *20*(2), 289–290. https://doi.org/10.1093/bioinformatics/btg412

Parks, D. H., Chuvochina, M., Rinke, C., Mussig, A. J., Chaumeil, P.-A., & Hugenholtz, P. (2022). GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Research*, *50*(D1), D785–D794. https://doi.org/10.1093/nar/gkab776

Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., & Tyson, G. W. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research*, *25*(7), 1043–1055. https://doi.org/10.1101/gr.186072.114

Petushkova, E., Mayorova, E., & Tsygankov, A. (2021). TCA Cycle Replenishing Pathways in Photosynthetic Purple Non-Sulfur Bacteria Growing with Acetate. *Life* , *11*(7). https://doi.org/10.3390/life11070711

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., & Glöckner, F. O. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research*, *41*(Database issue), D590–D596. https://doi.org/10.1093/nar/gks1219

Rawlings, N. D., Barrett, A. J., & Bateman, A. (2010). MEROPS: the peptidase database. *Nucleic Acids Research*, *38*(Database issue), D227–D233. https://doi.org/10.1093/nar/gkp971

Richardson, L., Allen, B., Baldi, G., Beracochea, M., Bileschi, M. L., Burdett, T., Burgin, J., Caballero-Pérez, J., Cochrane, G., Colwell, L. J., Curtis, T., Escobar-Zepeda, A., Gurbich, T. A., Kale, V., Korobeynikov, A., Raj, S., Rogers, A. B., Sakharova, E., Sanchez, S., … Finn, R. D. (2023). MGnify: the microbiome sequence data analysis resource in 2023. *Nucleic Acids Research*, *51*(D1), D753–D759. https://doi.org/10.1093/nar/gkac1080

Risso, C., Van Dien, S. J., Orloff, A., Lovley, D. R., & Coppi, M. V. (2008). Elucidation of an alternate isoleucine biosynthesis pathway in Geobacter sulfurreducens. *Journal of Bacteriology*, *190*(7), 2266–2274. https://doi.org/10.1128/JB.01841-07

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* , *26*(1), 139–140. https://doi.org/10.1093/bioinformatics/btp616

Rychlik, I. (2020). Composition and Function of Chicken Gut Microbiota. *Animals*. https://www.mdpi.com/2076-2615/10/1/103

Rzeznitzeck, J., Breves, G., Rychlik, I., Hoerr, F. J., von Altrock, A., Rath, A., & Rautenschlein, S. (2022). The effect of Campylobacter jejuni and Campylobacter coli colonization on the gut morphology, functional integrity, and microbiota composition of female turkeys. *Gut Pathogens*, *14*(1), 33. https://doi.org/10.1186/s13099-022-00508-x

Sahoo, M., Panigrahi, C., & Aradwad, P. (2022). Management strategies emphasizing advanced food processing approaches to mitigate food borne zoonotic pathogens in food system. *Food Frontiers*, *3*(4), 641–665. https://doi.org/10.1002/fft2.153

Salmon-Divon, M., He, Y. O., Kornspan, D., & Wen, Z. T. (2022). Editorial: Omics approach to study the biology and virulence of microorganisms causing zoonotic diseases. *Frontiers in Microbiology*, *13*. https://doi.org/10.3389/fmicb.2022.988983

Schubert, M., Lindgreen, S., & Orlando, L. (2016). AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Research Notes*, *9*, 88. https://doi.org/10.1186/s13104-016-1900-2

Shaffer, M., Borton, M. A., McGivern, B. B., Zayed, A. A., La Rosa, S. L., Solden, L. M., Liu, P., Narrowe, A. B., Rodríguez-Ramos, J., Bolduc, B., Gazitúa, M. C., Daly, R. A., Smith, G. J., Vik, D. R., Pope, P. B., Sullivan, M. B., Roux, S., & Wrighton, K. C. (2020). DRAM for distilling microbial metabolism to automate the curation of microbiome function. *Nucleic Acids Research*, *48*(16),

8883–8900. https://doi.org/10.1093/nar/gkaa621

Shen, W., Le, S., Li, Y., & Hu, F. (2016). SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. *PloS One*, *11*(10), e0163962. https://doi.org/10.1371/journal.pone.0163962

Smith, K. M., Machalaba, C. C., Seifman, R., Feferholtz, Y., & Karesh, W. B. (2019). Infectious disease and economics: The case for considering multi-sectoral impacts. *One Health (Amsterdam, Netherlands)*, *7*, 100080. https://doi.org/10.1016/j.onehlt.2018.100080

Szymanski, M., Zielezinski, A., Barciszewski, J., Erdmann, V. A., & Karlowski, W. M. (2016). 5SRNAdb: an information resource for 5S ribosomal RNAs. *Nucleic Acids Research*, *44*(D1), D180–D183. https://doi.org/10.1093/nar/gkv1081

Thibodeau, A., Fravalo, P., Yergeau, É., Arsenault, J., Lahaye, L., & Letellier, A. (2015). Chicken Caecal Microbiome Modifications Induced by Campylobacter jejuni Colonization and by a Non-Antibiotic Feed Additive. *PloS One*, *10*(7), e0131978. https://doi.org/10.1371/journal.pone.0131978

Tous, N., Marcos, S., Boroojeni, F., Pérez de Rozas, A., Zentek, J., Estonba, A., Sandvang, D., Gilbert, M. T. P., Esteve-Garcia, E., Finn, R., Alberdi, A., & Tarradas, J. (2022). Novel Strategies to Improve Chicken Performance and Welfare by Unveiling Host-Microbiota Interactions through Hologenomics. *Frontiers in Physiology*. https://doi.org/10.3389/fphys.2022.884925

UniProt Consortium. (2019). UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Research*, *47*(D1), D506–D515. https://doi.org/10.1093/nar/gky1049

Vacca, M., Celano, G., Calabrese, F. M., Portincasa, P., Gobbetti, M., & De Angelis, M. (2020). The Controversial Role of Human Gut Lachnospiraceae. *Microorganisms*, *8*(4). https://doi.org/10.3390/microorganisms8040573

Venables, W. N., & Ripley, B. D. (2002). *Modern Applied Statistics with S, Springer, New York: ISBN 0-387-95457-0*.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., … Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., Fu, X., Liu, S., Bo, X., & Yu, G. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation (Cambridge (Mass.))*, *2*(3), 100141. https://doi.org/10.1016/j.xinn.2021.100141

Wyszyńska, A. K., & Godlewska, R. (2021). Lactic Acid Bacteria - A Promising Tool for Controlling Chicken Campylobacter Infection. *Frontiers in Microbiology*, *12*, 703441. https://doi.org/10.3389/fmicb.2021.703441

Yan, W., Zhou, Q., Yuan, Z., Fu, L., Wen, C., Yang, N., & Sun, C. (2021). Impact of the gut microecology on Campylobacter presence revealed by comparisons of the gut microbiota from chickens raised on litter or in individual cages. *BMC Microbiology*, *21*(1), 290. https://doi.org/10.1186/s12866-021-02353-5

Yu, G., Smith, D. K., Zhu, H., Guan, Y., & Lam, T. T.-Y. (2017). Ggtree: An r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution / British Ecological Society*, *8*(1), 28–36. https://doi.org/10.1111/2041-210x.12628

Zamarreño Beas, J., Videira, M. A. M., & Saraiva, L. M. (2022). Regulation of bacterial haem biosynthesis. *Coordination Chemistry Reviews*, *452*, 214286. https://doi.org/10.1016/j.ccr.2021.214286

Supplementary material

Index

Supplementary Table S1:
Overrepresented source metabolites for *Campylobacter jejuni* in the distinct enterotype.

Gene	p-value	Standard mean value	Distinct mean value	Difference	Adjusted_ p-value
COPROPORPHYRIN_III	2.6728E-16	0.02	0.31	0.29	1.5101E-14
CPD.31	1.6891E-07	0.02	0.30	0.28	1.0046E-06
SULFATE	1.5273E-09	0.07	0.35	0.28	1.4382E-08
MOCS3.Cysteine	1.1046E-06	0.04	0.31	0.27	5.221E-06
Thiocarboxylated.MPT.synthases	1.1046E-06	0.04	0.31	0.27	5.221E-06
D.ERYTHRO.IMIDAZOLE.GLYCEROL.P	2.1499E-11	0.06	0.33	0.27	7.815E-10
FMNH2	1.4416E-08	0.08	0.34	0.26	1.0181E-07
Dihydro.Lipoyl.Proteins	1.4661E-06	0.05	0.31	0.26	6.3718E-06
TRP.tRNAs	2.6447E-10	0.13	0.38	0.25	3.3206E-09
ASP.tRNAs	2.6447E-10	0.13	0.38	0.25	3.3206E-09
TYR.tRNAs	2.6447E-10	0.13	0.38	0.25	3.3206E-09
Protein.Histidines	2.6066E-20	0.15	0.38	0.24	2.9454E-18
TYR	2.7664E-11	0.16	0.39	0.24	7.815E-10
ALA.tRNAs	2.4764E-10	0.16	0.39	0.22	3.3206E-09
XANTHINE	0.00016904	0.44	0.65	0.21	0.0005306
CPD.597	8.9924E-10	0.20	0.40	0.20	1.0161E-08
N.acetyl.D.mannosamine	1.1089E-06	0.27	0.46	0.20	5.221E-06
HS	0.00036103	0.45	0.64	0.19	0.00109149
L.PANTOATE	8.5958E-09	0.36	0.55	0.19	6.4755E-08
Red.Glutaredoxins	0.00042238	0.48	0.67	0.19	0.00119324
ETOH	8.3666E-09	0.23	0.42	0.19	6.4755E-08
MANNOSE.1P	1.1755E-07	0.33	0.50	0.17	7.8139E-07
X2.PG	0.00390177	0.50	0.67	0.17	0.00938084
CHOLATE	0.00012236	0.41	0.57	0.16	0.00039505
CPD.15189	0.00012236	0.41	0.57	0.16	0.00039505
GLUTATHIONE	0.00650179	0.53	0.68	0.15	0.01469404
NIACINE	0.00036705	0.48	0.63	0.14	0.00109149
LACTALD	2.4602E-05	0.49	0.62	0.13	9.2668E-05
X3.Hydroxy.Terminated.DNAs	9.9857E-06	0.66	0.74	0.07	4.0299E-05
DCTP	0.02557008	0.73	0.79	0.06	0.04897321
CPD.4211	0.0218647	0.04	0.04	0.00	0.04350291

Supplementary Table S2:
Differentially expressed genes in caecal host mucosa at days 7 and 35 between standard and distinct enterotypes.

Day	Gene	logFC	FDR	description
7	LOC107055361	-1.813867045	0.01946043684	
7	HBA1	1.394133619	0.03030980116	hemoglobin subunit alpha 1
7	HBE1	1.882651162	0.03797486673	hemoglobin subunit epsilon 1
7	H1F0	1.368723298	0.04274410218	H1 histone family member 0
35	NPTX2	1.260895776	0.002013652073	
35	WDR72	1.406510168	0.00248805704	
35	SYPL1	1.47579761	0.005875019077	synaptophysin like 1
35	SLC23A1	-1.591967739	0.008630811	
35	LOC124417815	-1.448936666	0.008913914643	
35	ABCB5	1.529901554	0.01015709323	
35	TNFRSF13B	-1.405821071	0.01015709323	TNF receptor superfamily member 13B
35	POU2AF1	-1.237562167	0.01351103227	POU class 2 homeobox associating factor 1
35	LOC107055347	-1.297939747	0.01351103227	
35	LOC112532123	1.484612487	0.01520477153	
35	IGLL1	-1.392582362	0.01682239206	immunoglobulin lambda-like polypeptide 1
35	CYP1A1	3.118820256	0.01823410302	cytochrome P450 family 1 subfamily A polypeptide 1
35	DCLK2	1.222715467	0.02210893979	
35	PLA2G12B	1.451253504	0.02270499991	
35	LOC121109172	2.872526104	0.02322920014	
35	AOX2	1.771952135	0.02322920014	aldehyde oxidase 2
35	LSAMP	-2.516523807	0.02336753155	limbic system-associated membrane protein
35	SCN3B	1.258063206	0.02342815201	
35	UNC13C	1.339991104	0.03344018545	
35	POU3F4	-1.432191081	0.03449976386	
35	CYP1A2	2.179678621	0.03968125009	cytochrome P450 family 1 subfamily A polypeptide 2
35	TENM2	-1.822964898	0.03968125009	teneurin transmembrane protein 2
35	LOC101751887	-1.268576734	0.03968125009	
35	SERPINB2	1.890363862	0.04185810366	
35	XKR9	1.236415045	0.04238328498	XK related 9
35	CXCL13L2	-1.279297801	0.04277567648	C-X-C motif chemokine ligand 13-like 2
35	GATD3AL2	-1.576733854	0.04324779248	
35	TAAR1	-1.287719979	0.04348330823	
35	SLC2A9	-1.35637258	0.04390555035	
35	MMP27	-1.753122383	0.04497641138	matrix metallopeptidase 1
35	LAG3	-1.29466731	0.04633592068	
35	KCNA3	-1.228759901	0.04791602393	potassium voltage-gated channel subfamily A member 3
35	NPW	-1.250985281	0.04949297101	
35	GNLY	-1.240782685	0.04950569715	granulysin
35	EVA1CL	-1.752173587	0.04987489453	
35	CPT1A	1.20428051	0.04989214558	carnitine palmitoyltransferase 1A

Supplementary Table S3:
Nomenclature, classification and definitions of metabolic pathways employed in the functional distillation analysis using the R package distillR.

Pathway	Element	Function	Definition
B010101	Inosinic acid (IMP)	Nucleic acid biosynthesis	K00764 (K01945,K11787,K11788,K13713) (K00601,K11175,K08289,K11787,K01492) (K01952,(K23269+K23264+K23265),(K23270+K23265)) (K01933,K11787,(K11788 (K01587,K11808,(K01589 K01588)))) (K01923,K01587,K13713) K01756 (K00602,(K01492,(K06863 K11176)))
B010201	Uridylic acid (UMP)	Nucleic acid biosynthesis	(K11540,((K11541 K01465),((K01954,(K01955+K01956)) ((K00609+K00610),K00608) K01465))) (K00226,K00254,K17828) (K13421,(K00762 K01591))
B010301	UDP/UTP	Nucleic acid biosynthesis	(K13800,K13809,K09903)
B010401	CDP/CTP	Nucleic acid biosynthesis	(K00940,K18533) K01937
B010501	ADP/ATP	Nucleic acid biosynthesis	K01939 K01756 (K00939,K18532,K18533,K00944) K00940
B010601	GDP/GTP	Nucleic acid biosynthesis	K00088 K01951 K00942 (K00940,K18533)
B020401	Serine	Amino acid biosynthesis	K00058 K00831 (K01079,K02203,K22305,K25528)
B020501	Threonine	Amino acid biosynthesis	(K00928,K12524,K12525,K12526) K00133 (K00003,K12524,K12525) (K00872,K02204,K02203) K01733
B020601	Cysteine	Amino acid biosynthesis	(K00640,K23304) (K01738,K13034,K17069)
B020602	Cysteine	Amino acid biosynthesis	(K01697,K10150) K01758
B020603	Cysteine	Amino acid biosynthesis	K00789 K17462 K01243 K07173 K17216 K17217
B020701	Methionine	Amino acid biosynthesis	(K00928,K12524,K12525) K00133 (K00003,K12524,K12525) (K00651,K00641) K01739 (K01760,K14155) (K00548,K24042,K00549)
B020801	Valine	Amino acid biosynthesis	(K01652+(K01653,K11258)) K00053 K01687 K00826
B020901	Isoleucine	Amino acid biosynthesis	(K01703+K01704) K00052
B020902	Isoleucine	Amino acid biosynthesis	(K17989,K01754) (K01652+(K01653,K11258)) K00053 K01687 K00826
B021001	Leucine	Amino acid biosynthesis	K01649 (K01702,(K01703+K01704)) K00052
B021101	Lysine	Amino acid biosynthesis	(K00928,K12524,K12525,K12526) K00133 K01714 K00215 K00674 (K00821,K14267) K01439 K01778 (K01586,K12526)
B021102	Lysine	Amino acid biosynthesis	K00928 K00133 K01714 K00215 K05822 K00841 K05823 K01778 K01586
B021103	Lysine	Amino acid biosynthesis	(K00928,K12524,K12525,K12526) K00133 K01714 K00215 K03340 (K01586,K12526)
B021104	Lysine	Amino acid biosynthesis	(K00928,K12524,K12525,K12526) K00133 K01714 K00215 K10206 K01778 (K01586,K12526)
B021105	Lysine	Amino acid biosynthesis	K01655 ((K17450 K01705),(K16792+K16793)) K05824

B021106	Lysine	Amino acid biosynthesis	K05827 K05828 K05829 K05830 K05831
B021201	Arginine	Amino acid biosynthesis	K00611 K01940 (K01755,K14681)
B021202	Arginine	Amino acid biosynthesis	K22478 K00145 K00821 K09065 K01438 K01940 K01755
B021301	Proline	Amino acid biosynthesis	((K00931 K00147),K12657) K00286
B021401	Glutamate	Amino acid biosynthesis	K00673 K01484 K00840 K06447 K05526
B021402	Glutamate	Amino acid biosynthesis	K01745 K01712 K01468 (K01479,K00603,K13990,(K05603 K01458))
B021501	Histidine	Amino acid biosynthesis	K00765 ((K01523 K01496),K11755,K14152) (K01814,K24017) ((K02501+K02500),K01663) ((K01693 K00817 (K04486,K05602,K18649)),(K01089 K00817)) (K00013,K14152)
B021601	Tryptophan	Amino acid biosynthesis	((((K01657+K01658),K13503,K13501,K01656) K00766),K13497) (((K01817,K24017) (K01656,K01609)),K13498,K13501) ((K01695+(K01696,K06001)),K01694)
B021701	Phenylalanine	Amino acid biosynthesis	(((K01850,K04092,K14187,K04093,K04516,K06208,K06209) (K01713,K04518,K05359)),K14170) (K00832,K00838)
B021801	Tyrosine	Amino acid biosynthesis	(((K01850,K04092,K14170,K04093,K04516,K06208,K06209) (K04517,K00211)),K14187) (K00832,K00838)
B021802	Tyrosine	Amino acid biosynthesis	(K01850,K04092,K14170) (K00832,K15849) (K00220,K24018,K15227)
B021901	GABA	Amino acid biosynthesis	K09470 K09471 K09472 K09473
B022001	Beta-alanine	Amino acid biosynthesis	(K00207,(K17722+K17723)) K01464 (K01431,K06016)
B022002	Beta-alanine	Amino acid biosynthesis	6.2.1.17 1.3.8.1 4.2.1.116 3.1.2.4 1.1.159 2.6.1.18
B022101	Ornithine	Amino acid biosynthesis	(K00618,K00619,K14681,K14682,K00620,K22477,K22478) (((K00930,K22478) K00145),K12659) (K00818,K00821) (K01438,K14677,K00620)
B022102	Ornithine	Amino acid biosynthesis	K19412 K05828 K05829 K05830 K05831
B022103	Ornithine	Amino acid biosynthesis	2.3.1.1 2.7.2.8 1.2.1.38 2.6.1.11 3.5.1.16
B030201	Betaine	Amino acid derivative biosynthesis	1.1.99.1 1.2.1.8
B030202	Betaine	Amino acid derivative biosynthesis	2.1.1.156 2.1.1.157
B030301	Ectoine	Amino acid derivative biosynthesis	K00928 K00133 K00836 K06718 K06720
B030701	Spermidine	Amino acid derivative biosynthesis	(K01583,K01584,K01585,K02626) K01480
B030901	Putrescine	Amino acid derivative	K01476 K01581

		biosynthesis	
B031001	Tryptamine	Amino acid derivative biosynthesis	(4.1.1.28,4.1.1.105)
B040101	Acetate	SCFA biosynthesis	(K00625,K13788,K15024) K00925
B040103	Acetate	SCFA biosynthesis	K01067
B040104	Acetate	SCFA biosynthesis	5.4.3.2 5.4.3.3 1.4.1.11 2.3.1.247 1.3.1.109 2.8.3.9 (2.3.1.9,2.3.1.16) 2.3.1.8 (2.7.2.1,2.7.2.15)
B040105	Acetate	SCFA biosynthesis	1.21.4.2 (2.7.2.1,2.7.2.15)
B040106	Acetate	SCFA biosynthesis	(1.2.7.1,1.2.1.104) 2.3.1.8 (2.7.2.1,2.7.2.15)
B040201	Butyrate	SCFA biosynthesis	1.2.7.1 (2.3.1.9,2.3.1.16) 1.1.1.35 4.2.1.150 1.3.1.109 2.3.1.19 (2.7.2.7,2.7.2.14)
B040202	Butyrate	SCFA biosynthesis	2.3.1.8 (2.7.2.1,2.7.2.15) (2.8.3.1,2.8.3.8)
B040203	Butyrate	SCFA biosynthesis	(2.3.1.9,2.3.1.16) 1.1.1.36 4.2.1.55 1.3.1.109 (2.8.3.1,2.8.3.8)
B040204	Butyrate	SCFA biosynthesis	1.4.1.2 1.1.1.399 2.8.3.12 4.2.1.167 7.2.4.5 1.3.1.109 (2.8.3.1,2.8.3.8)
B040205	Butyrate	SCFA biosynthesis	5.4.3.2 5.4.3.3 1.4.1.11 2.3.1.247 1.3.1.109 2.8.3.9
B040206	Butyrate	SCFA biosynthesis	2.8.3.18 1.2.1.76 1.1.1.61 2.8.3.M6 4.2.1.120 1.3.1.109 (2.8.3.1,2.8.3.8)
B040207	Butyrate	SCFA biosynthesis	1.4.1.2 1.1.1.399 2.8.3.12 4.2.1.167 7.2.4.5 1.3.1.109 (2.8.3.1,2.8.3.8)
B040208	Butyrate	SCFA biosynthesis	1.4.1.2 1.1.1.399 2.8.3.12 4.2.1.167 7.2.4.5 1.3.1.109 (2.8.3.1,2.8.3.8)
B040301	Propionate	SCFA biosynthesis	(4.2.1.28,1.1.1.1) 1.2.1.87 2.3.1.222 (2.7.2.1,2.7.2.7,2.7.2.14,2.7.2.15)
B040302	Propionate	SCFA biosynthesis	4.3.1.19 2.3.1.222 (2.7.2.1,2.7.2.7,2.7.2.14,2.7.2.15)
B040304	Propionate	SCFA biosynthesis	2.1.3.1 1.1.1.37 4.2.1.2 1.3.5.1 2.8.3.27
B040305	Propionate	SCFA biosynthesis	2.8.3.1 4.2.1.54 1.3.1.95 2.8.3.1
B040306	Propionate	SCFA biosynthesis	2.6.1.2 1.1.1.28 2.8.3.1 4.2.1.54 1.3.1.95 2.8.3.1
B050101	Indole-3-acetate	Indolic compound biosynthesis	1.13.12.3 3.5.1.4
B050102	Indole-3-acetate	Indolic compound biosynthesis	4.2.1.84 3.5.1.4
B050103	Indole-3-acetate	Indolic compound biosynthesis	3.5.5.1
B050104	Indole-3-acetate	Indolic compound biosynthesis	(2.6.1.1,2.6.1.27) 4.1.1.74 1.2.3.7
B050105	Indole-3-acetate	Indolic compound biosynthesis	(4.1.1.28,4.1.1.105) 1.4.3.4 1.2.3.7
B060401	indole-3-lactate	Organic anion biosynthesis	1.1.1.27
B060402	L-lactate	Organic anion biosynthesis	1.1.1.22
B060501	D-lactate	Organic anion biosynthesis	1.1.1.28
B060101	Succinate	Organic anion biosynthesis	(K01647,K05942) (K01681,K01682) (K00031,K00030) ((((K00164+K00658),K01616)+K00382),(K00174+K00175)) ((K01902+K01903),(K01899+K01900),K18118) ((K00234+K00235+K00236+(K00237,K25801)),(K00239+K00240+K00241),(K00244+K00245+K00246)) (K01676,K01679,(K01677+K01678)) (K00026,K00025,K00024,K00116)
B060102	Succinate	Organic anion biosynthesis	((((K00164+K00658),K01616)+K00382),K00174) ((K01902+K01903),(K01899+K01900),K18118) ((K00234+K00235+K00236+(K00237,K25801)),(K00239+K00240+K00241),(

			K00244+K00245+K00246)) (K01676,K01679,(K01677+K01678)) (K00026,K00025,K00024,K00116)
B060103	Succinate	Organic anion biosynthesis	K01580 (K13524,K07250,K00823,K16871) (K00135,K00139,K17761)
B060104	Succinate	Organic anion biosynthesis	(K00169+K00170+K00171+K00172) K01007 K01595 K00024 (K01677+K01678) (K00239+K00240) (K01902+K01903) (K15038,K15017) K14465 (K14467,K18861) K14534 K15016 K00626
B060105	Succinate	Organic anion biosynthesis	(K02160+K01961+K01962+K01963) K14468 K14469 K15052 K05606 (K01847,(K01848+K01849)) (K14471+K14472) (K00239+K00240+K00241) K01679 K08691 K14449 K14470 K09709
B060106	Succinate	Organic anion biosynthesis	(K00169+K00170+K00171+K00172) (K01959+K01960) K00024 (K01677+K01678) (K18209+K18210) (K01902+K01903) (K00174+K00175+K00176+K00177)
B060201	Fumarate	Organic anion biosynthesis	(K01647,K05942) (K01681,K01682) (K00031,K00030) ((((K00164+K00658),K01616)+K00382),(K00174+K00175)) ((K01902+K01903),(K01899+K01900),K18118) ((K00234+K00235+K00236+(K00237,K25801)),(K00239+K00240+K00241),(K00244+K00245+K00246)) (K01676,K01679,(K01677+K01678)) (K00026,K00025,K00024,K00116)
B060202	Fumarate	Organic anion biosynthesis	((((K00164+K00658),K01616)+K00382),K00174) ((K01902+K01903),(K01899+K01900),K18118) ((K00234+K00235+K00236+(K00237,K25801)),(K00239+K00240+K00241),(K00244+K00245+K00246)) (K01676,K01679,(K01677+K01678)) (K00026,K00025,K00024,K00116)
B060203	Fumarate	Organic anion biosynthesis	K01948 K00611 K01940 (K01755,K14681) K01476
B060204	Fumarate	Organic anion biosynthesis	(K00815,K00838,K00832,K03334) K00457 K00451 K01800 (K01555,K16171)
B060205	Fumarate	Organic anion biosynthesis	K00241+(K00242,K18859,K18860)+K00239+K00240
B060206	Fumarate	Organic anion biosynthesis	K00244+K00245+K00246+K00247
B060207	Fumarate	Organic anion biosynthesis	(K00169+K00170+K00171+K00172) K01007 K01595 K00024 (K01677+K01678) (K00239+K00240) (K01902+K01903) (K15038,K15017) K14465 (K14467,K18861) K14534 K15016 K00626
B060208	Fumarate	Organic anion biosynthesis	(K02160+K01961+K01962+K01963) K14468 K14469 K15052 K05606 (K01847,(K01848+K01849)) (K14471+K14472) (K00239+K00240+K00241) K01679 K08691 K14449 K14470 K09709
B060209	Fumarate	Organic anion biosynthesis	(K00169+K00170+K00171+K00172) (K01959+K01960) K00024 (K01677+K01678) (K18209+K18210) (K01902+K01903) (K00174+K00175+K00176+K00177)
B060210	Fumarate	Organic anion biosynthesis	(K18029+K18030) K14974 K18028 K15357 K13995 K01799
B060211	Fumarate	Organic anion biosynthesis	K00611 K01940 (K01755,K14681)
B060212	Fumarate	Organic anion biosynthesis	K22478 K00145 K00821 K09065 K01438 K01940 K01755
B060213	Fumarate	Organic anion biosynthesis	2.6.1.1,2.6.1.5,2.6.1.27,2.6.1.57 1.13.11.27 1.13.11.5 5.2.1.2 3.7.1.2
B060301	Citrate	Organic anion biosynthesis	(K01647,K05942) (K01681,K01682) (K00031,K00030) ((((K00164+K00658),K01616)+K00382),(K00174+K00175)) ((K01902+K01903),(K01899+K01900),K18118) ((K00234+K00235+K00236+(K00237,K25801)),(K00239+K00240+K00241),(K00244+K00245+K00246)) (K01676,K01679,(K01677+K01678)) (K00026,K00025,K00024,K00116)
B060302	Citrate	Organic anion biosynthesis	(K01647,K05942) (K01681,K01682) (K00031,K00030)
B060303	Citrate	Organic anion biosynthesis	K01647 (K01681,K01682) K01637 (K01638,K19282) (K00026,K00025,K00024)
B060304	Citrate	Organic anion	K01647 K01681 K00031 K00261 (K19268+K01846) K04835 K19280 K14449 K19281 K19282 K00024

		biosynthesis	
B070101	Thiamine (B1)	Vitamin biosynthesis	(((K03148+K03154) K03151),(K03150 K03149)) K03147 ((K00941 K00788),K14153,K21219) K00946
B070102	Thiamine (B1)	Vitamin biosynthesis	(((K03148+K03154) K03151),(K03153 K03149 K10810)) K03147 K00941 K00788 K00946
B070103	Thiamine (B1)	Vitamin biosynthesis	(K22699,K03147) ((K00941 (K00788,K21220)),K21219) K00946
B070104	Thiamine (B1)	Vitamin biosynthesis	(K00941 K00788),K14153,K21219
B070201	Riboflavin (B2)	Vitamin biosynthesis	(((K01497,K14652) ((K01498 K00082),K11752) (K22912,K20860,K20861,K20862,K21063,K21064)),(K02858,K14652)) K00794 K00793 ((K20884 K22949),K11753)
B070301	Niacin (B3)	Vitamin biosynthesis	3.6.1.22 (3.2.2.6,3.2.2.4) 3.4.1.19
B070401	Pantothenate (B5)	Vitamin biosynthesis	((K00826 K00606 K00077),K01579) (K01918,K13799)
B070402	Pantothenate (B5)	Vitamin biosynthesis	((K00606 K00077),(K13367 K00128)) K01918
B070501	Pyridoxal-P (B6)	Vitamin biosynthesis	K03472 K03473 K00831 K00097 K03474 K00275
B070502	Pyridoxal-P (B6)	Vitamin biosynthesis	K06215 K08681
B070601	Biotin (B7)	Vitamin biosynthesis	K00652 (((K00833,K19563) K01935),K19562) K01012
B070602	Biotin (B7)	Vitamin biosynthesis	K00652 K25570 K01935 K01012
B070603	Biotin (B7)	Vitamin biosynthesis	K16593 K00652 K19563 K01935 K01012
B070604	Biotin (B7)	Vitamin biosynthesis	K01906 K00652 (K00833,K19563) K01935 K01012
B070701	Tetrahydrofolate (B9)	Vitamin biosynthesis	(K01495,K09007,K22391) (K01077,K01113,(K08310,K19965)) ((K13939,((K13940,(K01633 K00950)) K00796)),(K01633 K13941)) (K11754,K20457) (K00287,K13998)
B070702	Tetrahydrofolate (B9)	Vitamin biosynthesis	K14652 K22100 K01633 K13941 K22099 K00287
B070801	Cobalamin (B12)	Vitamin biosynthesis	(K02302,((K02303,K13542) (K02304,K24866))) (K02190,K03795,K22011) K03394 (K05934,K13541,K21479) K05936 (K02189,K13541) K02188 K05895 ((K02191 K03399),K00595) K06042 K02224
B070802	Cobalamin (B12)	Vitamin biosynthesis	(K02303,K13542) (K03394,K13540) K02229 (K05934,K13540,K13541) K05936 K02228 K05895 K00595 K06042 K02224 K02230+K09882+K09883
B070803	Cobalamin (B12)	Vitamin biosynthesis	(K00798,K19221) K02232 (K02225,K02227) K02231 K00768 (K02226,K22316) K02233
B070901	Tocopherol/tocotorienol (E)	Vitamin biosynthesis	K09833 (K12502,K18534) K09834 K05928
B071001	Phylloquinone (K1)	Vitamin biosynthesis	((K02552 K02551 K08680 K02549),K14759) (K01911,K14760) K01661 (K19222,K12073) K23094 K17872 K23095
B071101	Menaquinone (K2)	Vitamin biosynthesis	K02552 K02551 K08680 K02549 K01911 K01661 K19222 K02548 K03183
B071102	Menaquinone (K2)	Vitamin biosynthesis	K11782 K18285 (K18286,K20810) K11783 K11784 K11785

B071103	Menaquinone (K2)	Vitamin biosynthesis	K11782 K18285 K18284 K11784 K11785
B071201	Ubiquinone (Q10)	Vitamin biosynthesis	(K03181,K18240) K03179 (K03182+K03186) K18800 K00568 K03185 K03183 (K03184,K06134) K00568
B071202	Ubiquinone (Q10)	Vitamin biosynthesis	K06125 K06126 K00591 K06127 K06134 K00591
B080101	Salicylate	Aromatic compound biosynthesis	5.4.4.2 4.2.99.21
B080201	Gallate	Aromatic compound biosynthesis	4.2.1.10
B080301	Chorismate	Aromatic compound biosynthesis	4.2.1.10 1.1.1.25 2.7.1.71 2.5.1.19 4.2.3.5
B080302	Chorismate	Aromatic compound biosynthesis	2.5.1.54 3.2.3.4 4.2.1.10 1.1.1.25 2.7.1.71 2.5.1.19 4.2.3.5
B080303	Chorismate	Aromatic compound biosynthesis	2.7.2.4 1.2.1.11 2.2.1.10 1.4.1.24 4.2.1.10 1.1.1.25 2.7.1.71 2.5.1.19 4.2.3.5
B080404	Dipicolinate	Aromatic compound biosynthesis	2.7.2.4 1.2.1.11 4.3.3.7
B090101	Staphyloferrin	Metallophore biosynthesis	K21898 K23446 K23447
B090102	Staphyloferrin	Metallophore biosynthesis	(5.1.1.10,5.1.1.12) 6.3.2.58 6.3.2.57
B090103	Staphyloferrin	Metallophore biosynthesis	K23371 K21949 K21721 K23372 K23373 K23374 K23375
B090104	Staphyloferrin	Metallophore biosynthesis	2.7.1.225 2.5.1.140 1.5.1.51 6.3.2.54 4.1.1.117 6.3.2.55 6.3.2.56
B090201	Aerobactin	Metallophore biosynthesis	K03897 K03896 K03894 K03895
B090202	Aerobactin	Metallophore biosynthesis	1.14.13.59 2.3.1.102 6.3.2.38 6.3.2.39
B090301	Staphylopine	Metallophore biosynthesis	(5.1.1.10,5.1.1.24) 2.5.1.152 1.5.1.52
B100402	Bacilysin	Antibiotic biosynthesis	5.4.99.5 4.1.1.100 5.3.3.19 ((5.3.3.19 1.3.1.aa),1.3.1.aa) 1.1.1.385 6.3.2.49
B100601	Carbapenem-3-carboxylate	Antibiotic biosynthesis	K18317 K18316 K18315
B100801	Clavaminate	Antibiotic biosynthesis	K12673 K12674 K12675 K12676
B100802	Clavaminate	Antibiotic biosynthesis	2.5.1.66 6.3.3.4 1.14.11.21 3.5.3.22 1.14.11.21
B101101	Erythromycin	Antibiotic biosynthesis	2.1.1.254 1.14.13.154

B101102	Erythromycin	Antibiotic biosynthesis	2.3.1.94 1.14.15.35 2.4.1.328 2.4.1.278
B101202	Fosfomycin	Antibiotic biosynthesis	5.4.2.9 4.1.1.82 1.1.1.309 2.7.7.104 2.1.1.308 1.11.1.23
B101401	Kanosamine	Antibiotic biosynthesis	K18652 K18653 K18654
B101402	Kanosamine	Antibiotic biosynthesis	1.1.1.361 2.6.1.104 3.1.3.92
B102101	Novobiocin	Antibiotic biosynthesis	6.3.1.15 2.1.1.284 2.4.1.302 2.1.1.285 2.1.3.12
B102201	Paromamine	Antibiotic biosynthesis	4.2.3.124 2.6.1.100 (1.1.1.329,1.1.99.38) 2.6.1.101 2.4.1.283 3.5.1.112
B102401	Pentalenolactone	Antibiotic biosynthesis	K12250 K15907 K18056 K17747 K18091 K18057 K17476
B102402	Pentalenolactone	Antibiotic biosynthesis	4.2.3.7 1.4.15.32 1.14.11.35 1.1.1.340 1.14.13.170 1.14.11.36 1.14.19.8
B102601	Prodigiosin	Antibiotic biosynthesis	((K21780+K21781) K21782 K21783 K21784 K21785 K21786) (K21428 K21778 K21779) K21787
B102801	Pyocyanin	Antibiotic biosynthesis	K13063 K20261 K06998 K20260 K20262 K21103 K20940
B102802	Pyocyanin	Antibiotic biosynthesis	2.1.1.327 1.14.13.218
B102901	Pyrrolnitrin	Antibiotic biosynthesis	K14266 K19981 K14257 K19982
B104101	Validamycin A	Antibiotic biosynthesis	K19969 K20431 K20432 K20433 K20434 K20435 K20436 K20437 K20438
B104102	Validamycin A	Antibiotic biosynthesis	4.2.3.152 5.1.3.33 2.7.1.214 (2.6.1.M1,2.7.7.91) 2.5.1.135 3.1.3.101 2.4.1.338 1.14.11.52
B104201	Violacein	Antibiotic biosynthesis	K20086 (K20087+K20088) K20089 K20090
B104202	Violacein	Antibiotic biosynthesis	1.4.3.23 1.21.98.2 1.14.13.217 1.14.13.224
D010101	Triglyceride	Lipid degradation	(K01046,K12298,K16816,K13534,K14073,K14074,K14075,K14076,K22283, K14452,K22284,K14674,K14675,K17900) (K01054,K25824)
D010102	Triglyceride	Lipid degradation	(3.1.1.3,3.1.1.34) (3.1.1.34,3.1.1.79,3.1.1.116) (3.1.1.23,3.1.1.79)
D010201	Fatty acid	Lipid degradation	(K01897,K15013) (K00232,K00249,K00255,K06445,K09479) (((K01692,K07511,K13767) (K00022,K07516)),K01825,K01782,K07514,K07515,K10527) (K00632,K07508,K07509,K07513)
D010301	Oleate	Lipid degradation	6.2.1.3 1.3.8.8 4.2.1.17 1.1.1.35 2.3.1.16 1.3.8.8 4.2.1.17 1.1.1.35 2.3.1.16 1.3.8.8 4.2.1.17 (1.1.1.35,1.1.1.211) 2.3.1.16 5.3.3.8 4.2.1.74
D010401	Dicarboxylic acids	Lipid degradation	6.2.1.5 1.3.8.7 4.2.1.17 1.1.1.35 2.3.1.174
D020101	Cellulose	Polysaccharide degradation	(3.2.1.4,3.2.1.176,3.2.1.132,3.2.1.73) (3.2.1.176,3.2.1.4,3.2.1.14) (1.14.99.54,1.14.99.56,1.14.99.53) (1.14.99.54,1.14.99.53) 1.14.99.54 (1.14.99.54,1.14.99.56) (1.14.99.54,1.14.99.56,1.14.99.53) (3.2.1.4,3.2.1.8,3.2.1.21,3.2.1.25,3.2.1.45,3.2.1.58,3.2.1.73,3.2.1.74,3.2.1.75 ,3.2.1.78,3.2.1.91,3.2.1.104,3.2.1.123,3.2.1.132,3.2.1.149,3.2.1.151,3.2.1.16 4,3.2.1.168,3.2.1.73,3.2.1.39,3.2.1.52,3.2.1.132,3.2.1.146) (3.2.1.4,3.2.1.91) (3.2.1.4,3.2.1.176,3.2.1.132,3.2.1.73) (3.2.1.132,3.2.1.4,3.2.1.73,3.2.1.8,3.2.1.156) (3.2.1.4,3.2.1.6,3.2.1.21,3.2.1.73,3.2.1.74,3.2.1.91,3.2.1.151,3.2.1.165) (3.2.1.8,3.2.1.32,3.2.1.4) 3.2.1.4 (3.2.1.4,3.2.1.151,3.2.1.73,2.4.1.207) (3.2.1.4,3.2.1.151) (3.2.1.4,3.2.1.151,3.2.1.78) (3.2.1.4,3.2.1.8,3.2.1.21,3.2.1.25,3.2.1.45,3.2.1.58,3.2.1.73,3.2.1.74,3.2.1.75 ,3.2.1.78,3.2.1.91,3.2.1.104,3.2.1.123,3.2.1.132,3.2.1.149,3.2.1.151,3.2.1.16

			4,3.2.1.168,3.2.1.73,3.2.1.39,3.2.1.52,3.2.1.132,3.2.1.146) (3.2.1.4,3.2.1.91) (3.2.1.4,3.2.1.176,3.2.1.132,3.2.1.73) (3.2.1.132,3.2.1.4,3.2.1.73,3.2.1.8,3.2.1.156) (3.2.1.4,3.2.1.6,3.2.1.21,3.2.1.73,3.2.1.74,3.2.1.91,3.2.1.151,3.2.1.165) (3.2.1.8,3.2.1.32,3.2.1.4) 3.2.1.4 (3.2.1.4,3.2.1.151,3.2.1.73,2.4.1.207) (3.2.1.4,3.2.1.151) (3.2.1.4,3.2.1.151,3.2.1.78)
D020201	Xyloglucan	Polysaccharide degradation	(3.2.1.4,3.2.1.8,3.2.1.21,3.2.1.25,3.2.1.45,3.2.1.58,3.2.1.73,3.2.1.74,3.2.1.75,3.2.1.78,3.2.1.91,3.2.1.104,3.2.1.123,3.2.1.132,3.2.1.149,3.2.1.151,3.2.1.164,3.2.1.168,3.2.1.73,3.2.1.39,3.2.1.52,3.2.1.132,3.2.1.146) (3.2.1.4,3.2.1.6,3.2.1.21,3.2.1.73,3.2.1.74,3.2.1.91,3.2.1.151,3.2.1.165) (3.2.1.4,3.2.1.151,3.2.1.73,2.4.1.207) (2.4.1.207,3.2.1.103,3.2.1.39,3.2.1.6,3.2.1.73,3.2.1.81,3.2.1.83,3.2.1.151,3.2.1.181,3.2.1.178,3.2.1.35,3.2.1.181) (3.2.1.4,3.2.1.151) (3.2.1.4,3.2.1.151,3.2.1.78) (3.2.1.176,3.2.1.4,3.2.1.14) (3.2.1.4,3.2.1.150,3.2.1.151) (1.14.99.54,1.14.99.56) (3.2.1.20,3.2.1.22,3.2.1.24,3.2.1.84,3.2.1.48,3.2.1.10,3.2.1.177,4.2.2.13,2.4.1.161) (3.2.1.37,3.2.1.55,3.2.1.8,3.2.1.99,3.2.1.145,3.2.1.146) (3.2.1.8,3.2.1.32,3.2.1.4) (3.2.1.23,3.2.1.25,3.2.1.31,3.2.1.55,3.2.1.152,3.2.1.165,3.2.1.37,3.2.1.146)
D020301	Starch	Polysaccharide degradation	(3.2.1.1,3.2.1.41,2.4.1.19,3.2.1.54,3.2.1.93,3.2.1.10,3.2.1.133,3.2.1.135,3.2.1.20,3.2.1.60,3.2.1.68,3.2.1.70,3.2.1.98,3.2.1.116,2.4.1.18,5.4.99.16,2.4.1.25,2.4.1.4,2.4.1.7,3.2.1.141,5.4.99.11,5.4.99.15,3.2.1.33,2.4.99.16) 3.2.1.2 (3.2.1.1,3.2.1.22,3.2.1.41,3.2.1.54,2.4.1.18,2.4.1.25) 3.2.1.1 3.2.1.33 (3.2.1.3,3.2.1.70,3.2.1.28,2.4.1.2) (3.2.1.3,3.2.1.20,3.2.1.22)
D020401	Chitin	Polysaccharide degradation	(3.2.1.14,3.2.1.17,3.2.1.96) (3.2.1.14,3.2.1.17) (3.2.1.17,4.2.2.n1,3.2.1.14) (3.2.1.17,3.2.1.96) (3.2.1.52,3.2.1.140) (3.2.1.21,3.2.1.37,3.2.1.45,3.2.1.52,3.2.1.55,3.2.1.58,3.2.1.74,3.2.1.120,3.2.1.126) (3.2.1.4,3.2.1.8,3.2.1.21,3.2.1.25,3.2.1.45,3.2.1.58,3.2.1.73,3.2.1.74,3.2.1.75,3.2.1.78,3.2.1.91,3.2.1.104,3.2.1.123,3.2.1.132,3.2.1.149,3.2.1.151,3.2.1.164,3.2.1.168,3.2.1.73,3.2.1.39,3.2.1.52,3.2.1.132,3.2.1.146) (3.2.1.52,3.2.1.35,3.2.1.169) (3.2.1.21,3.2.1.37,3.2.1.45,3.2.1.52) (1.14.99.54,1.14.99.56,1.14.99.53) (1.14.99.54,1.14.99.53) (3.1.1.72,3.5.1.41)
D020501	Pectin	Polysaccharide degradation	(3.2.1.15,3.2.1.40,3.2.1.67,3.2.1.82,3.2.1.171,3.2.1.173) (4.2.2.2,4.2.2.9,4.2.2.10) (4.2.2.2,4.2.2.9) (4.2.2.2,4.2.2.9) 4.2.2.2 (4.2.2.23,4.2.2.24) 4.2.2.6 4.2.2.24 4.2.2.23 (3.2.1.40,3.2.1.174) 3.2.1.173 3.1.1.11 3.2.1.172 (3.2.1.122,3.2.1.20,3.2.1.22,3.2.1.86,3.2.1.139,3.2.1.67) 3.1.1.72 (3.2.1.23,3.2.1.25,3.2.1.31,3.2.1.55,3.2.1.152,3.2.1.165,3.2.1.37,3.2.1.146)
D020601	Alpha galactan	Polysaccharide degradation	3.2.1.49 (3.2.1.22,3.2.1.49,3.2.1.94,3.2.1.88) 3.2.1.22 (3.2.1.122,3.2.1.20,3.2.1.22,3.2.1.86,3.2.1.139,3.2.1.67) (3.2.1.20,3.2.1.22,3.2.1.24,3.2.1.84,3.2.1.48,3.2.1.10,3.2.1.177,4.2.2.13,2.4.1.161) (3.2.1.22,3.2.1.49,2.4.1.67,2.4.1.82) (3.2.1.3,3.2.1.20,3.2.1.22)
D020701	Beta-galactan	Polysaccharide degradation	3.2.1.89 (3.2.1.23,3.2.1.25,3.2.1.31,3.2.1.55,3.2.1.152,3.2.1.165,3.2.1.37,3.2.1.146) (3.2.1.23,3.2.1.165) 3.2.1.23 (3.2.1.21,3.2.1.23,3.2.1.25,3.2.1.31,3.2.1.37,3.2.1.38,3.2.1.62,3.2.1.74,3.2.1.85,3.2.1.86,3.2.1.105,3.2.1.108,3.2.1.117,3.2.1.118,3.2.1.119,3.2.1.125,3.2.1.147,3.2.1.149,3.2.1.161,3.2.1.175,3.2.1.182) (3.2.1.23,3.2.1.46)
D020801	Mixed-Linkage glucans	Polysaccharide degradation	3.2.1.71 (3.2.1.8,3.2.1.31,3.2.1.37,3.2.1.38,3.2.1.45,3.2.1.75,3.2.1.136) (3.2.1.39,3.2.1.58,3.2.1.73,3.2.1.175) (3.2.1.4,3.2.1.176,3.2.1.132,3.2.1.73) (3.2.1.132,3.2.1.4,3.2.1.73,3.2.1.8,3.2.1.156) (3.2.1.4,3.2.1.6,3.2.1.21,3.2.1.73,3.2.1.74,3.2.1.91,3.2.1.151,3.2.1.165) (3.2.1.4,3.2.1.8,3.2.1.21,3.2.1.25,3.2.1.45,3.2.1.58,3.2.1.73,3.2.1.74,3.2.1.75,3.2.1.78,3.2.1.91,3.2.1.104,3.2.1.123,3.2.1.132,3.2.1.149,3.2.1.151,3.2.1.164,3.2.1.168,3.2.1.73,3.2.1.39,3.2.1.52,3.2.1.132,3.2.1.146) (3.2.1.58,3.2.1.39) 3.2.1.39 (3.2.1.21,3.2.1.37,3.2.1.45,3.2.1.52,3.2.1.55,3.2.1.58,3.2.1.74,3.2.1.120,3.2.1.126) (2.4.1.207,3.2.1.103,3.2.1.39,3.2.1.6,3.2.1.73,3.2.1.81,3.2.1.83,3.2.1.151,3.2.1.181,3.2.1.178,3.2.1.35,3.2.1.181) (3.2.1.39,3.2.1.58,3.2.1.73,3.2.1.175) (3.2.1.58,3.2.1.39) (3.2.1.21,3.2.1.23,3.2.1.25,3.2.1.31,3.2.1.37,3.2.1.38,3.2.1.62,3.2.1.74,3.2.1.85,3.2.1.86,3.2.1.105,3.2.1.108,3.2.1.117,3.2.1.118,3.2.1.119,3.2.1.125,3.2.1.147,3.2.1.149,3.2.1.161,3.2.1.175,3.2.1.182)
D020901	Xylans	Polysaccharide degradation	(3.2.1.8,3.2.1.32,3.2.1.4) (3.2.1.78,3.2.1.100,3.2.1.32,3.2.1.73) (3.2.1.132,3.2.1.4,3.2.1.73,3.2.1.8,3.2.1.156) (3.2.1.4,3.2.1.8,3.2.1.21,3.2.1.25,3.2.1.45,3.2.1.58,3.2.1.73,3.2.1.74,3.2.1.75,3.2.1.78,3.2.1.91,3.2.1.104,3.2.1.123,3.2.1.132,3.2.1.149,3.2.1.151,3.2.1.164,3.2.1.168,3.2.1.73,3.2.1.39,3.2.1.52,3.2.1.132,3.2.1.146) (3.2.1.8,3.2.1.32) (3.2.1.8,3.2.1.31,3.2.1.37,3.2.1.38,3.2.1.45,3.2.1.75,3.2.1.136) (3.2.1.102,3.2.1.8,3.2.1.8) (3.2.1.51,3.2.1.8) (3.2.1.76,3.2.1.37) (3.2.1.21,3.2.1.37,3.2.1.45,3.2.1.52,3.2.1.55,3.2.1.58,3.2.1.74,3.2.1.120,3.2.1.126) (3.2.1.20,3.2.1.22,3.2.1.24,3.2.1.84,3.2.1.48,3.2.1.10,3.2.1.177,4.2.2.13,2.4.1.161) 3.2.1.37 (3.2.1.139,3.2.1.131)
D021001	Beta-mannan	Polysaccharide degradation	3.2.1.78 (3.2.1.78,3.2.1.100,3.2.1.32,3.2.1.73) (3.2.1.4,3.2.1.8,3.2.1.21,3.2.1.25,3.2.1.45,3.2.1.58,3.2.1.73,3.2.1.74,3.2.1.75,3.2.1.78,3.2.1.91,3.2.1.104,3.2.1.123,3.2.1.132,3.2.1.149,3.2.1.151,3.2.1.16

			4,3.2.1.168,3.2.1.73,3.2.1.39,3.2.1.52,3.2.1.132,3.2.1.146) (3.2.1.78,3.2.1.100,3.2.1.32,3.2.1.73) (2.4.1.281,2.4.1.319,2.4.1.320) 3.2.1.78 3.1.1.72 (3.2.1.23,3.2.1.25,3.2.1.31,3.2.1.55,3.2.1.152,3.2.1.165,3.2.1.37,3.2.1.146)
D021101	Alpha-mannan	Polysaccharide degradation	3.2.1.130 (3.2.1.101,3.2.1.20) (3.2.1.101,3.2.1.20) (3.2.1.113,3.2.1.24) (3.2.1.24,3.2.1.113,3.2.1.114,3.2.1.170) (3.2.1.106,3.2.1.84,3.2.1.20,3.2.1.170,3.2.1.208) 3.2.1.113 (2.4.1.281,2.4.1.319,2.4.1.320)
D021201	Arabinan	Polysaccharide degradation	(3.2.1.37,3.2.1.55,3.2.1.8,3.2.1.99,3.2.1.145,3.2.1.146) (3.2.1.11,3.2.1.57,3.2.1.95) (3.2.1.37,3.2.1.55,3.2.1.8,3.2.1.99,3.2.1.145,3.2.1.146) (3.2.1.4,3.2.1.8,3.2.1.37,3.2.1.55,3.2.1.73) (3.2.1.21,3.2.1.37,3.2.1.45,3.2.1.52,3.2.1.55,3.2.1.58,3.2.1.74,3.2.1.120,3.2.1.126) (3.2.1.55,3.2.1.37) 3.2.1.55
D021301	Mucin	Polysaccharide degradation	3.2.1.97 (3.2.1.22,3.2.1.49,3.2.1.94,3.2.1.88) 3.2.1.49 (2.4.1.211,2.4.1.247) (3.2.1.20,3.2.1.22,3.2.1.24,3.2.1.84,3.2.1.48,3.2.1.10,3.2.1.177,4.2.2.13,2.4.1.161) (3.2.1.22,3.2.1.49,2.4.1.67,2.4.1.82)
D030101	Lactose	Sugar degradation	3.2.1.85 5.3.1.26 2.7.1.144 4.1.2.40
D030201	Sucrose	Sugar degradation	2.7.1.211 3.2.1.48 2.7.1.4
D030302	D-Apiose	Sugar degradation	1.1.1.420 3.1.1.115
D030401	D-Arabinose	Sugar degradation	5.3.1.3 2.7.1.47
D030402	D-Arabinose	Sugar degradation	5.3.1.3 2.7.1.51 4.1.2.17 1.2.1.21
D030501	D-Mannose	Sugar degradation	2.7.1.191 5.3.1.8
D030502	D-Mannose	Sugar degradation	2.7.1.7 5.3.1.8
D030601	D-Xylose	Sugar degradation	5.3.1.5 2.7.1.17
D030602	D-Xylose	Sugar degradation	1.1.1.9 2.7.1.17
D030603	D-Xylose	Sugar degradation	1.1.1.424 3.1.1.68 4.2.1.82 4.2.1.141 1.2.1.26
D030604	D-Xylose	Sugar degradation	1.1.1.359 3.1.1.110 4.2.1.82 4.1.2.28 1.1.1.26 2.3.3.9
D030605	D-Xylose	Sugar degradation	1.1.1.175 3.1.1.110 4.2.1.82 4.2.1.1.141 1.2.1.26
D030701	L-Fucose	Sugar degradation	5.1.3.29 5.3.1.25 2.7.1.51 4.1.2.17
D030702	L-Fucose	Sugar degradation	5.1.3.29 4.2.1.68 1.1.1.M68 3.7.1.26
D030801	L-Rhamnose	Sugar degradation	5.1.3.32 5.3.1.14 2.7.1.5 4.1.2.19
D030802	L-Rhamnose	Sugar degradation	(1.1.1.173,1.1.1.378) 3.1.1.65 4.2.1.90 4.1.2.53
D030803	L-Rhamnose	Sugar degradation	(1.1.1.173,1.1.1.378) 3.1.1.65 4.2.1.90 1.1.1.401 3.7.1.26
D030901	Galactose	Sugar degradation	K01785 K00849 K00965 K01784
D031001	NeuAc	Sugar degradation	4.1.3.3 5.1.3.8
D031002	NeuAc	Sugar degradation	4.1.3.3 2.7.1.60 5.1.3.9
D040101	Albumin	Protein degradation	A1A (S8A C13 (S1A (S8B S26B)))
D040201	Actin	Protein degradation	S1A (M10A A1A M35 (M9B S1B S1C C1A (M12A C14A A2A C13 M10C)))
D040301	Collagen	Protein degradation	C1A M10A (M12B (S1A A1A (C13 M35 (M12A S1D S8A A2A M4 S8B))))
D040401	Elastin	Protein degradation	M10A (S1A (M23A S01C C1A))
D040501	Glutelin	Protein degradation	S26B S8A S9A
D040601	Keratin	Protein degradation	S1D M12A S1A
D040701	Tropomyosin	Protein	S1B (C2A C1A S1A (M4 C13 M10A (S1D M12A)))

		degradation	
D040801	Troponin	Protein degradation	C2A (A2A M35 C14A)
D050101	Serine	Amino acid degradation	4.3.1.17,4.3.1.18
D050201	Threonine	Amino acid degradation	4.3.1.19
D050301	Cysteine	Amino acid degradation	4.4.1.1,4.4.1.28
D050302	Cysteine	Amino acid degradation	2.6.1.3 2.8.1.2
D050401	Methionine	Amino acid degradation	4.4.1.11
D050501	Valine	Amino acid degradation	2.6.1.42 1.2.1.25 1.3.8.5 4.2.1.150 3.1.2.4 1.1.1.31 1.2.1.27
D050502	Valine	Amino acid degradation	2.6.1.42 4.1.1.72 1.1.1.1
D050601	Isoleucine	Amino acid degradation	2.6.1.42 1.2.1.25 1.3.8.5 4.2.1.150 1.1.1.178 2.3.1.16
D050602	Isoleucine	Amino acid degradation	2.6.1.42 1.2.7.7
D050701	Leucine	Amino acid degradation	K00826 (((K00166+K00167),K11381)+K09699+K00382) (K00253,K00249) (K01968+K01969) (K05607,K13766) K01640
D050801	Lysine	Amino acid degradation	4.1.1.18 2.6.1.82 1.2.1.19 2.6.1.48 1.2.1.20 1.14.11.64 1.1.5.13
D050802	Lysine	Amino acid degradation	1.13.12.2 3.5.1.30 1.6.1.48 1.2.1.20 2.8.3.13
D050803	Lysine	Amino acid degradation	2.6.1.36 1.2.1.31
D050804	Lysine	Amino acid degradation	4.1.1.18 2.6.1.82 1.2.1.19 2.6.1.48 1.2.1.20 2.8.3.13
D050805	Lysine	Amino acid degradation	K01582 K09251 K00137 K07250 K00135 K15737 K15736
D050806	Lysine	Amino acid degradation	K00468 K01506 (K14268,K07250) K00135 ((K15737 K15736),(K01041 K00252 (K01692,K01825,K01782) (K01825,K01782) K00626))
D050901	Arginine	Amino acid degradation	4.1.1.19 3.5.3.11
D050902	Arginine	Amino acid degradation	4.1.1.19 3.5.3.12 3.5.1.53
D050903	Arginine	Amino acid degradation	1.13.12.1 3.5.1.4 3.5.3.7
D050904	Arginine	Amino acid degradation	K01476 K01581
D050905	Arginine	Amino acid degradation	K00613 K00542 K00933
D050906	Arginine	Amino acid degradation	(K01583,K01584,K01585,K02626) K01480 K01611 K00797

D051001	Proline	Amino acid degradation	5.1.1.4 1.21.4.1
D051101	Glutamate	Amino acid degradation	1.4.1.2,1.4.1.3
D051102	Glutamate	Amino acid degradation	2.6.1.1 4.3.1.1
D051103	Glutamate	Amino acid degradation	1.4.1.2 1.1.1.399 2.8.3.12 4.2.1.167 7.2.4.5
D051104	Glutamate	Amino acid degradation	5.4.99.1 4.3.1.2 4.2.1.34 4.1.3.22
D051105	Glutamate	Amino acid degradation	4.1.1.15
D051201	Histidine	Amino acid degradation	K01745 K01712 K01468 (K01479,K00603,K13990,(K05603 K01458))
D051301	Tryptophan	Amino acid degradation	(K00453,K00463) (K01432,K14263,K07130) K00486 K01556 K00452 K03392 (K10217,K23234)
D051302	Tryptophan	Amino acid degradation	4.1.99.1
D051601	Beta-alanine	Amino acid degradation	2.6.1.18 1.2.1.18
D051602	Beta-alanine	Amino acid degradation	2.6.1.120 1.1.1.298
D051701	Ornithine	Amino acid degradation	4.1.1.17 ((2.6.1.82 1.2.1.19),(6.3.1.11 1.4.3.M3 1.2.1.00 3.5.1.94))
D051801	GABA	Amino acid degradation	2.6.1.19 (1.2.1.24,1.2.1.16)
D051802	GABA	Amino acid degradation	2.6.1.19 1.1.1.61 2.8.3.M6 4.2.1.120 1.3.1.109 (2.8.3.1,2.8.3.8)
D060101	Nitrate	Nitrogen compound degradation	((K00370+K00371+K00374),(K02567+K02568)) ((K00362+K00363),(K03385+K15876))
D060102	Nitrate	Nitrogen compound degradation	1.7.5.1 1.7.2.1 1.7.2.5 1.7.2.4
D060103	Nitrate	Nitrogen compound degradation	1.9.6.1 1.7.2.2
D060105	Nitrate	Nitrogen compound degradation	1.7.7.2 1.7.7.1
D060201	Urea	Nitrogen compound degradation	6.3.4.6 3.5.1.54
D060202	Urea	Nitrogen compound degradation	3.5.1.5
D060301	Urate	Nitrogen compound	1.7.3.3 3.5.2.17 4.1.1.97

		degradation	
D060302	Urate	Nitrogen compound degradation	1.14.13.113 3.5.2.17 4.1.1.97
D060401	GlcNAc	Nitrogen compound degradation	2.7.1.59 3.5.1.25 3.5.99.6
D060402	GlcNAc	Nitrogen compound degradation	3.5.1.25 3.5.99.6
D060601	Allantoin	Nitrogen compound degradation	3.5.2.5 3.5.3.9 3.5.3.26 (1.1.1.350,1.1.1.154) 2.1.3.5
D060602	Allantoin	Nitrogen compound degradation	3.5.2.5 3.5.3.4 4.3.2.3
D060603	Allantoin	Nitrogen compound degradation	3.5.2.5 3.5.3.9 3.5.3.26 4.3.2.3
D060701	Creatinine	Nitrogen compound degradation	3.5.4.21 3.5.2.14 3.5.1.59 1.5.3.1
D060801	Betaine	Nitrogen compound degradation	2.1.1.5 1.5.3.10 (1.5.3.24,1.5.3.1)
D060901	L-carnitine	Nitrogen compound degradation	1.14.13.239 1.2.1.4 1.1.1.38
D061001	Methylamine	Nitrogen compound degradation	1.4.9.1
D061002	Methylamine	Nitrogen compound degradation	6.3.4.12 2.1.1.21 1.5.99.5
D061101	Phenylethylamine	Nitrogen compound degradation	(1.4.3.4,1.4.3.21) 1.2.1.39
D061201	Hypotaurine	Nitrogen compound degradation	2.6.1.77 1.2.1.3
D061301	Taurine	Nitrogen compound degradation	2.6.1.77
D061303	Taurine	Nitrogen compound degradation	2.5.1.55
D070101	2,3-Butanediol	Alcohol degradation	1.1.1.4,1.1.1.76
D070201	Ethanol	Alcohol degradation	1.1.1.1 1.2.1.10

D070202	Ethanol	Alcohol degradation	1.1.1.1 1.2.1.3 6.2.1.1
D070401	Glycerol	Alcohol degradation	2.7.1.30 1.1.5.3
D070402	Glycerol	Alcohol degradation	1.1.1.6 2.7.1.29
D070403	Glycerol	Alcohol degradation	4.2.1.30 1.1.1.202
D070404	Glycerol	Alcohol degradation	1.1.1.6 2.7.1.121
D070501	Propylene glycol	Alcohol degradation	4.2.1.28 (1.1.1.1,(1.2.1.87 2.3.1.222 (2.7.2.1,2.7.2.7,2.7.2.14,2.7.2.15)))
D070601	Ethylene glycol	Alcohol degradation	1.1.1.77 1.2.1.21
D070801	Phytol	Alcohol degradation	1.1.1.1 1.2.1.3 6.2.1.3 1.3.1.38
D070901	Polyvinyl alcohol	Alcohol degradation	1.1.2.6
D080101	Toluene	Xenobiotic degradation	(K15760+K15761+K15763+K15764) K00055 K00141
D080103	Toluene	Xenobiotic degradation	K07540 (K07543+K07544) K07545 K07546 (K07547+K07548) (K07549+K07550)
D080201	Xylene	Xenobiotic degradation	(K15757+K15758) K00055 K00141
D080402	Benzene	Xenobiotic degradation	K16249+K16243+K16244+K16242+K16245+K16246
D080501	Benzoate	Xenobiotic degradation	(K05549+K05550+K05784) K05783
D080502	Benzoate	Xenobiotic degradation	K04116 K04117 K07534 K07535 K07536
D080601	Anthranilate	Xenobiotic degradation	(K05599+K05600+K11311),(K16319+K16320+K18248+K18249)
D080701	Catechol	Xenobiotic degradation	K03381 K01856 K03464 (K01055,K14727)
D080702	Catechol	Xenobiotic degradation	(K00446,K07104) ((K10217 K01821 K01617),K10216) (K18364,K02554) (K18365,K01666) (K18366,K04073)
D080801	Cumate	Xenobiotic degradation	(K10619+K16303+K16304+K18227) K10620 K10621 K10622 K10623
D080901	Biphenyl	Xenobiotic degradation	(K08689+K15750+K18087+K18088) K08690 K00462 K10222
D081001	Carbazole	Xenobiotic degradation	K15751 (K15754+K15755) K15756
D081101	Benzoyl-CoA	Xenobiotic degradation	((K04112+K04113+K04114+K04115),(K19515+K19516)) K07537 K07538 K07539
D081201	Naphthalene	Xenobiotic degradation	(K14579+K14580+K14578+K14581) K14582 K14583 K14584 K14585 K00152
D081301	Salicylate	Xenobiotic	K18242+K18243+K14578+K14581

		degradation	
D081401	Terephthalate	Xenobiotic degradation	(K18074+K18075+K18077) K18076
D081501	Phthalate	Xenobiotic degradation	(K18068+K18069) K18067 K04102
D081601	Phenylacetate	Xenobiotic degradation	K01912 (K02609+K02610+K02611+K02612+K02613) K15866 K02618 K02615 K01692 K00074
D081701	Trans-cinnamate	Xenobiotic degradation	(((K05708+K05709+K05710+K00529) K05711),K05712) K05713 K05714 K02554 K01666 K04073
D081801	Caffeine	Xenobiotic degradation	K21722 K21723 K21724
D081901	Mercury	Xenobiotic degradation	4.99.1.2 1.16.1.1
D090101	Penicillin	Antibiotic degradation	K18698,K18699,K18796,K18767,K18797,K19097,K19317,K18768,K18970, K19316,K22346,K18795,K19218,K19217,K17836,K18766
D090201	Carbapenem	Antibiotic degradation	K17837,K18782,K18781,K18780,K19099,K19216
D090301	Cephalosporin	Antibiotic degradation	K19095,K19096,K19100,K19101,K19214,K19215,K20319,K20320,K01467
D090401	Oxacillin	Antibiotic degradation	K17838,K18790,K18791,K19098,K18792,K19213,K21276,K18793,K18971, K22352,K19209,K18976,K18973,K18794,K18972,K21277,K19210,K19211, K19212,K22335,K19319,K22331,K22351,K19320,K19318,K19321,K19322, K21266,K22334,K22333,K22332
D090501	Streptogramin	Antibiotic degradation	K19349,K19350
D090601	Fosfomycin	Antibiotic degradation	K21252
D090701	Tetracycline	Antibiotic degradation	K08151,K08168,K18214,K18218,K18220,K18221
D090801	Macrolide	Antibiotic degradation	K06979,K08217,K18230,K18231,K21251
D091001	Chloramphenicol	Antibiotic degradation	K00638,K08160,K18552,K18553,K18554,K19271
D091101	Lincosamide	Antibiotic degradation	K18236,K19349,K19350,K19545
D091201	Streptothricin	Antibiotic degradation	K19273,K20816

Figure S1:
Top 2% bacteria drivers of the enterotypes identified by Dirichlet Multinomial Models. Each bacteria is given a value for the contribution it makes to the community type. The model identified two community types at days 7 and 21, but recognised 3 community types at day 35. Trial C individuals were always clustered together except for day 7, where some animals from trial C grouped with animals from trials A and B, as can be seen in the RDA (Figure 1b). We defined the distinct enterotype based on the community type 2, where Campylobacterota and Bacteroidota strains are more abundant and all animals belong to trial C. The rest of the communities were defined as the standard enterotype.

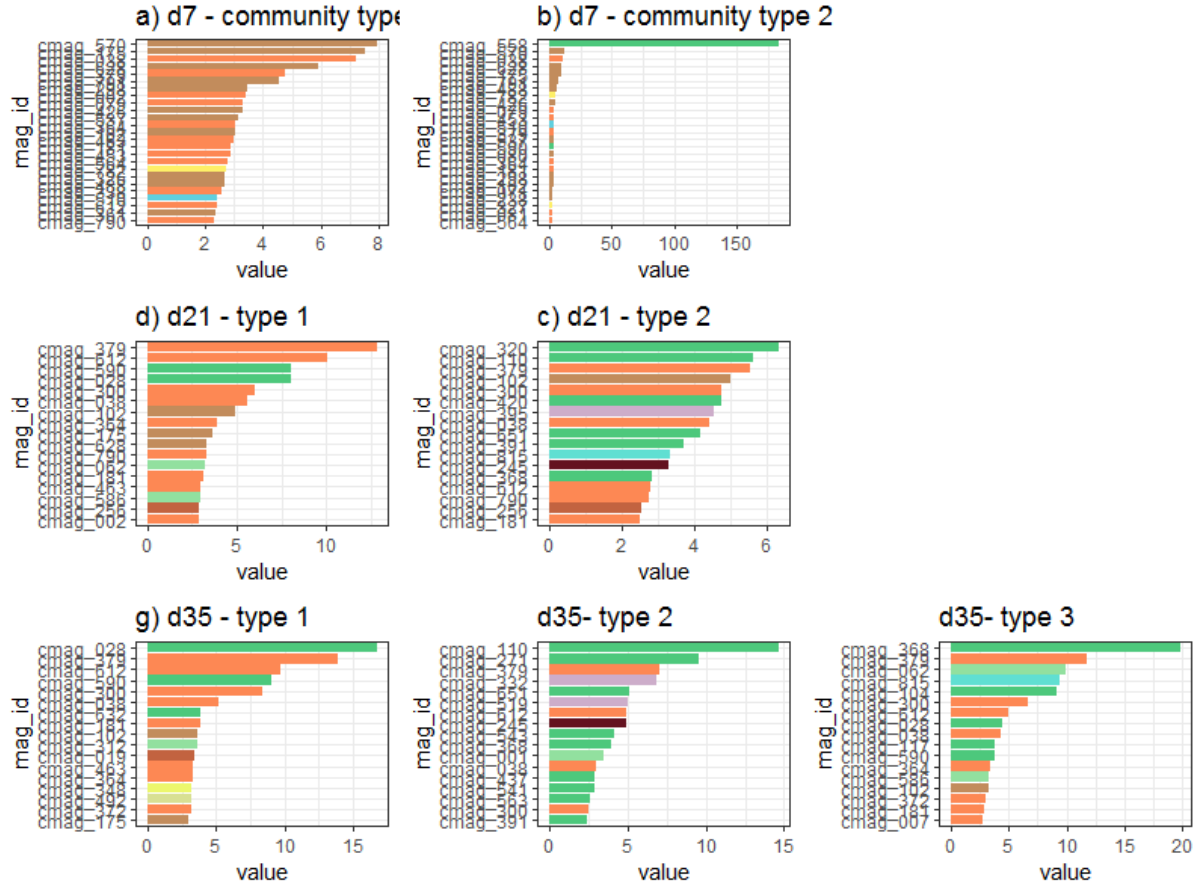


Figure S2:

Relative abundances of each bacterial strain for the standard and distinct enterotypes for the three sampling points, 7, 21 and 35.

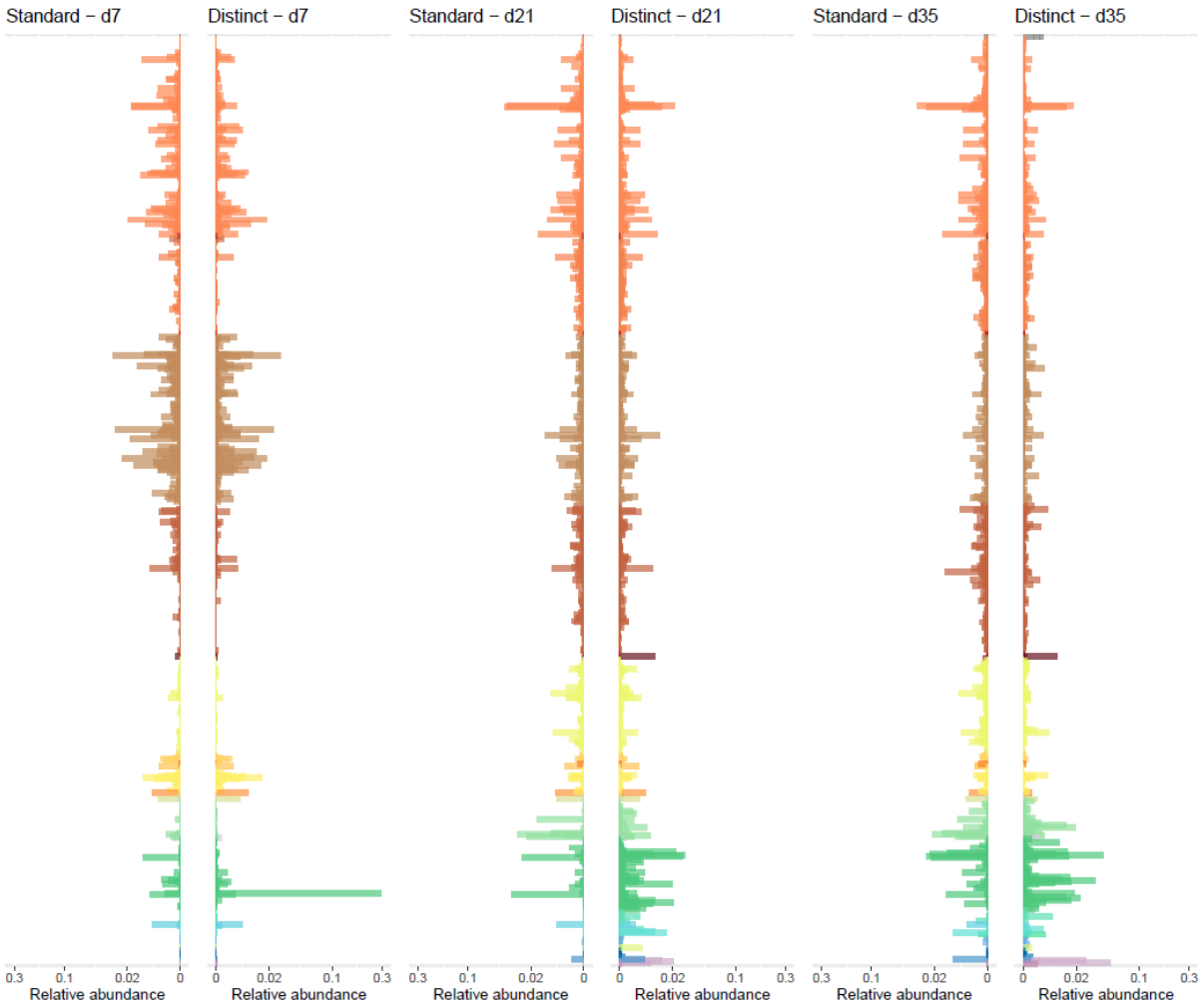


Figure S3:

Spanning tree of joint genome-scale metabolic network model between *Campylobacter jejuni* and *C. coli*. Campylobacter_D jejuni: ERR4836918_bin_11 Campylobacter_D coli: ERR4836965_bin_9.

https://fluxer.umbc.edu/model?id=f7dd951a9e005ee384ef401dda3e7c111a89d711_2d5898fafefe218fd2c2bb5aba37f1a7e75398404_obj_merge#

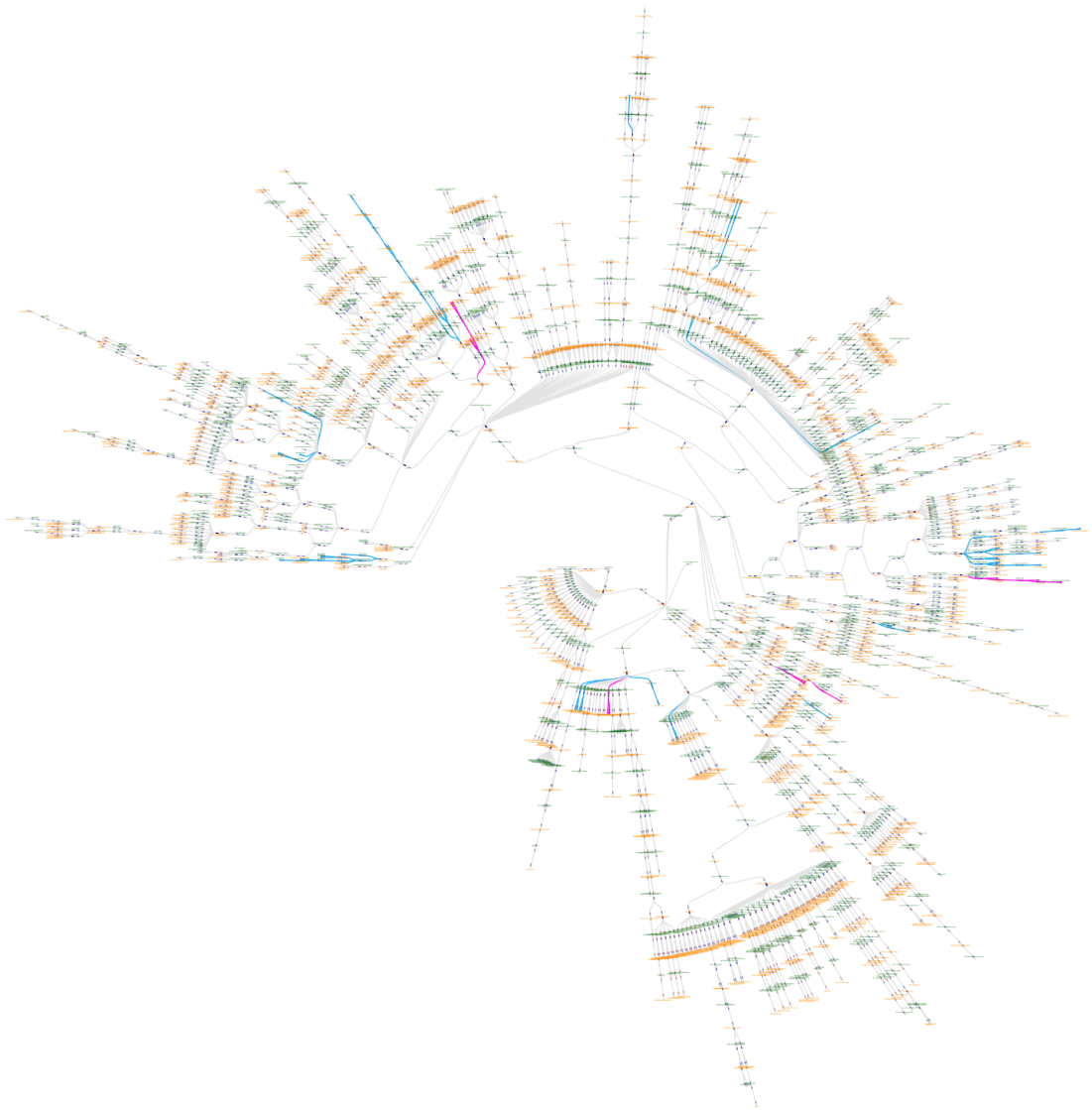
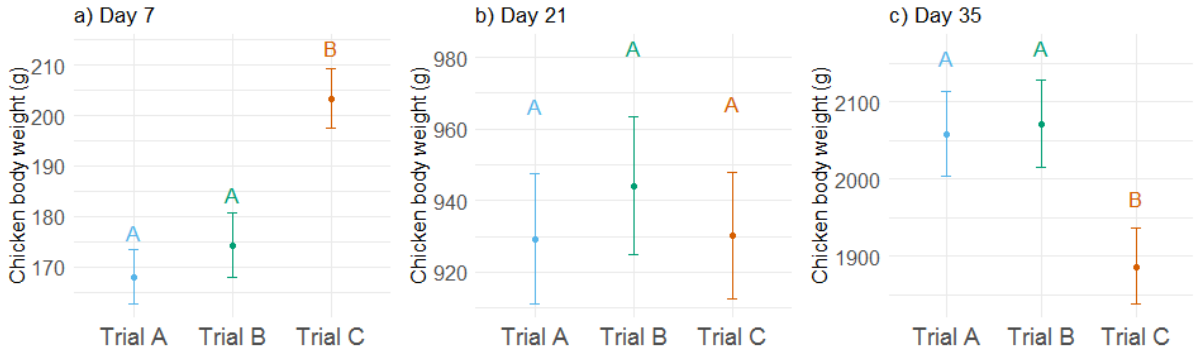


Figure S4:
Body weight changes between experimental trials across the three sampling time points, days 7, 21 and 35. We used linear mixed effect models as implemented in the R package lme4 (Bates et al. 2014). Chicken body weight (g) was used as response variable and trial, chicken age, sex and genetic line were included as fixed explanatory variables. To account for the fact that chickens were nested within pens we included a pen-level random intercept (1|pen). Models were plotted separated by sampling time (a-c) and together (d). Different letters in panel a-c represent significant differences (*p*-value < 0.05) between groups, based on a Tukey's HSD *post hoc* test.



Supplementary Figure S5:
Volcano plots of Differential Gene Expression analyses carried out between both enterotypes in each one of the sampling periods. Points in red represent genes that are differentially expressed, i.e., had a corrected p-value under 0.05 through the Benjamini-Hochberg method.

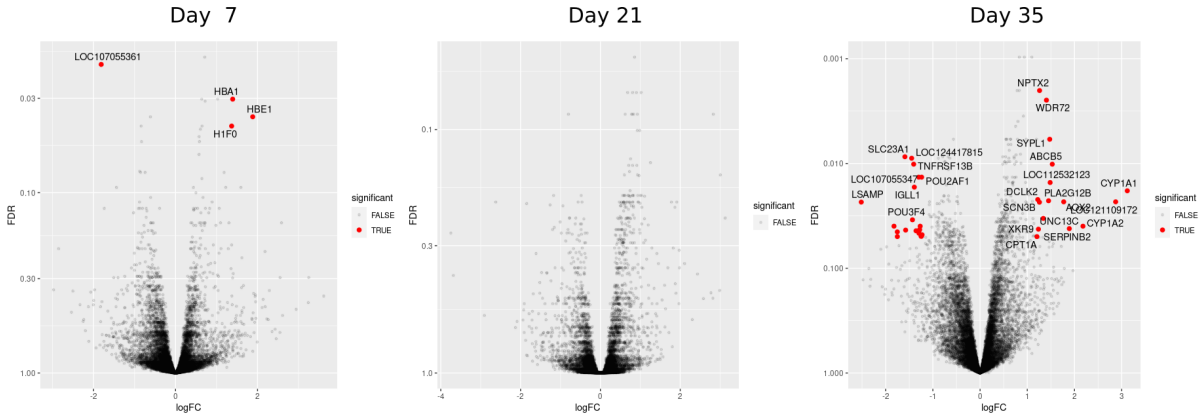
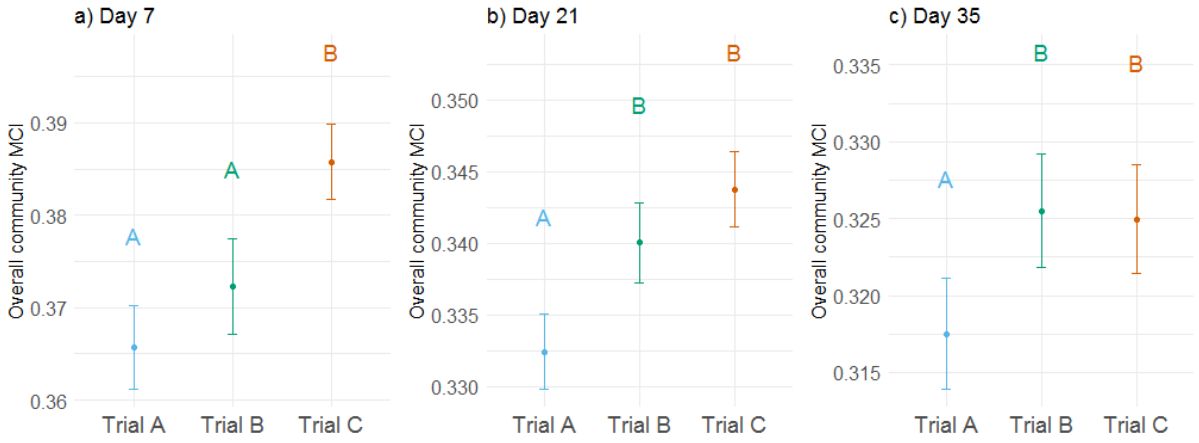


Figure S6:
Overall community MCI values between experimental trials across the three sampling time points, days 7, 21 and 35. We used linear mixed effect models as implemented in the R package glmmPQL() of R package MASS (Venables and Ripley 2002). Overall community MCI was used as response variable and trial, chicken age, sex and genetic line were included as fixed explanatory variables. We used a quasibinomial distribution with logit link function because the response variable was fractional (i.e. values were between 0 and 1). To account for the fact that chickens were nested within pens we included a pen-level random intercept (1|pen). Models were plotted separated by sampling time (a-c). Different letters in panel a-c represent significant differences (p-value < 0.05) between groups, based on a Tukey's HSD *post hoc* test.



General discussion and concluding remarks

Chapter 6

Overview of thesis outcomes

The advent of omic technologies, together with a higher awareness of microbial communities associated with animals and plants, has led to a rapid expansion of hologenomic studies (Rosenberg & Zilber-Rosenberg, 2018). Today, hologenomics is applied to understand a myriad of biological processes, including dietary adaptations (Cheng et al., 2023), phenotypic plasticity (Baldo et al., 2023; Fontaine & Kohl, 2023), and evolutionary processes (Rasmussen et al., 2023). The food industry is equally immersed in understanding animal-microbiota systems to improve animal performance and welfare, and thereby acquiring sustainable practices (He et al., 2023; X. Wang et al., 2019; Xue et al., 2020). However, there were many challenges and limitations related to studies on host-microbiota systems when the project started back in 2019, as discussed in Annex 1. HoloFood aimed to develop a holo-omic framework for gaining greater knowledge of host-microbiota interactions in broiler chickens reared under intensive farming conditions. Within the project, this doctoral thesis intended to investigate host-microbiota interplay through integrating (meta)genomics and (meta)transcriptomics.

In Chapter 3, my colleagues and I came to the conclusion that trial-and-error strategies need to be replaced by knowledge-based approaches that aim to comprehend the intricate biological processes unfolding chicken performance. Such knowledge should not be limited to results provided by traditional methods used to monitor animal health and nutrition, but should leverage the novel omic techniques that provide a much more detailed overview of the biomolecular mechanisms taking place in the organism.

Based on multi-omic data, Chapter 4 revealed that we still lack a proper understanding of the functional dynamics of chickens' caecal microbiota. We found that while functional diversity increased as chickens grew, the overall functional capacity of the community decreased. By clustering gene annotations into functional traits, we for the first time saw that metabolic capacity was inversely correlated with animal performance. Analysing functional characteristics of microbial communities opens up new avenues to understand the mechanistic interactions between both domains.

Lastly, we determined in Chapter 5 that zoonotic agents need to be studied in combination with the host and its microbiota to better understand the tripartite interactions among them. Genome-scale metabolic networks revealed that the distinct enterotype was metabolically favourable for *Campylobacter* spp., and functional activity of the microbial community could explain the reduced chicken growth. These findings pave the way for exploring strategies to manipulate early-life microbiota and prevent *Campylobacter* colonisation.

General discussion

Searching for signs of hologenomic variation

This thesis aimed to study host-microbiota variability in intensively bred broilers. For this purpose, 3 replicate trials were performed, where chickens from 2 genetic lines and 2 sexes were distributed in pens and grown under 3 dietary treatments. When the study was

designed, we assumed that most of the differences in broiler performance would be explained by the experimental groups. Nevertheless, the analyses explained in Chapter 3 revealed that the differences in performance between pens were negligible compared to the inter-individual and inter-experimental variabilities observed in the trials. This was mainly due to the limited effect of treatments, which could be partly explained by the high complexity of the study design, which was not oriented towards maximising treatment outcomes. But even if we had simplified the study, the effect sizes would still be very limited, and probably barely significant for production purposes. In addition, the analyses explained in Chapter 4 showed no significant effect of treatment on microbiome composition, nor evidence of the probiotic strains in the bacterial genome catalogue, which prevented further attempts to explore the reasons for their poor success. Thus, we were unable to address issues related to *in vivo* experiments that are currently under debate: “How do additives modulate gut microbiota?”, “To what extent do host-microbe characteristics influence the outcome of additives?”, and “Does host colonisation success influence the outcome of probiotics?” (Suez et al., 2019, 2020). However, we faced this caveat as an opportunity to explore host-microbiota interactions related to the inter-individual and inter-experiment variabilities that emerged in our study. W(Bilal et al., 2021; Rehman et al., 2020) usually considered biological noise and assumed to be uncontrollable in animal trials (Bilal et al., 2021; Rehman et al., 2020) to study the relationship between the two domains at the animal level. In addition, the unexpected differences between experiments prompted us to delve into the connections between zoonotic agents, the rest of the microbiota, and the host.

Hologenomic perspective revealed novel avenues for microbiome monitoring, selecting probiotic candidates and preventing zoonotic agents

Caecal genome-resolved metagenomics enabled us to assemble 822 bacterial genomes, of which 435 constituted new species after comparing the generated catalogue to the most extensive one available at the time (Gilroy et al., 2021). Moreover, multi-omic datasets generated for 613 chicken individuals across the three sampling points, a sample collection frequency similar to previous studies (Ocejo et al., 2019; Zhou et al., 2021), revealed compositional and functional dynamics of the caecal microbiome at unprecedented levels. With the intention of delving into the possible implications of the microbiota on host performance, we distilled the millions of microbial gene annotations into 170 genome-inferred functional traits (GIFTs) (Koziol et al., 2023), which also enabled us to generate quantitative estimations of gene expression associated with these GIFTs, yielding a unique functional overview of the metabolic capacity and activity of gut microbiomes. Such a data clustering and filtering strategy drastically increased our interpretative capacity, by decreasing computational requirements and gaining statistical power for downstream analyses (Shaffer et al., 2020). For instance, functional distillation was essential to unveil discordances between functional diversity and functional capacity, as shown in Chapter 4, as well as a functional specialisation of the distinct microbiome as a result of a compositional divergence from the standard microbiome I included in Chapter 5. In fact, both observations were for the first time correlated with chicken performance. This provided initial observations to the

questions posed after analysing performance and health parameters, which were “Are chicken caecal microbiomes correlated with chicken performance, and if so why?” and “Is *Campylobacter* spp. responsible for the reduced chicken weight?”. While the implications of the microbiome in host biological processes are largely acknowledged, there is still much debate about what is a beneficial microbiota in a production context (Bindari & Gerber, 2022; Diaz Carrasco et al., 2019). Our microbial multi-omic analyses significantly contributed to this discussion not only unveiling community-wide trends, but also species-level interactions, associated with chicken growth.

In addition, multi-omics casted light onto hitherto understudied taxa, such as the recently described TANB77, UBA1242, RF39, and RF32 clades comprising uncultured microorganisms. These bacterial taxa share many characteristics, including reduced-genomes and several auxotrophies, that have raised interest in the multiple metabolic dependencies they may have with the host (Chklovski et al., 2023). Indeed, they have been found in bacterial catalogues of dozens of vertebrate species (Antton Alberdi, pers. comm.), which could suggest a close evolutionary trajectory with vertebrates. In salmonids, which possess less complex and diverse microbiomes, a study shed more light on the symbiotic relation these bacteria might have with the host by studying *Mycoplasma*, a genus from the phylum Firmicutes like RF39, and with genomic features similar to the latter (Rasmussen et al., 2021). Further analyses showed that bacterial genomic variations were better explained by the genetic background of the host rather than the sampling location, suggesting a joint evolutionary trajectory (Rasmussen et al., 2023). In our study, we employed hierarchical modelling of species communities (HMSC) framework, a very versatile approach and gives the opportunity to easily handle big metagenomic datasets for later associating (Tikhonov et al., 2020), to explore their temporal dynamics across chicken development. Based on the results, we classified them as *increasers* or *decreasers* and then associated them with chicken performance. The mentioned taxa increased with chicken age, and positively correlated with chicken body weight in the last stages of production. Furthermore, their active role in producing SCFAs gives us clues about the potential benefits for the host. In fact, some of these taxa were correlated with host metabolic balance in chickens (Jiang et al., 2023) and mice (L. Wang et al., 2017; Wu et al., 2023). It could be that these taxa may even stop the colonisation of opportunistic or pathogenic bacteria (Bozzi et al., 2021). Therefore, in a context where broiler chickens have reduced microbial community because of the characteristics of intensive production practices (Rychlik, 2020), it seems pertinent to focus on those bacteria that may have closer evolutionary relationship with the host rather than administering non-indigenous strains (Anee et al., 2021). Future analyses about their phylogenomics, comparative genomics and metabolic network reconstructions, while taking into account the genomic background of the host, could greatly benefit the research community to better understand their shared evolution and potential benefits in production.

Lastly, we discovered not a single but three Campylobacteraceae strains and their colonisation dynamics, which illustrates how opportunistic or pathogenic colonisations are often orchestrated by several strains (Diaz Caballero et al., 2023). *Campylobacter jejuni*

peaked at day 21, but then declined to give rise to other two strains *C. coli* and *Helicobacter pullorum*. However, we are not certain whether this is a case of competition or commensalism between them, as some individuals showed all three strains at days 21 and 35. On the origin of the colonisation, our study showed a change in the microbial composition already for day 7. This leaves more room for action to redirect the microbiome before *Campylobacter* takes hold, as the window of opportunity of *Campylobacter* is considered to be between days 14-21 (Ijaz et al., 2018). Moreover, identifying *Bacteroides fragilis_A* as the primary driver of the distinct microbiome before *Campylobacter* colonisation is a major step forward, as it can be used as an early biomarker. Unfortunately, we could not know whether the excessive colonisation of *B. fragilis_A* was due to a cross-contamination on the farm or if it happened during transport from the breeding companies (Mota-Gutierrez et al., 2022). Metabolites obtained from genome-scale metabolic networks (Karp et al., 2021), revealed that the distinct microbiome provided numerous essential metabolites to *Campylobacter*, which most of them were acquired from *B. fragilis_A*. This *B. fragilis_A-Campylobacter* interaction could be just one of the many possibilities that can exist. Moreover, metabolite exchange could occur in the opposite direction, thus being an interdependent relationship rather than a codependent one (Culp & Goodman, 2023). It is also unclear the possible effects *Campylobacter* has on the host, as we did not see any evidence of an immune response in the intestinal samples. Transcriptomics of the spleen, which was collected for future analyses as detailed in Chapter 3, may provide more information about this matter. In summary, future works studying pathogen/opportunistic colonisations could benefit from a host-microbiome-coloniser perspective.

Technical challenges and opportunities of holo-omics

Being part of an European project means that this thesis benefited from a massive collective effort for gathering samples, generating data and analysing results; but it was not without technical constraints. We expected high host contamination in ileal and caecal metagenomic samples (Rasmussen et al., 2021). However, we obtained highly uneven host DNA for both intestinal sections. With more than 90% of host DNA in ileum samples, we ultimately decided to remove them from downstream analyses. Studying both sections would have enabled us to have a broader view of host-microbiota interactions through the gastrointestinal tract (Huang et al., 2018; Zhang et al., 2022). Alternatively, the disparate host percentages from both intestinal sections was used to explore a strategy to recover host genomes from metagenomics samples in Annex 2. This strategy was tested because host genomics was initially assessed to see whether host genetic variations could partly explain microbiota changes with noteworthy impact on performance. For this purpose, we made an effort to generate reference genomes for both chicken genetic lines (Rhie et al., 2021), but their assembly was delayed due to Covid-19. Despite unexpected setbacks, Edward Rice and colleagues successfully published a chicken pangenome in Annex 3. In retrospect, analysing genetic variations for the purposes of this thesis would increase the interpretative capacity, but after seeing the robust trends of the microbiome, as well as the small differences between chicken genomic backgrounds in the annexes, much contribution from the host would not be expected.

Lastly, this dissertation successfully employed a multi-staged approach that facilitated the formulation of hypotheses with a clear directionality (Graw et al., 2021). We combined metagenomics, metatranscriptomics and host transcriptomics with host performance parameters, and obtained significant results despite the noise, likely generated by the high variability of the experimental design. The results were very consistent across chicken genetic lines, origins and sexes. That is, the positive association between low-capacity bacteria and weight was robust between experimental groups, but at the same time we did not see significant differences in the abundance of either *Campylobacter* spp. or *B.fragilis_A* strains between pens. In addition, we observed these clear temporal trends at the population level despite having a cross-sectional study, as we prioritised invasive samples instead of repeatedly collecting faecal samples from the same individual. Therefore, we have reasonable evidence to believe that these results could be generalised to other genetic lines, as long as the particularities of the administered diet are taken into account.

Future perspectives

Holo-omics has indeed emerged as a powerful tool to unveil biological connections that would remain hidden with targeted methods (Alberdi et al., 2021). However, the effectiveness of this approach is counterbalanced by the high costs of data generation, as well as the advanced computer and statistical skills for integrating multiple omic layers (Graw et al., 2021). Therefore, conducting exploratory metadata analyses and preliminary assessments on a limited set of individuals becomes crucial for deciding the feasibility of employing holo-omics, and if so, identifying the optimal omic layers for the study (Nyholm et al., 2020). Despite these challenges, our study highlights the need to move from trial-and-error experiments to knowledge-based strategies for testing novel feeds, additives, or drugs.

This dissertation comprehensively explores the caecal microbiomes of the chicken from a hologenomic perspective, focusing on their temporal trends and their relation with the host. Throughout the analytical process, specific taxa were highlighted that played major roles in host energy balance. By harnessing available resources, future studies have the opportunity to integrate the generated data with publicly available datasets, enabling a comprehensive exploration of the bacterial genomic attributes of highlighted taxa across various vertebrates.

References

Alberdi, A., Andersen, S. B., Limborg, M. T., Dunn, R. R., & Gilbert, M. T. P. (2021). Disentangling host–microbiota complexity through hologenomics. *Nature Reviews. Genetics*, 1–17. https://doi.org/10.1038/s41576-021-00421-0

Anee, I. J., Alam, S., Begum, R. A., Shahjahan, R., & Khandaker, A. M. (2021). The role of probiotics on animal health and nutrition. *The Journal of Basic and Applied Zoology*, *82*(1), 52. https://doi.org/10.1186/s41936-021-00250-x

Baldo, L., Tavecchia, G., Rotger, A., Igual, J. M., & Riera, J. L. (2023). Insular holobionts: persistence and seasonal plasticity of the Balearic wall lizard (Podarcis lilfordi) gut microbiota. *PeerJ*, *11*, e14511. https://doi.org/10.7717/peerj.14511

Bilal, M., Si, W., Barbe, F., Chevaux, E., Sienkiewicz, O., & Zhao, X. (2021). Effects of novel probiotic

strains of Bacillus pumilus and Bacillus subtilis on production, gut health, and immunity of broiler chickens raised under suboptimal conditions. *Poultry Science*, *100*(3), 100871. https://doi.org/10.1016/j.psj.2020.11.048

Bindari, Y. R., & Gerber, P. F. (2022). Centennial Review: Factors affecting the chicken gastrointestinal microbial composition and their association with gut health and productive performance. *Poultry Science*, *101*(1), 101612. https://doi.org/10.1016/j.psj.2021.101612

Bozzi, D., Rasmussen, J. A., Carøe, C., Sveier, H., Nordøy, K., Gilbert, M. T. P., & Limborg, M. T. (2021). Salmon gut microbiota correlates with disease infection status: potential for monitoring health in farmed animals. *Animal Microbiome*, *3*(1), 30. https://doi.org/10.1186/s42523-021-00096-2

Cheng, S.-C., Liu, C.-B., Yao, X.-Q., Hu, J.-Y., Yin, T.-T., Lim, B. K., Chen, W., Wang, G.-D., Zhang, C.-L., Irwin, D. M., Zhang, Z.-G., Zhang, Y.-P., & Yu, L. (2023). Hologenomic insights into mammalian adaptations to myrmecophagy. *National Science Review*, *10*(4), nwac174. https://doi.org/10.1093/nsr/nwac174

Chklovski, A., Parks, D. H., Woodcroft, B. J., & Tyson, G. W. (2023). CheckM2: a rapid, scalable and accurate tool for assessing microbial genome quality using machine learning. *Nature Methods*, *20*(8), 1203–1212. https://doi.org/10.1038/s41592-023-01940-w

Culp, E. J., & Goodman, A. L. (2023). Cross-feeding in the gut microbiome: Ecology and mechanisms. *Cell Host & Microbe*, *31*(4), 485–499. https://doi.org/10.1016/j.chom.2023.03.016

Diaz Caballero, J., Wheatley, R. M., Kapel, N., López-Causapé, C., Van der Schalk, T., Quinn, A., Shaw, L. P., Ogunlana, L., Recanatini, C., Xavier, B. B., Timbermont, L., Kluytmans, J., Ruzin, A., Esser, M., Malhotra-Kumar, S., Oliver, A., & MacLean, R. C. (2023). Mixed strain pathogen populations accelerate the evolution of antibiotic resistance in patients. *Nature Communications*, *14*(1), 4083. https://doi.org/10.1038/s41467-023-39416-2

Diaz Carrasco, J. M., Casanova, N. A., & Fernández Miyakawa, M. E. (2019). Microbiota, Gut Health and Chicken Productivity: What Is the Connection? *Microorganisms*, *7*(10). https://doi.org/10.3390/microorganisms7100374

Fontaine, S. S., & Kohl, K. D. (2023). The microbiome buffers tadpole hosts from heat stress: a hologenomic approach to understand host–microbe interactions under warming. *The Journal of Experimental Biology*. https://journals.biologists.com/jeb/article-abstract/226/1/jeb245191/286604

Gilroy, R., Ravi, A., Getino, M., Pursley, I., Horton, D. L., Alikhan, N.-F., Baker, D., Gharbi, K., Hall, N., Watson, M., Adriaenssens, E. M., Foster-Nyarko, E., Jarju, S., Secka, A., Antonio, M., Oren, A., Chaudhuri, R. R., La Ragione, R., Hildebrand, F., & Pallen, M. J. (2021). Extensive microbial diversity within the chicken gut microbiome revealed by metagenomics and culture. *PeerJ*, *9*(e10941), e10941. https://doi.org/10.7717/peerj.10941

Graw, S., Chappell, K., Washam, C. L., Gies, A., Bird, J., Robeson, M. S., 2nd, & Byrum, S. D. (2021). Multi-omics data integration considerations and study design for biological systems and disease. *Molecular Omics*, *17*(2), 170–185. https://doi.org/10.1039/d0mo00041h

He, Z., Liu, R., Wang, M., Wang, Q., Zheng, J., Ding, J., Wen, J., Fahey, A. G., & Zhao, G. (2023). Combined effect of microbially derived cecal SCFA and host genetics on feed efficiency in broiler chickens. *Microbiome*, *11*(1), 198. https://doi.org/10.1186/s40168-023-01627-6

Huang, P., Zhang, Y., Xiao, K., Jiang, F., Wang, H., Tang, D., Liu, D., Liu, B., Liu, Y., He, X., Liu, H., Liu, X., Qing, Z., Liu, C., Huang, J., Ren, Y., Yun, L., Yin, L., Lin, Q., … Zeng, J. (2018). The chicken gut metagenome and the modulatory effects of plant-derived benzylisoquinoline alkaloids. *Microbiome*, *6*(1), 211. https://doi.org/10.1186/s40168-018-0590-5

Ijaz, U. Z., Sivaloganathan, L., McKenna, A., Richmond, A., Kelly, C., Linton, M., Stratakos, A. C., Lavery, U., Elmi, A., Wren, B. W., Dorrell, N., Corcionivoschi, N., & Gundogdu, O. (2018). Comprehensive Longitudinal Microbiome Analysis of the Chicken Cecum Reveals a Shift From Competitive to Environmental Drivers and a Window of Opportunity for Campylobacter. *Frontiers in Microbiology*, *9*, 2452. https://doi.org/10.3389/fmicb.2018.02452

Jiang, X., Zhang, B., Lan, F., Zhong, C., Jin, J., Li, X., Zhou, Q., Li, J., Yang, N., Wen, C., & Sun, C. (2023). Host genetics and gut microbiota jointly regulate blood biochemical indicators in chickens. *Applied Microbiology and Biotechnology*. https://doi.org/10.1007/s00253-023-12814-8

Karp, P. D., Midford, P. E., Billington, R., Kothari, A., Krummenacker, M., Latendresse, M., Ong, W. K.,

Subhraveti, P., Caspi, R., Fulcher, C., Keseler, I. M., & Paley, S. M. (2021). Pathway Tools version 23.0 update: software for pathway/genome informatics and systems biology. *Briefings in Bioinformatics*, *22*(1), 109–126. https://doi.org/10.1093/bib/bbz104

Koziol, A., Odriozola, I., Leonard, A., Eisenhofer, R., San José, C., Aizpurua, O., & Alberdi, A. (2023). Mammals show distinct functional gut microbiome dynamics to identical series of environmental stressors. *mBio*, e0160623. https://doi.org/10.1128/mbio.01606-23

Mota-Gutierrez, J., Lis, L., Lasagabaster, A., Nafarrate, I., Ferrocino, I., Cocolin, L., & Rantsiou, K. (2022). Campylobacter spp. prevalence and mitigation strategies in the broiler production chain. *Food Microbiology*, *104*, 103998. https://doi.org/10.1016/j.fm.2022.103998

Nyholm, L., Koziol, A., Marcos, S., Botnen, A. B., Aizpurua, O., Gopalakrishnan, S., Limborg, M. T., Gilbert, M. T. P., & Alberdi, A. (2020). Holo-Omics: Integrated Host-Microbiota Multi-omics for Basic and Applied Biological Research. *iScience*, *23*(8), 101414. https://doi.org/10.1016/j.isci.2020.101414

Ocejo, M., Oporto, B., & Hurtado, A. (2019). 16S rRNA amplicon sequencing characterization of caecal microbiome composition of broilers and free-range slow-growing chickens throughout their productive lifespan. *Scientific Reports*, *9*(1), 2506. https://doi.org/10.1038/s41598-019-39323-x

Rasmussen, J. A., Kiilerich, P., Madhun, A. S., Waagbø, R., Lock, E.-J. R., Madsen, L., Gilbert, M. T. P., Kristiansen, K., & Limborg, M. T. (2023). Co-diversification of an intestinal Mycoplasma and its salmonid host. *The ISME Journal*, *17*(5), 682–692. https://doi.org/10.1038/s41396-023-01379-z

Rasmussen, J. A., Villumsen, K. R., Duchêne, D. A., Puetz, L. C., Delmont, T. O., Sveier, H., von Gersdorff Jørgensen, L., Præbel, K., Martin, M. D., Bojesen, A. M., Gilbert, M. T. P., Kristiansen, K., & Limborg, M. T. (2021). Genome-resolved metagenomics suggests a mutualistic relationship between Mycoplasma and salmonid hosts. *Communications Biology*, *4*(1), 1–10. https://doi.org/10.1038/s42003-021-02105-1

Rehman, A., Arif, M., Sajjad, N., Al-Ghadi, M. Q., Alagawany, M., Abd El-Hack, M. E., Alhimaidi, A. R., Elnesr, S. S., Almutairi, B. O., Amran, R. A., Hussein, E. O. S., & Swelum, A. A. (2020). Dietary effect of probiotics and prebiotics on broiler performance, carcass, and immunity. *Poultry Science*, *99*(12), 6946–6953. https://doi.org/10.1016/j.psj.2020.09.043

Rhie, A., McCarthy, S. A., Fedrigo, O., Damas, J., Formenti, G., Koren, S., Uliano-Silva, M., Chow, W., Fungtammasan, A., Kim, J., Lee, C., Ko, B. J., Chaisson, M., Gedman, G. L., Cantin, L. J., Thibaud-Nissen, F., Haggerty, L., Bista, I., Smith, M., … Jarvis, E. D. (2021). Towards complete and error-free genome assemblies of all vertebrate species. *Nature*, *592*(7856), 737–746. https://doi.org/10.1038/s41586-021-03451-0

Rosenberg, E., & Zilber-Rosenberg, I. (2018). The hologenome concept of evolution after 10 years. *Microbiome*, *6*(1), 78. https://doi.org/10.1186/s40168-018-0457-9

Rychlik, I. (2020). Composition and Function of Chicken Gut Microbiota. *Animals*. https://www.mdpi.com/2076-2615/10/1/103

Shaffer, M., Borton, M. A., McGivern, B. B., Zayed, A. A., La Rosa, S. L., Solden, L. M., Liu, P., Narrowe, A. B., Rodríguez-Ramos, J., Bolduc, B., Gazitúa, M. C., Daly, R. A., Smith, G. J., Vik, D. R., Pope, P. B., Sullivan, M. B., Roux, S., & Wrighton, K. C. (2020). DRAM for distilling microbial metabolism to automate the curation of microbiome function. *Nucleic Acids Research*, *48*(16), 8883–8900. https://doi.org/10.1093/nar/gkaa621

Shah, T. M., Patel, J. G., Gohil, T. P., Blake, D. P., & Joshi, C. G. (2019). Host transcriptome and microbiome interaction modulates physiology of full-sibs broilers with divergent feed conversion ratio. *NPJ Biofilms and Microbiomes*, *5*, 24. https://doi.org/10.1038/s41522-019-0096-3

Suez, J., Zmora, N., & Elinav, E. (2020). Probiotics in the next-generation sequencing era. *Gut Microbes*, *11*(1), 77–93. https://doi.org/10.1080/19490976.2019.1586039

Suez, J., Zmora, N., Segal, E., & Elinav, E. (2019). The pros, cons, and many unknowns of probiotics. *Nature Medicine*. https://www.nature.com/articles/s41591-019-0439-x

Tikhonov, G., Opedal, Ø. H., Abrego, N., Lehikoinen, A., de Jonge, M. M. J., Oksanen, J., & Ovaskainen, O. (2020). Joint species distribution modelling with the r-package Hmsc. *Methods in Ecology and Evolution / British Ecological Society*, *11*(3), 442–447. https://doi.org/10.1111/2041-210X.13345

Wang, L., Jacobs, J. P., Lagishetty, V., Yuan, P.-Q., Wu, S. V., Million, M., Reeve, J. R., Jr, Pisegna, J.

R., & Taché, Y. (2017). High-protein diet improves sensitivity to cholecystokinin and shifts the cecal microbiome without altering brain inflammation in diet-induced obesity in rats. *American Journal of Physiology. Regulatory, Integrative and Comparative Physiology*, *313*(4), R473–R486. https://doi.org/10.1152/ajpregu.00105.2017

Wang, X., Tsai, T., Deng, F., Wei, X., Chai, J., Knapp, J., Apple, J., Maxwell, C. V., Lee, J. A., Li, Y., & Zhao, J. (2019). Longitudinal investigation of the swine gut microbiome from birth to market reveals stage and growth performance associated bacteria. *Microbiome*, *7*(1), 109. https://doi.org/10.1186/s40168-019-0721-7

Wu, Y., Pham, Q., Wang, Y., Huang, H., Jiang, X., Li, R. W., Yu, L., Luo, Y., Wang, J., & Wang, T. T. Y. (2023). Red cabbage microgreen modulation of gut microbiota is associated with attenuation of diet-induced obesity risk factors in a mouse model. *Food & Function*, *14*(14), 6654–6664. https://doi.org/10.1039/d3fo01249b

Xue, M.-Y., Sun, H.-Z., Wu, X.-H., Liu, J.-X., & Guan, L. L. (2020). Multi-omics reveals that the rumen microbiome and its metabolome together with the host metabolome contribute to individualized dairy cow performance. *Microbiome*, *8*(1), 64. https://doi.org/10.1186/s40168-020-00819-8

Zhang, Y., Jiang, F., Yang, B., Wang, S., Wang, H., Wang, A., Xu, D., & Fan, W. (2022). Improved microbial genomes and gene catalog of the chicken gut from metagenomic sequencing of high-fidelity long reads. *GigaScience*, *11*. https://doi.org/10.1093/gigascience/giac116

Zhou, Q., Lan, F., Li, X., Yan, W., Sun, C., Li, J., Yang, N., & Wen, C. (2021). The Spatial and Temporal Characterization of Gut Microbiota in Broilers. *Frontiers in Veterinary Science*, *8*, 712226. https://doi.org/10.3389/fvets.2021.712226

Holo-omics: integrated host-microbiota multi-omics for basic and applied biological research

Annex 1

iScience

CellPress
OPEN ACCESS

Perspective

Holo-Omics: Integrated Host-Microbiota Multi-omics for Basic and Applied Biological Research

Lasse Nyholm,[1,*] Adam Koziol,[1] Sofia Marcos,[2] Amanda Bolt Botnen,[1] Ostaizka Aizpurua,[1] Shyam Gopalakrishnan,[1,3] Morten T. Limborg,[1] M.Thomas P. Gilbert,[1,4] and Antton Alberdi[1]

SUMMARY

From ontogenesis to homeostasis, the phenotypes of complex organisms are shaped by the bidirectional interactions between the host organisms and their associated microbiota. Current technology can reveal many such interactions by combining multi-omic data from both hosts and microbes. However, exploring the full extent of these interactions requires careful consideration of study design for the efficient generation and optimal integration of data derived from (meta)genomics, (meta)transcriptomics, (meta)proteomics, and (meta)metabolomics. In this perspective, we introduce the holo-omic approach that incorporates multi-omic data from both host and microbiota domains to untangle the interplay between the two. We revisit the recent literature on biomolecular host-microbe interactions and discuss the implementation and current limitations of the holo-omic approach. We anticipate that the application of this approach can contribute to opening new research avenues and discoveries in biomedicine, biotechnology, agricultural and aquacultural sciences, nature conservation, as well as basic ecological and evolutionary research.

Research conducted over the last decade has fundamentally changed how we perceive the biology and underlying genetic properties of macroorganisms, from looking at individuals as isolated genetic entities to recognizing how they interact with their associated microorganisms in a myriad of biological processes. These microorganisms associated with plants and animals are now acknowledged as relevant—even essential—assets to many basic biological processes, including nutrient acquisition (Falcinelli et al., 2015), immune response (Wu and Wu, 2012), development (Rudman et al., 2019), biomolecule synthesis (Nicholson et al., 2012), and behavior (Liang et al., 2018). This realization has promoted the notion of the **holobiont** (see Box 1 for definitions of this and other terms in bold), a term used to collectively describe the host organism and all its associated microorganisms.

Historically, the phenotypic variation of plants and animals has been attributed to the interplay between **genomic** properties (Koonin et al., 2000) and environmental factors (Schmid, 1992). However, a long history of research on some insects and domestic vertebrates suggested that microorganisms associated with host animals should also be included in the equation. For example, termites have long been known (Leidy, 1881) to require gut microbes to be able to digest their food. In the last decade, researchers have benefited from the rapid development of high-throughput sequencing technology to more intensively explore how the **metagenomic** features of host-associated microorganisms also shape plant and animal phenotypes (Gilbert et al., 2018; Stringlis et al., 2018). These advances have expanded our knowledge on the role of host-microbe interactions in the evolution and ecology of modern-day organisms and how knowledge of such interactions can be beneficial in applied sciences. They basically revealed the termite example to be closer to the norm than the exception. Although individually both genomic and metagenomic approaches have proven useful for understanding many biological processes, each type of study has typically ignored the effect of the other domain and, critically, their interplay. Hence, the knowledge gained through such approaches is, at the very least, incomplete. The recognition of the importance of these host-microbiota interactions has recently opened up new research avenues based on the integrated analysis of coupled genomic and metagenomic data (Limborg et al., 2018), which can be referred to as the research field of **hologenomics** (Figure 1A).

[1]Center for Evolutionary Hologenomics, GLOBE Institute, University of Copenhagen, Copenhagen 1353, Denmark

[2]Department of Genetics, Physical Anthropology and Animal Physiology, University of the Basque Country (UPV/EHU), Leioa 48940, Spain

[3]Department of Health Technology, Section for Bioinformatics, Technical University of Denmark, Kongens Lyngby 2800, Denmark

[4]Norwegian University of Science and Technology, University Museum, Trondheim 7491, Norway

*Correspondence: lasse.nyholm@sund.ku.dk

https://doi.org/10.1016/j.isci.2020.101414

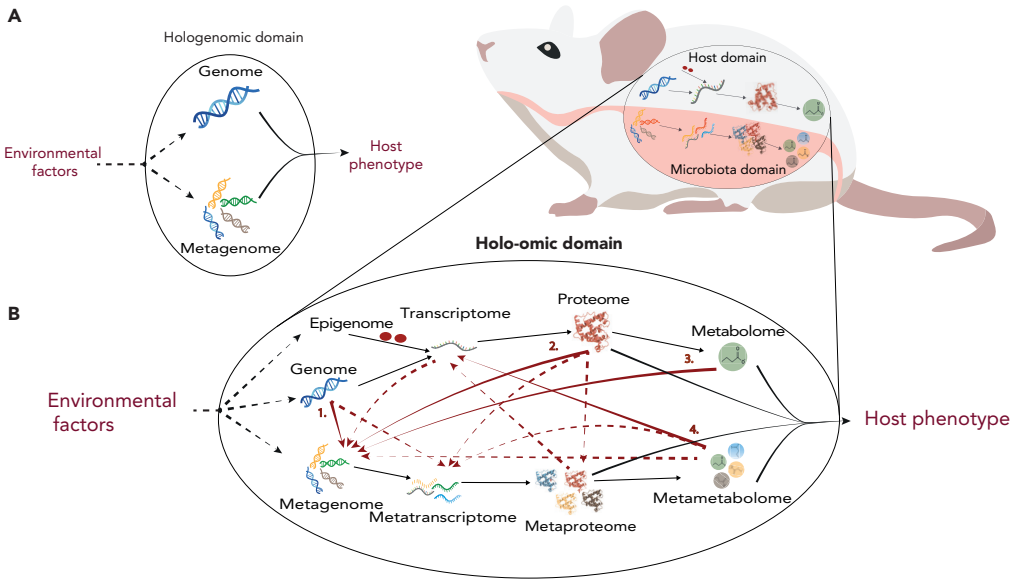


Figure 1. From Hologenomic to Holo-Omic

(A) Simplified visualization of the hologenomic domain.

(B) Host-microbiota interactions within the holo-omic domain here exemplified by zooming in on the luminal surface of the host intestine. Red arrows indicate host-microbiota holo-omic interactions. Solid red arrows indicate interactions supported in the primary literature (numbers refer to the publications listed in Table 1), whereas dashed red arrows indicate potential holo-omic interactions that, to the best of our knowledge, have not yet been documented. Solid black arrows indicate omic levels influencing host phenotype, and dashed black arrows indicate omic levels influenced by environmental factors.

Efforts to study the effects of host and microbial genes and their consequences have become embedded in layer upon layer of jargon. Because the concepts being discussed are new, some of these new terms are necessary, so as to have common reference points. But they only serve as effective reference points if they are well defined. Here we propose that hologenomics (the combined genetic content of the host and the **microbiota**) can be expanded to the **holo-omic** level by the incorporation of data from multiple **omic** levels from both host and microbiota domains (Limborg et al., 2018) (Figure 1B). This approach is inspired by elements originating from **systems biology** (e.g., metagenomics systems biology [Greenblum et al., 2012] and the use of **multi-omic** data integration [Bersanelli et al., 2016; Heintz-Buschart et al., 2016; Liu et al., 2020]). However, multi-omics implies omic data from only one domain, whereas holo-omics is defined by the incorporation of both host and microbial data. In theory, implementing a holo-omic approach would allow researchers to reveal a range of biomolecular interactions responsible for shaping the phenotype of complex organisms, using a variety of molecular tools, and would ultimately provide great potential for application across many different fields of research. The holo-omic toolbox requires both methodological and analytical tools. Within the methodological tools are the nucleic acid sequencing and mass spectrometry technologies that enable tracking the biomolecular pathways linking host and microbial genomic sequences with biomolecular phenotypes by generating (**meta**)**transcriptomes**, (**meta**)**proteomes**, and (**meta**)**metabolomes**. The same technologies also enable **epigenomic** and **exposomic** profiling, which can further contribute to disentangling the biochemical associations between host-microbiota-environment interactions and their effect on host phenotypes (Kumar et al., 2014; Rogler and Vavricka, 2015). The analytical tools required to extract useful information from the enormous amount of highly complex data generated by current high-throughput technologies are still limited. Association studies—identifying correlations between genetic variants and phenotypes—have been used to detect the genetic contributions to complex phenotypes (Welter et al., 2014). This approach has been extended to metabolomic profiles (Luo, 2015) and metagenomic variants (Blekhman et al., 2015; Qin et al., 2012), but methods that jointly leverage the multiple omic levels to infer the causal pathways between genomic processes and phenotypes are still scarce.

In this context, the technology to generate large amounts of data to be used in a holo-omic context is already available, but the analytical tools to reveal and identify host-microbiota interactions are still limited.

As a consequence, only a handful of research groups worldwide have been able to effectively implement the holo-omic approach. To contribute to the development of this new field, in this perspective we first revisit the available evidence for the biological importance of host-microbiota interactions. Second, we present how the holo-omic toolbox can be used to study host-microbiota interactions at varying levels of complexity to guide researchers through applying the holo-omic approach. Third, we showcase the potential provided by the holo-omic approach to host-microbiota interactions in both basic and applied biological sciences and finally we identify the limiting factors that currently prevent the widespread implementation of the holo-omic approach and discuss possible solutions to overcome them.

HOST-MICROBIOTA INTERACTIONS IN LIGHT OF HOLO-OMICS

The holo-omic approach to host-microbiota interactions relies on three major assumptions: (1) host-associated microorganisms interact not only with each other but also with their host (Bredon et al., 2018; Fischer et al., 2017; Stringlis et al., 2018; Vaishnava et al., 2011); (2) these interactions affect, either positively or negatively, central biological processes of hosts and microorganisms (Wu and Wu, 2012); and (3) the interplay can be traced using biomolecular tools (Bansal et al., 2010; Bredon et al., 2018; Kelly et al., 2015; Virtue et al., 2019).

It has been estimated that the number of host-associated microbial cells and genes greatly outnumber that of their hosts' (Gilbert et al., 2018; Stringlis et al., 2018). These microorganisms do not passively inhabit the surfaces of their hosts but instead continuously interact with each other and their hosts through a myriad of complex feedback processes (e.g., Falcinelli et al., 2015; Kelly et al., 2015; Stringlis et al., 2018). For example, host genomic features are co-responsible for shaping the microbiota composition (Suzuki et al., 2019) through the differential biosynthesis of antibacterial peptides (Carvalho et al., 2012), differential composition of intestinal mucosa (Vaishnava et al., 2011), or differential release of nutrients (Reese et al., 2018). Gene expression interdependencies are also common between hosts and microorganisms. For instance, administration of *Lactobacillus rhamnosus* increases the uptake of fatty acids in zebrafish by down-regulating the transcription of host genes related to cholesterol and triglycerides metabolism (Falcinelli et al., 2015). Similarly, the metabolism of microbiota-derived butyrate in epithelial cells stabilizes the function of the hypoxia-inducible transcription factor, which regulates the expression of a number of genes related to host immunity (Kelly et al., 2015). Further examples of similar causal relationships between different omic levels from hosts and microorganisms are compiled in Table 1, and undoubtedly, many more will be revealed in the years to come.

Host-microbiota interactions can have both positive and/or negative influences on host fitness. This has, for instance, been illustrated in studies on relatively well-defined bacteria-insect interactions. Such studies have revealed that the nature of these influences are often context dependent (Fry et al., 2004; Werren et al., 2008) and that these interactions can have both negative and positive influences on evolutionary adaptations (Bennett and Moran, 2015). For other, less studied and more complex host-microbiota consortiums, it has been found that positive interactions can, for instance, lead to increases in nutrient uptake through the degradation of recalcitrant organic compounds (Bredon et al., 2018), increase survival through modulating the resistance toward infectious diseases (Rosshart et al., 2017), or lengthen lifespan through modulating the aging process (Kim and Jazwinski, 2018). On the contrary, host-microbiota interactions can also have negative outcomes for the host. This is most obvious in the context of pathogens that cause infectious diseases (Fei and Zhao, 2013), but it is also apparent, for example, in the context of **dysbiosis** associated with chronic diseases such as inflammatory bowel syndrome (IBS) (Imhann et al., 2018). The origin of such microbial imbalances remains a cause of contention due to difficulty determining whether a disrupted microbiota is the cause or effect of a given illness (Walker, 2017) and it seems likely that such dysbioses have many different causes in different host species, genotypes, and contexts. This debate raises the question of how to determine what constitutes a healthy microbiome, a question that is difficult to answer, especially for wild organisms, owing to inter-population variation caused by environmental and genetic factors as well as the lack of functional annotation of many microbial genes (Lloyd-Price et al., 2016).

All these examples highlight the relevance of acknowledging and understanding the biomolecular interactions occurring between different omic levels of hosts and microorganisms. In the following section we will describe how holo-omics can be implemented by addressing different methodological, experimental, and analytical approaches.

Omic Levels	Organism	Major Findings	Reference	Arrow in Figure 1
Genome, microbial 16S	Mouse	20 host genes are associated with microbiome composition	Suzuki et al. (2019)	1
Genome, microbial 16S	Human	Genetic disposition for inflammatory bowel disease is associated with a reduction in abundance of the genus *Roseburia* in the gut microbiome	Imhann et al. (2018)	1
Transcriptome, metagenome	Pill-bug (*Armadillidium vulgare*)	Potential collaboration between microbiota and pill-bug in degrading lignocellulose	Bredon et al. (2018)	–
Proteome, microbial 16S	Mouse	Lack of the TLR5 protein increases Proteobacteria and decreases Bacteroidetes in microbiome and promotes gut inflammation	Carvalho et al. (2012)	2
Metabolome, metagenome	Thale cress (*Arabidopsis thaliana*)	Beneficial rhizobacteria induce excretion of the metabolite scopoletin that stimulates iron uptake and suppresses soil-borne pathogens	Stringlis et al. (2018)	3
Metametabolome, transcriptome	Human epithelial cells	Metabolism of microbiota-derived butyrate stabilizes the HIF transcription factor in human epithelial cells	Kelly et al. (2015)	4
Metametabolome, transcriptome	Human epithelial cells	The presence of microbiota-derived indole stimulates the expression of host genes connecting to the formation of tight junctions with a resulting higher pathogen resistance	Bansal et al. (2010)	4
Metametabolome, transcriptome	Mouse	Microbiota-derived indole controls expression of host *miR-181* expression that regulates adiposity and insulin sensitivity	Virtue et al. (2019)	4

Table 1. Examples of Holo-Omic Studies in the Current Litterature

Examples of studies considering different omic levels from hosts and associated microorganisms at different levels of resolution. When evidence of host-microbiota interactions are available numbers link the table to the corresponding interaction in Figure 1.

IMPLEMENTING THE HOLO-OMIC APPROACH

The holo-omic approach can be implemented by using a range of different methodological tools in diverse experimental setups that might require a variety of analytical and statistical approaches (Figure 2). Regarding data generation, most studies linking the host and the microbiota domains have relied on targeted approaches (e.g., **amplicon sequencing**, **targeted RNA-sequencing**, and **western blotting**) to characterize the microbial domain. However, untargeted approaches (e.g., **shotgun DNA sequencing** and **shotgun proteomics**), which non-selectively provide a snapshot of nucleotides, proteins, and metabolites

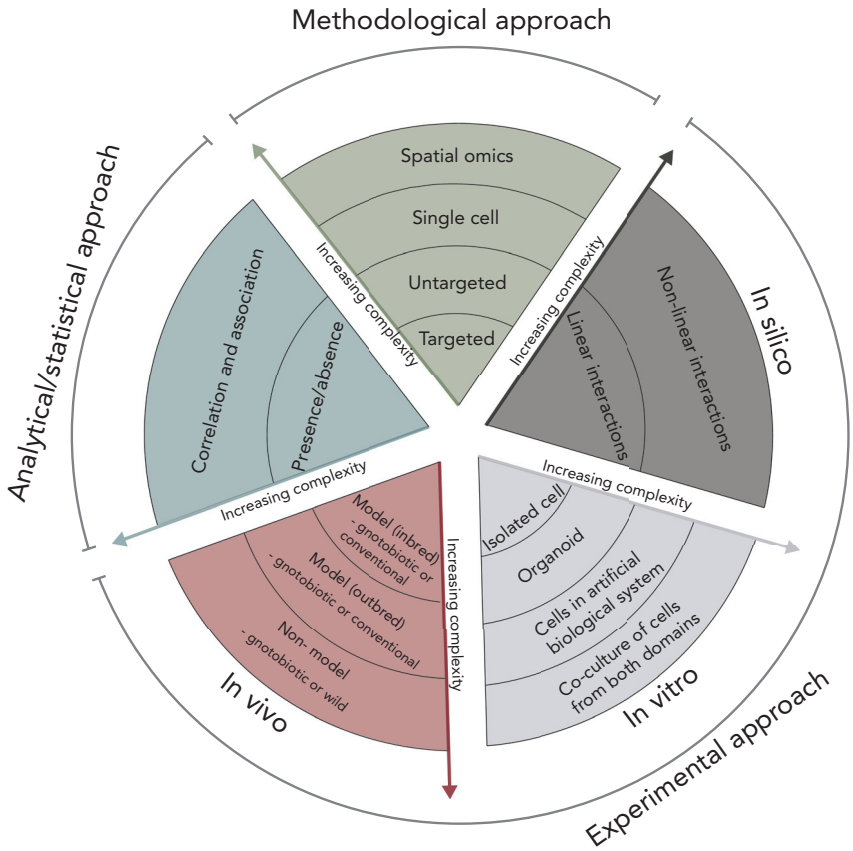


Figure 2. Overview of Different Approaches in Holo-Omics and Their Influence on the Level of Complexity
Approaches are divided into methodological, experimental, and statistical. Arrows indicate the level of complexity relative to each segment of the figure.

present in a sample, are progressively complementing or replacing targeted approaches. For instance, coupled untargeted host/microbe data from shotgun sequencing offers advantages over targeted approaches, such as the construction of metagenome assembled genomes (**MAGs**) from metagenomic data (Almeida et al., 2019) and the generation of individual genomic profiles (Blekhman et al., 2015). Furthermore, the (meta)genomic data needed for implementing the hologenomic approach to host-microbe interactions are often derived from samples containing DNA from both domains (Blekhman et al., 2015). At the same time, the ever-decreasing costs of sequencing coupled with increases in computational efficiency are expected to boost this trend toward shotgun sequencing (Quince et al., 2017). In recent years, **single cell sequencing** has expanded our ability to link specific genetic properties to single cells (Xu and Zhou, 2018), which could be used to study the interactions between *in vitro* cultures of eukaryotic and prokaryotic cells in great detail. In addition to this, the use of **spatial metagenomics** is capable of resolving the geographical distribution of individual microbes within a community (Sheth et al., 2019), and we foresee that this method will prove valuable in the future of holo-omics to highlight the effect of relative spatial orientation between host and microbial cells. In 10 years, incorporating a range of approaches in a single study with massive replication will probably be trivial from a cost perspective. In this context, the burden (and key challenge) is combining theoretical insight and analytical clarity.

If the metagenomic data include some proportion of host DNA, often considered as host contamination, *in silico* approaches can be used to also profile the host genotype and screen for potential associations between genetic markers and microbial traits (Blekhman et al., 2015). *In vitro* approaches, in which the host environment is reproduced in simpler physical models such as miniature organs grown from stem cells (i.e., epithelial organoids), might provide the required resolution when trying to uncover the interaction between well-defined binary interactions, e.g., the effect of microbiota-produced butyrate on host

transcriptomics in epithelial cells (Kelly et al., 2015). *In vivo* approaches using single-symbiont or gnotobiotic organisms are chosen when trying to uncover the complete effect of a symbiont, beyond the effect of a single molecule (Koch and McFall-Ngai, 2019), whereas wild organisms might provide the most direct evidence about the effect of host-microbiota interactions in natural processes (Alberdi et al., 2016).

The implementation of a full holo-omic approach with multiple omic levels from both hosts and microorganisms begins with the generation of high-dimensional data. Depending on the aims of the study, data from each sample in the study can encompass measurements on genes, genomes, transcripts, proteins, or metabolites. Specifically, the microbiota can be characterized by hundreds of MAGs, thousands of gene orthologs, or millions of genes. The number of independent measurements and the high dimensionality of the resulting data pose significant challenges to traditional statistical approaches, such as correlation-based methods and linear models. One possible approach to reducing the complexity of the problem is to use some form of dimensionality reduction, such as clustering MAGs by taxonomy or ecological guilds (Zhao et al., 2018), or grouping genes by their functional properties (Qin et al., 2010). Although such dimensionality reduction simplifies the analyses and reduces computational complexity, it can lead to loss of biologically relevant information (Wang et al., 2019).

Pioneering studies in hologenomics have relied on association analyses to identify correlations between hosts and related microorganisms. Genome-wide association studies (**GWASs**) have linked specific loci in the host genome to the presence of pathogenic or beneficial microbes (Blekhman et al., 2015; Imhann et al., 2018). Similar approaches have been used in the study of epigenomes (Wan et al., 2016), metabolomes (Sekula et al., 2016), and proteomes (Okada et al., 2016). GWASs served as inspiration for metagenome-wide association studies (**MGWASs**) linking specific genes in the metagenome to phenotypic traits of interest in the host (Qin et al., 2012). So far, most methods used to integrate multi-omic data from both host and microbiota domains have relied on standard statistical methods, such as general linear models and linear mixed models in GWASs and MGWASs (Blekhman et al., 2015; Imhann et al., 2018; Qin et al., 2012). These methods are often hampered by the high-dimensional nature of the metagenomic data, highlighting the need for specialized methods to deal with highly complex holo-omic data (Wang et al., 2019).

Aiming to advance holo-omic research beyond association analyses, we recently introduced a methodological framework proposing a two-step approach to reveal the mechanisms underlying phenotypic variance modulated by the interactions between the host and related microorganisms (Limborg et al., 2018): an initial association phase based on GWAS and MGWAS analysis, followed by an interaction phase to identify bidirectional interactions at different omic levels. The initial association phase can identify variants (SNPs) within the genome and metagenome (e.g., amplicon sequence variants, operational taxonomic units, MAGs, or genes) associated with certain host phenotypes. In the following interaction phase, the effects of the associated GWAS variants on other omic domains are explored, thus identifying the important aspects of the molecular machinery that lead from genotypic variation to phenotypic variation. Although the two-step approach allows us to dig deeper into the interactions between the different omic domains that affect the phenotype, we are still limited by the power of the GWAS performed in the first step in identifying causal variants. In essence, the first step acts as a dimensionality reduction step, reducing the space of interactions that need to be interrogated. The problem of integrated inference by leveraging different omics data is a difficult one, and the development of computational methods in this field have been hindered by the inherent complexity of holo-omic data and the biological process underlying them. The current state of the art in integrating different omics dataset relies either on network-based methods (Langfelder and Horvath, 2008), regularized regression-based methods (Rohart et al., 2017), or other niche tools (Hernandez-Ferrer et al., 2017). However, none of these methods were designed for the analysis of metagenomic, metatranscriptomic, or metametabolomic data.

The methodological, experimental, and analytical approaches mentioned above are challenged by the high costs of data generation and the complexity of downstream analyses. This requires that researchers consider at least three fundamental questions about the system under study before taking on a holo-omic study (Box 2).

APPLYING THE HOLO-OMIC APPROACH ACROSS BIOLOGICAL SCIENCES

The holo-omic approach outlined above can be implemented in many basic and applied biological research fields to address relevant scientific questions concerning host-microbiota interactions. In the

Box 1. Glossary

Amplicon sequencing:PCR amplification-based targeted sequencing of a specific genetic region.
Dysbiosis: Any change to the components of resident commensal microbial communities relative to the community found in healthy individuals.
Epigenome: The heritable alteration of DNA or proteins associated with DNA that changes gene expression levels in a cell or tissue without modifying the sequence of DNA.
Epigenotype:The pattern of epigenetic modification (alteration of DNA or proteins that changes gene expression) in a cell or tissue.
Exposome: Every exposure that an organism is subjected to throughout its lifetime.
Genome: The complete set of genetic material of an organism.
Genome-wide association study(GWAS): An examination of a genome-wide set of genetic variations associated with a trait of interest.
Holobiont:A host organism and its associated microorganisms.
Hologenome: The combined genetic content of the host and its associated microbiota.
Holo-omics: The analysis of multiple omic levels from both host and associated microbiota domains.
Hologenome theory of evolution: The theory that posits host, symbionts, and their associated hologenome, acting in consortium, function as a biological entity and as a level of selection in evolution.
Metagenome-assembled genome (MAG): Genome assembled from shotgun sequencing data generated from the entire genetic content present in a given environment.
Metabolome: The entire pool of metabolites present in an organism.
Metagenome: The entire genetic content present in a given environment.
Metametabolome: The entire pool of metabolites present in an environmental sample.
Metaproteome: The complete set of proteins/peptides present in an environmental sample.
Metatranscriptome: The entire pool of mRNA in an environmental sample.
Metagenome-wide association study (MGWAS): An examination of a metagenome-wide set of genetic variations associated with a trait of interest.
Microbiome: The sum of genetic material in a microbial community.
Microbiota: The ecological community of microorganisms.
Multi-omics: The analysis of multiple types of omic data (e.g., metagenome and metaproteome).
Omic: Term used to describe any level of multi-omics (i.e., (meta)genomics, epigenomics, (meta)transcriptomics, (meta)proteomics, and (meta)metabolomics).
Proteome: The entire pool of proteins present in an organism.
Shotgun DNA sequencing: The non-targeted sequencing of the entire genetic content of a sample.
Shotgun proteomics: The direct analysis of complex protein mixtures to generate global profiles of proteins within a sample.
Single cell sequencing: Sequencing of the nucleic acid content within a single cell.
Spatial metagenomics: Characterization of the spatial orientation of microbes in their environment by fixation in a matrix followed by either amplicon sequencing or shotgun sequencing.
Systems biology:A holistic approach, often employing quantitative modeling, to study biological systems that cannot be reduced to the sum of the systems individual parts.
Targeted RNA sequencing: Sequencing of specific RNA molecules using probes complementing the transcript of interest.
Transcriptome: The sum of RNA transcripts produced by a single organism.
Western blotting: Separation and identification of proteins in a gel matrix using antibodies.

following, we showcase and discuss the application of the holo-omic approach in diverse fields of biological sciences. For some fields, we include boxes containing case examples to better illustrate its potential implementation.

Agricultural and Aquacultural Sciences

The holo-omic approach could prove a meaningful tool in developing animal and plant production as microorganisms are increasingly considered essential assets to improve efficiency and sustainability (Małyska et al., 2019). Among other strategies, animal feed and feed additives are used to modulate animal gut microbiota and improve host growth and health. More sustainable feed formulas are being developed such as the use of seaweed to decrease dairy cows methane emissions (Machado et al., 2014). It has also been suggested that piscivorous fish can be fed with a plant-based diet in aquaculture systems to replace fish meal (Gatlin et al., 2007). On the other hand, feed additives, such as probiotics, prebiotics, and synbiotics, are extensively used in animal breeding owing to many attributed benefits including protection against

pathogens, stimulation of the immunological response, and increment of production capacity, but yet, little is known about their specific mode of action (Markowiak and Śliżewska, 2018). For instance, positive effects were reported on the use of probiotics to control diarrhea syndrome in post-weaning piglets (Kyriakis et al., 1999) and have been found to result in decreased mortality in rainbow trout (Irianto and Austin, 2002). The implementation of the holo-omic approach can help us unveil how feed, microbiota, and the host interact in the intestinal environment, which could prove essential for optimizing the production of host organisms and improving management practices (Box 3). A similar initiative implemented for plants could aim at enhancing adaptation and response to rapid climate change.

Biotechnology

The holo-omic approach could also contribute to developing and optimizing biotechnological solutions. For instance, it could be used to better understand host-microbiota systems capable of enzymatically degrading complex polysaccharides (Ni and Tokuda, 2013) in the search for novel sustainable ways of transforming organic waste compounds into industrially relevant biomolecules and biofuels. Many wood-feeding organisms are capable of partially digesting lignocellulose into glucose, but to complete the

Box 2. Three Main Questions that Researchers Need to Consider to Maximize the Outcome of a Holo-Omic Study

(1) Are host-microbiota interactions relevant in the system under study?

Researchers must assess whether host-microbiota interactions are relevant for understanding the system under study. The impact of microorganisms associated with complex hosts is now regarded as almost universal (Barko et al., 2018), but the effect sizes can vary from low (Kong et al., 2019) to high (Rosshart et al., 2017) values. Hence, an initial screening of the variability of hosts' phenotypic traits and microbial communities associated with them is recommendable to elucidate potential correlations. This could be done using a cost-effective targeted gene sequencing approach to later study the system in more detail using non-targeted approaches.

(2) Is it meaningful to implement a holo-omic approach?

It is necessary to evaluate whether the implementation of a holo-omic approach is reasonable given the properties of the biological system and its environment. Holo-omics relies on the premise that genomic and metagenomic differences across individuals, treatments, populations, or species affect biological processes and phenotypic outcomes. Thus, the existence of genomic or metagenomic variation in the system is essential. It is also necessary to bear in mind that the capacity to recover genomic and metagenomic signatures is largely affected by environmental variables (Figure 3). The background noise introduced by these variables contains information on how the environment influences the dependent variables (Figure 3), but as they are often difficult to measure or control in non-laboratory settings they will often complicate signal recovery. Factors extrinsic to the host (diet, temperature, humidity, etc.) are known to affect both the composition of the microbiota and the expression of its genes (Cernava et al., 2019; David et al., 2014; Moran and Yun, 2015). The level and structure of (meta)genomic and environmental variation will therefore dictate the biological meaning and design of any holo-omic study (Figure 3).

Assessing the economic and technical feasibility of the study is also paramount. This includes acknowledging the genome size of the host, as the genome of some species can be magnitudes larger than others, e.g., amphibians (Nowoshilow et al., 2018) versus birds (Zhang et al., 2014), or questioning whether optimal sample preservation conditions can be ensured, especially critical for (meta)transcriptomics (Ferreira et al., 2018). Assessing the biomolecular properties of the samples (e.g., host:microbiota DNA/RNA ratios) is a relevant preliminary step that aids in outlining an optimal study design (Human Microbiome Project Consortium, 2012).

(3) Which omic levels are relevant and how to maximize the amount of useful data derived from them?

Lastly, researchers should identify which omic levels are the most relevant, both by considering the biological and technical features of the experimental system and their relevance to the research questions. The omic levels selected for analysis will largely determine the number of samples to include (Ching et al., 2014; Hong and Park, 2012), where and how to collect the samples (e.g., which part of the intestinal tract (Kokou et al., 2019), preservation and storage conditions (Ferreira et al., 2018; Hickl et al., 2019), sequencing depth, or how to maximize the amount of biological information coming from them (Quince et al., 2017). The ability of the downstream statistical analyses is dependent not just on these factors but also on the genetic architecture of the phenotype being studied. Prior knowledge of the functional basis of the phenotype can be used to markedly improve the experimental design and improve the power of the statistical analyses (Kichaev et al., 2019).

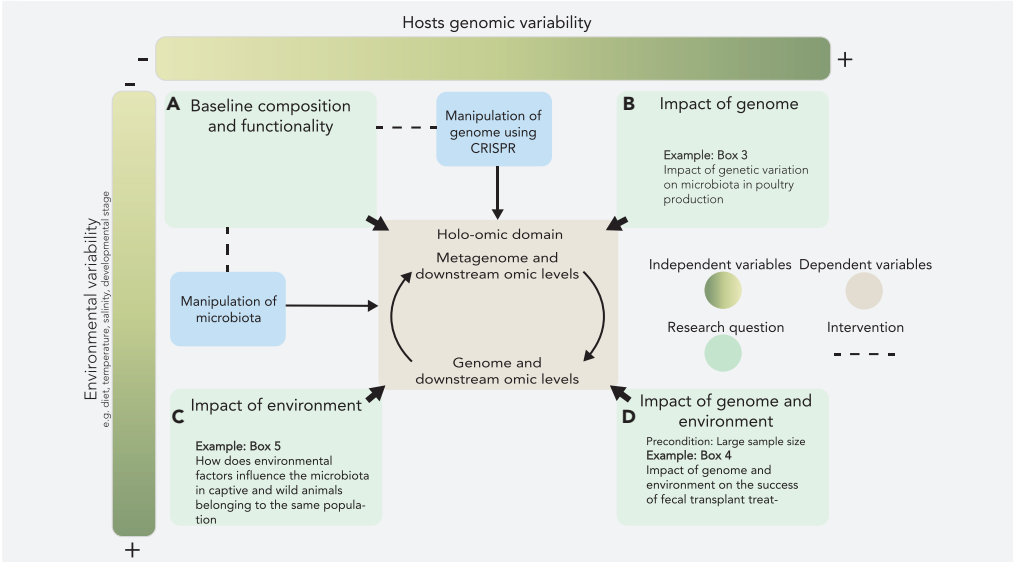


Figure 3. Overview of Different Variables that Will Impact Holo-Omic Studies

In this conceptualization, two independent variables, the environment and the host genome, affect dependent variables (center), the metagenome, and downstream omic levels and their interactions with the host genome and derived omic levels. Different combinations enable implementing different types of experimental approaches.
(A) When both genetic background and environment are constant (e.g., laboratory conditions) the underlying composition and functionality of the microbiota as well as the underlying interaction with the host domain can be determined. These conditions allow researchers to manipulate microbiota composition and functionality and to manipulate the host genome (e.g., using CRISPR-Cas9 genome editing technology).
(B) When the genetic background is variable and the environment is relatively consistent, the impact of genetic variants on downstream omic levels can be isolated.
(C) When the genetic background is similar and the environment is variable, the impact of environmental factors on the different omic levels can be studied.
(D) When both genetic background and environment are variable, the high level of variability will complicate the isolation of factors responsible for modifying the omic levels. Increasing sample size can mitigate this problem.

degradation they need the complementary enzymes produced by their resident microbes (Bredon et al., 2018). Similarly, several studies based on metagenomics and metaproteomics in termites have shown that the microbiota is responsible for producing some of the most complex enzymes involved in the degradation of lignocellulose (Ni and Tokuda, 2013). Most of such complex biochemical reactions occur under anaerobic environments; hence, setting up appropriate bioreactors tends to be a complex process. The holo-omic approach can assist in determining specific bioreactor conditions by ascertaining the enzymatic and metabolic contribution of microorganisms and animal hosts, thus facilitating the replication of the optimal chemical conditions that mimic the hosts' gut environment (Gutleben et al., 2018).

Biomedical Research

Incorporating the holo-omics approach to biomedical research offers an exciting new avenue toward better treatments of many modern human diseases. Most people in industrialized societies exhibit depauperate gut microbiotas (Gupta et al., 2017), which is often held co-responsible for the concomitant explosion in the rate of autoimmune diseases (Bach, 2002), all diseases that have been associated with a dysbiotic microbiome in patients including IBS (Imhann et al., 2018), diabetes, or colorectal cancer (Feng et al., 2015). Although we rarely know whether such a dysbiotic microbiome (Walker, 2017) is the cause or an effect of a disease trait, it is now clear that the field of holo-omics provides an attractive approach to better understand how such changes in host-microbiome interactions occur and potentially how they can be reverted to healthy states. Better understanding of how human genotypes and the exposome of an individual affect the interactions between patients and associated microorganisms would enable advances toward more

Box 3. Implementing the Holo-Omic Approach in Poultry Farming

Chickens are an important source of high-quality protein for a large proportion of the human population. The gut microbiota of broilers (chicken bred for meat production) is highly variable since they are slaughtered before reaching an age in which the microbial community dynamics stabilize (Rychlik, 2020). Although the administration of probiotics and prebiotics to modulate the gut microbiota is becoming increasingly popular, results are still inconclusive and further research is needed (Ducatelle et al., 2015).

(1) Are host-microbiota interactions relevant in the system under study?

The controlled environmental conditions in intensive poultry production systems, which use the same feeding strategy and environment for all individuals, indicate that the likely reason behind variation in the chicken performance and their gut microbiota when administered pre- and probiotics might be explained by microbial founding effects and microgenomic variation of broilers (Box 2. Figure 3B: Impact of Genome) not only across but also within breeds. An initial examination of the genotypes along with a targeted screening of the microbiota of each individual in the broiler population can allow researchers to discover any potential association between the two domains using GWASs and MGWASs with a particular focus on pre- and probiotics-related phenotypic responses (e.g., inflammatory markers, stress response molecules).

(2) Is it meaningful to implement a holo-omic approach?

Pre- and probiotics interact with native gut microorganisms as well as with the host. The gut microbiota of broilers is relatively simple because of the closed environment where the broilers are reared. The genetic diversity of conventionally bred broilers is low, yet even small interindividual differences can be crucial and might have wide implications on the response to pre- and probiotics. These system properties allow the successful application of the holo-omic framework for obtaining relevant microbe-microbe and microbe-host interactions, which can help researchers optimize feed additives design, production, and administration, thereby preventing production inefficiency driven by gut dysbiosis.

(3) Which omic levels are relevant and how to maximize the amount of useful data derived from them?

If associations are detected between the (meta)genome and host phenotypic traits, the study of transcriptome, metatranscriptome, and metabolome can unveil the nature of microbe-microbe and host-microbe interactions and how they affect the host. Detecting molecular pathways that are activated or deactivated in the presence of pre- and probiotics can enable researchers to identify production-related phenotypic changes in the host.

accurate personalized medicine (Ginsburg and McCarthy, 2001). A holo-omics-based personalized medicine would recognize not only the genetic and exposomic features of patients but also the associated microbiota (Box 4).

Ecology and Evolution

Implementation of a holo-omic approach holds the potential to address many basic questions regarding the ecology and evolution of species. Most pertinent of these are in regards to the holobiont and testing specific hypotheses derived from the **hologenome theory of evolution** (Rosenberg et al., 2007). For example, how does selection occur on the holobiont and what mechanisms underpin the cross talk between the host and microbiota axes. One potential application is to measure the impact of microorganisms in vertebrate adaptation and improve predictions from anthropogenic disturbances, such as climate change and habitat destruction, on species distributions. It has been proposed that metagenomes could confer enhanced adaptive capacity to their hosts (Alberdi et al., 2016), potentially enabling rapid adaptation to changing environmental conditions (Fontaine and Kohl, 2020). The adaptive capacity of hosts and their associated microbiota through linking specific host genotypes with metagenomes has been demonstrated with regard to toxicity resilience (Macke et al., 2017), heat tolerance (Moran and Yun, 2015), drought and desiccation (Cernava et al., 2019), disease resistance (Rosshart et al., 2017), and nutrient acquisition (Falcinelli et al., 2015). Through characterizing host-microbiota pathways it is possible to catalog these interactions and begin to assess evolutionary adaptations within the metagenome. This could enable metagenomic—rather than only genomic (e.g., Razgour et al. (2019))—adaptations to be considered when predicting species range shifts owing to climate change and potentially improves the predictive capacity of species distributions. Likewise, such an approach could also be adopted to improve predictions of the adaptive capacity relevant to modeling invasive species (Fontaine and Kohl, 2020), enabling better estimates of invasion trajectories and ecological impact forecasts.

Box 4. Holo-Omic Approach to Fecal Transplant Treatments

The use of fecal transplants, i.e., transferring fecal material from a healthy donor to a patient with a gastrointestinal disorder, is now becoming a promising treatment for multiple gastrointestinal disorders (Mcilroy et al., 2019). Although such treatments have shown some success, outcomes often vary among patients despite receiving the same treatment (Sbahi and Di Palma, 2016). Therefore, we hypothesize that the holo-omics approach can be applied to improve the success rate of such treatments by matching the genotype between fecal donors and recipients similar to procedures for organ transplants.

(1) Are host-microbiota interactions relevant in the system under study?

The success of fecal transplants relies on the capability of beneficial microbes from the donor fecal sample being able to colonize and establish themselves in the gut environment of the recipient. One can hypothesize that the probability of success relates to differences among patient gut "environments" that depends not only on the existing microorganism community but also on the genotype or **epigenotype** of the human host (Box 2. Figure 3D: Impact of Genome and Environment).

(2) Is it meaningful to implement a holo-omic approach?

The information gained from a holo-omic approach will ultimately lead to more efficient treatments by, for example, optimizing the biological match between a fecal donor and recipient. For example, a holo-omics analysis in a controlled cohort can reveal concrete genotypes of a host that are associated with the gut microbiota composition. Then, once we have accumulated knowledge about specific candidate genes directly associated with composition and function of gut microbiota, we can screen these genes to optimize the genetic match between donor and recipient, thereby improving the odds that the recipient is likely to adopt the healthy microbes from the donor and thereby counteract the negative effect from microbes such as *Clostridioides difficile* (Gough et al., 2011).

(3) Which omic levels are relevant and how to maximize the amount of useful data derived from them?

A holo-omic approach to identify the factors underlying the differential success of fecal transplants could include the patients' genomic and epigenomic features coupled with transcriptomic, metagenomic, metatranscriptomic, and metabolomic variation before and after a fecal transplant. Associating these features with the success of the treatment, and with each other, would shed light on the functional changes introduced by the transplant, which would enable identifying the factors leading to a success or failure of the treatment.

Species Conservation

Holo-omics could also be relevant for developing optimal active conservation actions, such as captivity breeding and animal translocations (Box 5). As captive conditions differ extensively from those experienced in the wild, many species kept in captivity diverge in their microbiota compositions compared with their wild counterparts (McKenzie et al., 2017). This could have implications for attempts to translocate species (i.e., introduction, re-introduction, and re-stocking) as the functionality of the microbiota might be compromised thereby diminishing the chance of successful translocation (Bahrndorff et al., 2016). Microbiota composition and functionality varies between local environments, and identifying the local variants can impact conservation effort success. Although conservationists have traditionally focused on a species genetic traits (Allendorf et al., 2010), the holo-omic approach posit to match this information with information on microbiota composition and functionality, to avoid mixing populations with different hologenomic adaptations to a given environment. Matching captive individuals with a "wild microbiota" prior to their release and monitoring their fitness and associated temporal changes of the microbiota in the wild could reveal the efficiency of the holo-omic approach in the field of conservation.

CURRENT LIMITATIONS AND FUTURE PERSPECTIVES

Although holo-omics represent a valuable tool for many fields, its implementation is still hampered by economic, technical and biological limitations. A main economic hurdle is the high cost of shotgun sequencing. Targeted sequencing or DNA microarrays approaches can be cost-effective alternatives for characterizing (meta)genomes in some cases, although shallow shotgun sequencing can in some instances recover higher taxonomic resolution at the same cost, while also providing direct inference about functionality (Hillmann et al., 2018). Targeted approaches might enable researchers to establish correlation between the presence of specific microorganisms and genetic or phenotypic traits of the host, but to infer causation the use of shotgun sequencing will often be necessary to provide whole genome resolution. Alternatively, a cost-effective approach, mostly useful when the microbial diversity is limited, is to combine targeted amplicon sequencing with deep shotgun sequencing on a subset of samples in a dataset (Lesker

Box 5. Implementing the Holo-Omic Approach in Conservation Biology

In winter, the Western capercaillie (*Tetrao urogallus*) feeds almost exclusively on conifer needles rich in resin and phenol and low in nutrients (Bryant, 1980). It has been proposed that the microbiota might be of major importance in aiding the metabolism of these hard-to-digest compounds (Wienemann et al., 2011). Failure of translocated captive-bred individuals to survive in the wild is suspected to be a consequence of the lack of specific microbes capable of digesting the toxic compounds in the diet (Wienemann et al., 2011).

(1) Are host-microbiota interactions relevant in the system under study?

The highly specialized diet with many hard-to-digest components of the capercaillie suggests that the digestion of these compounds might be facilitated by the microbiota. An initial screening using shallow shotgun sequencing will indicate microbial differences between wild and captive capercaillies to identify taxa and functions related to the degradation of resin and phenol that might be missing in captive individuals.

(2) Is it meaningful to implement a holo-omic approach?

If the captive bred individuals originate from the same population as they are meant to be released in, then the system is relatively simple with two similar genetic backgrounds (wild and captive-bred from the same wild population). This means that the effect of genetics is roughly the same for wild and captive conspecifics, which will allow researchers to study the impact of the environment (i.e., a diet of pine needles) on microbiota functionality (Box 2. Figure 3C: Impact of Environment). If captive bred individuals originating from one population are to be released to increase the number of animals in another population, then it becomes increasingly important to consider that host gene functionality between populations might vary and the contribution from the microbiota to these functions are likely to also vary. It is therefore important to consider if the genes or allelic variants necessary for an optimal digestion of conifer needles are present, either inherent to the host genome or in the metagenome.

(3) Which omic levels are relevant and how to maximize the amount of useful data derived from them?

If the initial screening of the metagenome indicates a lack of functions related to the metabolism of phenol and resin in captive capercaillies, the next step will be to gradually feed them more of their natural diet of pine needles and subsequently screen both the metagenome and (meta)transcriptome. Screening both the transcriptome and metatranscriptome will allow conservationists to uncover complementary interactions between host and microbiota genes. If the genes of interest are suddenly present and expressed then the dietary change has been enough to provide the captive capercaillies with a "wild microbiota" and released animals can then be monitored and their fitness compared with control animals with a captive microbiota. If captive individuals fail to acquire the needed functionalities through the gradual change to a more natural diet other vectors of enrichment should be tested (e.g., natural soil or feces from wild capercaillies).

et al., 2020). However, if the required resolution can only be achieved using shotgun approaches, it is essential to consider the costs of generating the required amount of data and to design the experiments and sampling strategies accordingly. One of the advantages of the holo-omic approach is that all generated data are useful in qualitative terms (i.e., host DNA is valuable information, rather than contamination). However, this does not imply that all generated data are quantitatively useful. The usefulness and cost-effectiveness are influenced by the proportion of host- or microbiota-derived nucleic acids, amino acids, or metabolites. These proportions change drastically across sample types (Marotz et al., 2018) and host taxa (Human Microbiome Project Consortium, 2012; Singh et al., 2014), and an incorrect estimation can require drastic budget adjustments.

The holo-omic approach faces essential challenges, such as those linked to the quantity and complexity of the data to be analyzed. The interactions between different microbes, each synthesizing and metabolizing a variety of molecules, and the interactions between microbes and host cells is extremely complex, with the nature of these interactions being far from uniform and linear. This demands an integrative approach that can account for the different data types under the same inference framework. Generative/mechanistic models exist for many of the individual omics data, such as transcriptome, proteome, and metagenome, but integrating these models under a single inference framework is challenging, given the different data types (compositional versus absolute abundance, discrete vs. continuous) and the vastly different biological processes that underlie them. Thus, developing mechanistic models for such data are an active area of research. In addition, in most current studies, the holo-omics data contain a lot of missing values, e.g., the transcriptomics and microbiome data may not come from the same individuals, and the generated data fall

under the small N (sample size), large P (features) paradigm. That is, the data contain a limited number of independent observations of a large number of features. In the case of holo-omics data, features can include millions of genomic variants, mRNA quantification for thousands of genes, abundance estimates of hundreds to thousands of taxa in the microbiome, and tens to hundreds of phenotypes such as health parameters and growth rates. Unfortunately, the large number of features (P) are not accompanied by a corresponding increase in sample sizes (N), owing to the high cost of generating such comprehensive data for a large number of individuals. Identifying the important determining features in such datasets can be very challenging given the limited number of independent observations. Statistical advances in the last decade including development of deep learning methods are helping address the challenges posed by the high dimensionality and complex correlation structure of the data. Development of such methods is an area of active research where several advances have been made in integrating host-microbiome data (Bersanelli et al., 2016; Heintz-Buschart et al., 2016; Liu et al., 2020).

CONCLUSION

Although still challenged by many limitations, the feasibility to conduct holo-omic research will only increase in the near future, aided by the continuous publication and improvement of macro- and microorganism genomes, the decrease of costs for DNA/RNA sequencing and mass spectrometry, the increase of computational capacities, and the uninterrupted development of analytical tools to analyze the huge amounts of data generated. These trends will allow a broader range of research groups to conduct holo-omic studies and as the need for detailed information on host-microbiota interactions increases in both applied and basic sciences there is no doubt that the holo-omic approach will gain popularity in the future.

ACKNOWLEDGMENTS

The authors thank the following for funding their research: The Danish National Research Foundation award to M.T.P.G. (DNRF143), Villum Fonden grant to M.T.P.G. (17417), Lundbeckfonden grant to A.A. (R250-2017-1351), Danish Council for Independent Research grants to A.A. (DFF 5051-00033) and M.T.L. (DFF 8022-00005), ERC Consolidator Grant to M.T.P.G. (681396-Extinction Genomics), The Norwegian Seafood Research Fund - FHF grant to M.T.P.G. and M.T.L. (901436-HoloFish), H2020 Marie Skłodowska-Curie Individual Fellowship grant to M.T.L. (745723-HappyFish) and the European Union's Horizon 2020 Research and Innovation Programme grant to M.T.P.G., A.A. and M.T.L. (Grant Agreement No 817729 - HoloFood). Furthermore, the authors would like to thank Rob Dunn for his input and discussions.

AUTHOR CONTRIBUTIONS

Conceptualization, L.N., M.T.P.G., and A.A.; Writing – Original Draft, L.N.; Writing – Review & Editing, all authors; Visualization, L.N., S.M., and A.A; Funding Acquisition, A.A., M.T.L., and M.T.P.G.

REFERENCES

Alberdi, A., Aizpurua, O., Bohmann, K., Zepeda-Mendoza, M.L., and Gilbert, M.T.P. (2016). Do vertebrate gut metagenomes confer rapid ecological adaptation? Trends Ecol. Evol. *31*, 689–699.

Allendorf, F.W., Hohenlohe, P.A., and Luikart, G. (2010). Genomics and the future of conservation genetics. Nat. Rev. Genet. *11*, 697–709.

Almeida, A., Mitchell, A.L., Boland, M., Forster, S.C., Gloor, G.B., Tarkowska, A., Lawley, T.D., and Finn, R.D. (2019). A new genomic blueprint of the human gut microbiota. Nature *568*, 499–504.

Bach, J.-F. (2002). The effect of infections on susceptibility to autoimmune and allergic diseases. N. Engl. J. Med. *347*, 911–920.

Bahrndorff, S., Alemu, T., Alemneh, T., and Lund Nielsen, J. (2016). The microbiome of animals: implications for conservation biology. Int. J. Genomics Proteomics *2016*, 5304028.

Bansal, T., Alaniz, R.C., Wood, T.K., and Jayaraman, A. (2010). The bacterial signal indole increases epithelial-cell tight-junction resistance and attenuates indicators of inflammation. Proc. Natl. Acad. Sci. U S A *107*, 228–233.

Barko, P.C., McMichael, M.A., Swanson, K.S., and Williams, D.A. (2018). The gastrointestinal microbiome: a review. J. Vet. Intern. Med. *32*, 9–25.

Bennett, G.M., and Moran, N.A. (2015). Heritable symbiosis: the advantages and perils of an evolutionary rabbit hole. Proc. Natl. Acad. Sci. U S A *112*, 10169–10176.

Bersanelli, M., Mosca, E., Remondini, D., Giampieri, E., Sala, C., Castellani, G., and Milanesi, L. (2016). Methods for the integration of multi-omics data: mathematical aspects. BMC Bioinformatics *17* (*Suppl 2*), 15.

Blekhman, R., Goodrich, J.K., Huang, K., Sun, Q., Bukowski, R., Bell, J.T., Spector, T.D., Keinan, A., Ley, R.E., Gevers, D., et al. (2015). Host genetic variation impacts microbiome composition across human body sites. Genome Biol. *16*, 191.

Bredon, M., Dittmer, J., Noël, C., Moumen, B., and Bouchon, D. (2018). Lignocellulose degradation at the holobiont level: teamwork in a keystone soil invertebrate. Microbiome *6*, 162.

Bryant, J.P. (1980). Selection of winter forage by subarctic browsing vertebrates: the role of plant chemistry. Annu. Rev. Ecol. Syst. *11*, 261–285.

Carvalho, F.A., Koren, O., Goodrich, J.K., Johansson, M.E., Nalbantoglu, I., Aitken, J.D., Su, Y., Chassaing, B., Walters, W.A., González, A., et al. (2012). Transient inability to manage proteobacteria promotes chronic gut inflammation in TLR5-deficient mice. Cell Host Microbe *12*, 139–152.

Cernava, T., Aschenbrenner, I.A., Soh, J., Sensen, C.W., Grube, M., and Berg, G. (2019). Plasticity of a holobiont: desiccation induces fasting-like metabolism within the lichen microbiota. ISME J. *13*, 547–556.

Ching, T., Huang, S., and Garmire, L.X. (2014). Power analysis and sample size estimation for

RNA-Seq differential expression. RNA *20*, 1684–1696.

David, L.A., Maurice, C.F., Carmody, R.N., Gootenberg, D.B., Button, J.E., Wolfe, B.E., Ling, A.V., Devlin, A.S., Varma, Y., Fischbach, M.A., et al. (2014). Diet rapidly and reproducibly alters the human gut microbiome. Nature *505*, 559–563.

Ducatelle, R., Eeckhaut, V., Haesebrouck, F., and Van Immerseel, F. (2015). A review on prebiotics and probiotics for the control of dysbiosis: present status and future perspectives. Animal *9*, 43–48.

Falcinelli, S., Picchietti, S., Rodiles, A., Cossignani, L., Merrifield, D.L., Taddei, A.R., Maradonna, F., Olivotto, I., Gioacchini, G., and Carnevali, O. (2015). Lactobacillus rhamnosus lowers zebrafish lipid content by changing gut microbiota and host transcription of genes involved in lipid metabolism. Sci. Rep. *5*, 9336.

Fei, N., and Zhao, L. (2013). An opportunistic pathogen isolated from the gut of an obese human causes obesity in germfree mice. ISME J. *7*, 880–884.

Feng, Q., Liang, S., Jia, H., Stadlmayr, A., Tang, L., Lan, Z., Zhang, D., Xia, H., Xu, X., Jie, Z., et al. (2015). Gut microbiome development along the colorectal adenoma-carcinoma sequence. Nat. Commun. *6*, 6528.

Ferreira, P.G., Muñoz-Aguirre, M., Reverter, F., Sá Godinho, C.P., Sousa, A., Amadoz, A., Sodaei, R., Hidalgo, M.R., Pervouchine, D., Carbonell-Caballero, J., et al. (2018). The effects of death and post-mortem cold ischemia on human tissue transcriptomes. Nat. Commun. *9*, 490.

Fischer, C.N., Trautman, E.P., Crawford, J.M., Stabb, E.V., Handelsman, J., and Broderick, N.A. (2017). Metabolite exchange between microbiome members produces compounds that influence Drosophila behavior. Elife *6*, e18855.

Fontaine, S.S., and Kohl, K.D. (2020). The gut microbiota of invasive bullfrog tadpoles responds more rapidly to temperature than a non-invasive congener. Mol. Ecol. https://doi.org/10.1111/mec.15487.

Fry, A.J., Palmer, M.R., and Rand, D.M. (2004). Variable fitness effects of Wolbachia infection in *Drosophila melanogaster*. Heredity *93*, 379–389.

Gatlin, D.M., Barrows, F.T., Brown, P., Dabrowski, K., Gaylord, T.G., Hardy, R.W., Herman, E., Hu, G., Krogdahl, Å., Nelson, R., et al. (2007). Expanding the utilization of sustainable plant products in aquafeeds: a review. Aquac. Res. *38*, 551–579.

Gilbert, J.A., Blaser, M.J., Caporaso, J.G., Jansson, J.K., Lynch, S.V., and Knight, R. (2018). Current understanding of the human microbiome. Nat. Med. *24*, 392–400.

Ginsburg, G.S., and McCarthy, J.J. (2001). Personalized medicine: revolutionizing drug discovery and patient care. Trends Biotechnol. *19*, 491–496.

Gough, E., Shaikh, H., and Manges, A.R. (2011). Systematic review of intestinal microbiota transplantation (fecal bacteriotherapy) for recurrent *Clostridium difficile* infection. Clin. Infect. Dis. *53*, 994–1002.

Greenblum, S., Turnbaugh, P.J., and Borenstein, E. (2012). Metagenomic systems biology of the human gut microbiome reveals topological shifts associated with obesity and inflammatory bowel disease. Proc. Natl. Acad. Sci. U. S. A. *109*, 594–599.

Gupta, V.K., Paul, S., and Dutta, C. (2017). Geography, ethnicity or subsistence-specific variations in human microbiome composition and diversity. Front. Microbiol. *8*, 1162.

Gutleben, J., Chaib De Mares, M., van Elsas, J.D., Smidt, H., Overmann, J., and Sipkema, D. (2018). The multi-omics promise in context: from sequence to microbial isolate. Crit. Rev. Microbiol. *44*, 212–229.

Heintz-Buschart, A., May, P., Laczny, C.C., Lebrun, L.A., Bellora, C., Krishna, A., Wampach, L., Schneider, J.G., Hogan, A., de Beaufort, C., et al. (2016). Integrated multi-omics of the human gut microbiome in a case study of familial type 1 diabetes. Nat. Microbiol. *2*, 16180.

Hernandez-Ferrer, C., Ruiz-Arenas, C., Beltran-Gomila, A., and González, J.R. (2017). MultiDataSet: an R package for encapsulating multiple data sets with application to omic data integration. BMC Bioinformatics *18*, 36.

Hickl, O., Heintz-Buschart, A., Trautwein-Schult, A., Hercog, R., Bork, P., Wilmes, P., and Becher, D. (2019). Sample preservation and storage significantly impact taxonomic and functional profiles in metaproteomics studies of the human gut microbiome. Microorganisms *7*, https://doi.org/10.3390/microorganisms7090367.

Hillmann, B., Al-Ghalith, G.A., Shields-Cutler, R.R., Zhu, Q., Gohl, D.M., Beckman, K.B., Knight, R., and Knights, D. (2018). Evaluating the information content of shallow shotgun metagenomics. mSystems *3*, https://doi.org/10.1128/mSystems.00069-18.

Hong, E.P., and Park, J.W. (2012). Sample size and statistical power calculation in genetic association studies. Genomics Inform. *10*, 117–122.

Human Microbiome Project Consortium (2012). A framework for human microbiome research. Nature *486*, 215–221.

Imhann, F., Vich Vila, A., Bonder, M.J., Fu, J., Gevers, D., Visschedijk, M.C., Spekhorst, L.M., Alberts, R., Franke, L., van Dullemen, H.M., et al. (2018). Interplay of host genetics and gut microbiota underlying the onset and clinical presentation of inflammatory bowel disease. Gut *67*, 108–119.

Irianto, A., and Austin, B. (2002). Use of probiotics to control furunculosis in rainbow trout, *Oncorhynchus mykiss* (Walbaum). J. Fish Dis. *25*, 333–342.

Kelly, C.J., Zheng, L., Campbell, E.L., Saeedi, B., Scholz, C.C., Bayless, A.J., Wilson, K.E., Glover, L.E., Kominsky, D.J., Magnuson, A., et al. (2015). Crosstalk between microbiota-derived short-chain fatty acids and intestinal epithelial HIF augments tissue barrier function. Cell Host Microbe *17*, 662–671.

Kichaev, G., Bhatia, G., Loh, P.-R., Gazal, S., Burch, K., Freund, M.K., Schoech, A., Pasaniuc, B., and Price, A.L. (2019). Leveraging polygenic functional enrichment to improve GWAS power. Am. J. Hum. Genet. https://doi.org/10.1016/j.ajhg.2018.11.008.

Kim, S., and Jazwinski, S.M. (2018). The gut microbiota and healthy aging. Gerontology *64*, 513–520.

Koch, E.J., and McFall-Ngai, M. (2019). Model systems for the study of how symbiotic associations between animals and extracellular bacterial partners are established and maintained. Drug Discov. Today Dis. Models. https://doi.org/10.1016/j.ddmod.2019.08.005.

Kokou, F., Sasson, G., Friedman, J., Eyal, S., Ovadia, O., Harpaz, S., Cnaani, A., and Mizrahi, I. (2019). Core gut microbial communities are maintained by beneficial interactions and strain variability in fish. Nat. Microbiol. https://doi.org/10.1038/s41564-019-0560-0.

Kong, H.G., Kim, H.H., Chung, J.H., Jun, J., Lee, S., Kim, H.M., Jeon, S., Park, S.G., Bhak, J., and Ryu, C.M. (2019). The Galleria mellonella hologenome supports microbiota-independent metabolism of long-chain hydrocarbon beeswax. Cell Rep. *26*, 2451–2464.e5.

Koonin, E.V., Aravind, L., and Kondrashov, A.S. (2000). The impact of comparative genomics on our understanding of evolution. Cell *101*, 573–576.

Kumar, H., Lund, R., Laiho, A., Lundelin, K., Ley, R.E., Isolauri, E., and Salminen, S. (2014). Gut microbiota as an epigenetic regulator: pilot study based on whole-genome methylation analysis. MBio *5*, https://doi.org/10.1128/mBio.02113-14.

Kyriakis, S.C., Tsiloyiannis, V.K., Vlemmas, J., Sarris, K., Tsinas, A.C., Alexopoulos, C., and Jansegers, L. (1999). The effect of probiotic LSP 122 on the control of post-weaning diarrhoea syndrome of piglets. Res. Vet. Sci. *67*, 223–228.

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics *9*, 559.

Leidy, J. (1881). Parasites of the Termites (Collins, Printer).

Lesker, T.R., Durairaj, A.C., Gálvez, E.J.C., Lagkouvardos, I., Baines, J.F., Clavel, T., Sczyrba, A., McHardy, A.C., and Strowig, T. (2020). An integrated metagenome catalog reveals new insights into the murine gut microbiome. Cell Rep. *30*, 2909–2922.e6.

Liang, S., Wu, X., and Jin, F. (2018). Gut-brain psychology: rethinking psychology from the microbiota-gut-brain axis. Front. Integr. Neurosci. *12*, 33.

Limborg, M.T., Alberdi, A., Kodama, M., Roggenbuck, M., Kristiansen, K., and Gilbert, M.T.P. (2018). Applied hologenomics: feasibility and potential in aquaculture. Trends Biotechnol. *36*, 252–264.

Liu, Z., Ma, A., Mathé, E., Merling, M., Ma, Q., and Liu, B. (2020). Network analyses in microbiome based on high-throughput multi-omics data. Brief. Bioinform. https://doi.org/10.1093/bib/bbaa005.

Lloyd-Price, J., Abu-Ali, G., and Huttenhower, C. (2016). The healthy human microbiome. Genome Med. *8*, 51.

Luo, J. (2015). Metabolite-based genome-wide association studies in plants. Curr. Opin. Plant Biol. *24*, 31–38.

Machado, L., Magnusson, M., Paul, N.A., de Nys, R., and Tomkins, N. (2014). Effects of marine and freshwater macroalgae on in vitro total gas and methane production. PLoS One *9*, e85289.

Macke, E., Callens, M., De Meester, L., and Decaestecker, E. (2017). Host-genotype dependent gut microbiota drives zooplankton tolerance to toxic cyanobacteria. Nat. Commun. *8*, 1608.

Małyska, A., Markakis, M.N., Pereira, C.F., and Cornelissen, M. (2019). The microbiome: a life science opportunity for our society and our planet. Trends Biotechnol. https://doi.org/10.1016/j.tibtech.2019.06.008.

Markowiak, P., and Śliżewska, K. (2018). The role of probiotics, prebiotics and synbiotics in animal nutrition. Gut Pathog. *10*, 21.

Marotz, C.A., Sanders, J.G., Zuniga, C., Zaramela, L.S., Knight, R., and Zengler, K. (2018). Improving saliva shotgun metagenomics by chemical host DNA depletion. Microbiome *6*, 42.

Mcilroy, J.R., Segal, J.P., Mullish, B.H., Nabil Quraishi, M., Gasbarrini, A., Cammarota, G., and Ianiro, G. (2019). Current and future targets for faecal microbiota transplantation. Hum. Microbiome J. *11*, 100045.

McKenzie, V.J., Song, S.J., Delsuc, F., Prest, T.L., Oliverio, A.M., Korpita, T.M., Alexiev, A., Amato, K.R., Metcalf, J.L., Kowalewski, M., et al. (2017). The effects of captivity on the mammalian gut microbiome. Integr. Comp. Biol. *57*, 690–704.

Moran, N.A., and Yun, Y. (2015). Experimental replacement of an obligate insect symbiont. Proc. Natl. Acad. Sci. U S A *112*, 2093–2096.

Nicholson, J.K., Holmes, E., Kinross, J., Burcelin, R., Gibson, G., Jia, W., and Pettersson, S. (2012). Host-gut microbiota metabolic interactions. Science *336*, 1262–1267.

Ni, J., and Tokuda, G. (2013). Lignocellulose-degrading enzymes from termites and their symbiotic microbiota. Biotechnol. Adv. *31*, 838–850.

Nowoshilow, S., Schloissnig, S., Fei, J.F., Dahl, A., Pang, A.W.C., Pippel, M., Winkler, S., Hastie, A.R., Young, G., Roscito, J.G., et al. (2018). The axolotl genome and the evolution of key tissue formation regulators. Nature *554*, 50–55.

Okada, H., Ebhardt, H.A., Vonesch, S.C., Aebersold, R., and Hafen, E. (2016). Proteome-wide association studies identify biochemical modules associated with a wing-size phenotype in *Drosophila melanogaster*. Nat. Commun. *7*, 12649.

Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K.S., Manichanh, C., Nielsen, T., Pons, N., Levenez, F., Yamada, T., et al. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. Nature *464*, 59–65.

Qin, J., Li, Y., Cai, Z., Li, S., Zhu, J., Zhang, F., Liang, S., Zhang, W., Guan, Y., Shen, D., et al. (2012). A metagenome-wide association study of gut microbiota in type 2 diabetes. Nature *490*, 55–60.

Quince, C., Walker, A.W., Simpson, J.T., Loman, N.J., and Segata, N. (2017). Shotgun metagenomics, from sampling to analysis. Nat. Biotechnol. *35*, 833–844.

Razgour, O., Forester, B., Taggart, J.B., Bekaert, M., Juste, J., Ibáñez, C., Puechmaille, S.J., Novella-Fernandez, R., Alberdi, A., and Manel, S. (2019). Considering adaptive genetic variation in climate change vulnerability assessment reduces species range loss projections. Proc. Natl. Acad. Sci. U S A *116*, 10418–10423.

Reese, A.T., Pereira, F.C., Schintlmeister, A., Berry, D., Wagner, M., Hale, L.P., Wu, A., Jiang, S., Durand, H.K., Zhou, X., et al. (2018). Microbial nitrogen limitation in the mammalian large intestine. Nat. Microbiol. *3*, 1441–1450.

Rogler, G., and Vavricka, S. (2015). Exposome in IBD: recent insights in environmental factors that influence the onset and course of IBD. Inflamm. Bowel Dis. *21*, 400–408.

Rohart, F., Gautier, B., Singh, A., and Lê Cao, K.-A. (2017). mixOmics: an R package for 'omics feature selection and multiple data integration. PLoS Comput. Biol. *13*, e1005752.

Rosenberg, E., Koren, O., Reshef, L., Efrony, R., and Zilber-Rosenberg, I. (2007). The role of microorganisms in coral health, disease and evolution. Nat. Rev. Microbiol. *5*, 355–362.

Rosshart, S.P., Vassallo, B.G., Angeletti, D., Hutchinson, D.S., Morgan, A.P., Takeda, K., Hickman, H.D., McCulloch, J.A., Badger, J.H., Ajami, N.J., et al. (2017). Wild mouse gut microbiota promotes host fitness and improves disease resistance. Cell *171*, 1015–1028.e13.

Rudman, S.M., Greenblum, S., Hughes, R.C., Rajpurohit, S., Kiratli, O., Lowder, D.B., Lemmon, S.G., Petrov, D.A., Chaston, J.M., and Schmidt, P. (2019). Microbiome composition shapes rapid genomic adaptation of *Drosophila melanogaster*. Proc. Natl. Acad. Sci. U S A *116*, 20025–20032.

Rychlik, I. (2020). Composition and function of chicken gut microbiota. Animals (Basel) *10*, 103.

Sbahi, H., and Di Palma, J.A. (2016). Faecal microbiota transplantation: applications and limitations in treating gastrointestinal disorders. BMJ Open Gastroenterol. *3*, e000087.

Schmid, B. (1992). Phenotypic variation in plants. Evol. Trends Plants *6*, 45–60.

Sekula, P., Goek, O.N., Quaye, L., Barrios, C., Levey, A.S., Römisch-Margl, W., Menni, C., Yet, I., Gieger, C., Inker, L.A., et al. (2016). A metabolome-wide association study of kidney function and disease in the general population. J. Am. Soc. Nephrol. *27*, 1175–1188.

Sheth, R.U., Li, M., Jiang, W., Sims, P.A., Leong, K.W., and Wang, H.H. (2019). Spatial metagenomic characterization of microbial biogeography in the gut. Nat. Biotechnol. *37*, 877–883.

Singh, K.M., Shah, T.M., Reddy, B., Deshpande, S., Rank, D.N., and Joshi, C.G. (2014). Taxonomic and gene-centric metagenomics of the fecal microbiome of low and high feed conversion ratio (FCR) broilers. J. Appl. Genet. *55*, 145–154.

Stringlis, I.A., Yu, K., Feussner, K., de Jonge, R., Van Bentum, S., Van Verk, M.C., Berendsen RL, Bakker, P.A.H.M., Feussner, I., and Pieterse, C.M.J. (2018). MYB72-dependent coumarin exudation shapes root microbiome assembly to promote plant health. Proc. Natl. Acad. Sci. U S A *115*, E5213–E5222.

Suzuki, T.A., Phifer-Rixey, M., Mack, K.L., Sheehan, M.J., Lin, D., Bi, K., and Nachman, M.W. (2019). Host genetic determinants of the gut microbiota of wild mice. Mol. Ecol. https://doi.org/10.1111/mec.15139.

Vaishnava, S., Yamamoto, M., Severson, K.M., Ruhn, K.A., Yu, X., Koren, O., Ley, R., Wakeland, E.K., and Hooper, L.V. (2011). The antibacterial lectin RegIIIg promotes the spatial segregation of microbiota and host in the intestine. Science *334*, https://doi.org/10.1126/science.1208930.

Virtue, A.T., McCright, S.J., Wright, J.M., Jimenez, M.T., Mowel, W.K., Kotzin, J.J., Joannas, L., Basavappa, M.G., Spencer, S.P., Clark, M.L., et al. (2019). The gut microbiota regulates white adipose tissue inflammation and obesity via a family of microRNAs. Sci. Transl. Med. *11*, eaav1892.

Walker, W.A. (2017). Dysbiosis. In The Microbiota in Gastrointestinal Pathophysiology: Implications for Human Health Prebiotics, Probiotics, and Dysbiosis, M.H. Floch, Y. Ringel, and W.A. Walker, eds. (Academic Press), pp. 227–232.

Wang, Q., Wang, K., Wu, W., Giannoulatou, E., Ho, J.W.K., and Li, L. (2019). Host and microbiome multi-omics integration: applications and methodologies. Biophys. Rev. *11*, 55–65.

Wan, Z.Y., Xia, J.H., Lin, G., Wang, L., Lin, V.C.L., and Yue, G.H. (2016). Genome-wide methylation analysis identified sexually dimorphic methylated regions in hybrid tilapia. Sci. Rep. *6*, 35903.

Welter, D., MacArthur, J., Morales, J., Burdett, T., Hall, P., Junkins, H., Klemm, A., Flicek, P., Manolio, T., Hindorff, L., et al. (2014). The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Res. *42*, D1001–D1006.

Werren, J.H., Baldo, L., and Clark, M.E. (2008). Wolbachia: master manipulators of invertebrate biology. Nat. Rev. Microbiol. *6*, 741–751.

Wienemann, T., Schmitt-Wagner, D., Meuser, K., Segelbacher, G., Schink, B., Brune, A., and

Berthold, P. (2011). The bacterial microbiota in the ceca of Capercaillie (*Tetrao urogallus*) differs between wild and captive birds. Syst. Appl. Microbiol. *34*, 542–551.

Wu, H.-J., and Wu, E. (2012). The role of gut microbiota in immune homeostasis and autoimmunity. Gut Microbes *3*, 4–14.

Xu, Y., and Zhou, X. (2018). Applications of single-cell sequencing for multiomics. Methods Mol. Biol. *1754*, 327–374.

Zhang, G., Li, C., Li, Q., Li, B., Larkin, D.M., Lee, C., Storz, J.F., Antunes, A., Greenwold, M.J., Meredith, R.W., et al. (2014). Comparative genomics reveals insights into avian genome evolution and adaptation. Science *346*, 1311–1320.

Zhao, L., Zhang, F., Ding, X., Wu, G., Lam, Y.Y., Wang, X., Fu, H., Xue, X., Lu, C., Ma, J., et al. (2018). Gut bacteria selectively promoted by dietary fibers alleviate type 2 diabetes. Science *359*, 1151–1156.

Recovering high-quality host genomes from gut metagenomic data through genotype imputation

Annex 2

ADVANCED GENETICS
Open Access
www.advgenet.com

Recovering High-Quality Host Genomes from Gut Metagenomic Data through Genotype Imputation

Sofia Marcos, Melanie Parejo, Andone Estonba, and Antton Alberdi**

Metagenomic datasets of host-associated microbial communities often contain host DNA that is usually discarded because the amount of data is too low for accurate host genetic analyses. However, genotype imputation can be employed to reconstruct host genotypes if a reference panel is available. Here, the performance of a two-step strategy is tested to impute genotypes from four types of reference panels built using different strategies to low-depth host genome data (≈2× coverage) recovered from intestinal samples of two chicken genetic lines. First, imputation accuracy is evaluated in 12 samples for which both low- and high-depth sequencing data are available, obtaining high imputation accuracies for all tested panels (>0.90). Second, the impact of reference panel choice in population genetics statistics on 100 chickens is assessed, all four panels yielding comparable results. In light of the observations, the feasibility and application of the applied imputation strategy are discussed for different species with regard to the host DNA proportion, genomic diversity, and availability of a reference panel. This method enables leveraging insofar discarded host DNA to get insights into the genetic structure of host populations, and in doing so, facilitates the implementation of hologenomic approaches that jointly analyze host and microbial genomic data.

1. Introduction

The large molecular datasets generated through shotgun DNA sequencing regularly contain useful information to characterize taxa, functions, and structures beyond the primary aim of the study. This is especially true in metagenomic datasets that often present mixtures of DNA from eukaryotic, prokaryotic, and viral origin.[1] While primarily used for characterizing the genomic architecture of microbial communities, metagenomic data generated from intestinal contents or feces can also be used for extracting useful genomic information of the animal host.[2] In fact, hologenomic approaches, that entail joint analysis of animal genomes along with metagenomes of host-associated microorganisms to study animal-microbiota interactions, can benefit from such optimization strategies.[3,4]

However, mining host genomic data from metagenomic datasets presents a number of challenges. The fraction of host sequences in the metagenomic mixture is often unpredictable, and can range from a negligible proportion (<5%) to an almost complete representation (>95%) of the sample,[5] even within a single taxon and sample type.[6] Hence, a given amount of metagenomic sequencing effort does not ensure that the desired depth of host DNA sequencing will be reached. When the host DNA fraction in the metagenomic mixture is low, achieving the desired sequencing depth requires increasing sequencing effort, with its respective economic burden. In consequence, the amount of host DNA sequences generated is often insufficient for accurate variant calling.

One useful strategy for efficient data mining of host genomic information is genotype imputation, which consists in estimating missing haplotypes of poorly characterized genomes using a reference panel of high-quality genotypes.[7] Thus, information gaps of genomes with very low sequencing depth can be reconstructed based on the haplotype information of a properly characterized representative panel. Genotype imputation of single nucleotide polymorphisms (SNPs) is a widely employed approach in association studies to increase the density of variants of genomic datasets.[8,9] The recent generation of large high-quality genomic databases, such as the human 1000 Genomes Project (1000G)[10] and the 1000 Bull Genomes Project,[11] has improved the accuracy of imputation and increased the statistical power of association analyses, especially for rare variants.[12,13] However, ideal reference panels are only available for a limited number of model and farm species, and they require high computational capacity.

When large reference panels are not available, an alternative strategy is to create a custom panel using a representative subset

S. Marcos, M. Parejo, A. Estonba
Applied Genomics and Bioinformatics
University of the Basque Country (UPV/EHU)
Leioa, Bilbao 48940, Spain
E-mail: sofia.marcos@ehu.eus
A. Alberdi
Center for Evolutionary Hologenomics
GLOBE Institute
University of Copenhagen
Copenhagen 1353, Denmark
E-mail: antton.alberdi@sund.ku.dk

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/ggn2.202100065

DOI: 10.1002/ggn2.202100065

of genomes of the studied population.[14,15] This approach can be more cost-efficient because when haplotype diversity is limited, genomic information of a subset of the population can successfully impute haplotype information to the rest of the population. Moreover, the study-specific panel can be combined with individuals from public databases,[14,15] which has been previously employed in sheep,[16] pig,[17] and chicken studies.[18]

Nevertheless, in addition to the size and diversity of the panel,[19] imputation strategy may also affect the accuracy of recovered genotypes.[20] In contrast to the standard imputation method, in which low density SNP arrays are imputed to high density based on a reference panel, shallow shotgun sequenced data display particular challenges, as no genotype is known with certainty. Recently, a two-step imputation strategy for ultra low-depth coverage samples (<1×) was introduced.[21] This approach relies on updating genotype likelihoods using a reference panel before imputing the missing genotypes in order to recover a higher density of SNPs with greater confidence. It was first proposed in human population genetics as an alternative to genotyping arrays,[21] and later applied to recover ancient human genomes.[22] To the best of our knowledge, such an imputation strategy has not been implemented yet in non-model animal populations with a limited number of available samples as a reference panel.

Here, we present a straightforward approach to recover high-quality host genomes from gut metagenomic data, showcased in farm chickens. We evaluate how the reference panel composition and sample depth of coverage affects imputation performance using four panels designed according to the resources scientists studying microbial metagenomics may have access to. We first calculate imputation accuracy between imputed and true genotypes in three chromosomes using 12 validation samples for which high-depth sequencing data are also available. Then, we employ a bigger dataset of 100 individuals to impute all autosomal chromosomes and explore how the choice of the reference panel affects commonly used population genetics statistics. Aiming at facilitating its implementation in other study systems, we provide the bioinformatic pipeline and discuss suitable panels and minimal depth thresholds required for successful imputation in light of the characteristics of the study system.

2. Experimental Section

2.1. Ethical Statement

Animal experiments were performed at IRTA's experimentation facilities in Tarragona under the permit FUE-2018-00813123 issued by the Government of Catalonia, in compliance with the Spanish Royal Decree on Animal Experimentation RD53/2013 and the European Union Directive 2010/63/EU about the protection of animals used in the experimentation.

2.2. Target Population and Reference Panels

2.2.1. Target Population

Genomic information of the target population of 100 chickens belonging to two broiler lines (Cobb500 and Ross308, hereafter simply Cobb and Ross) was generated from metagenomic DNA extracted from the cecum contents of the birds. In short, cecum content (≈100 mg) was collected right after euthanizing the animal, and preserved in E-matrix tubes with DNA/RNA Shield buffer (Zymo Research, Cat. No. BioSite-R1200-125) at −20 °C until extraction. After physical cell disruption through bead-beating using a Tissuelyser II machine (Qiagen, Cat. No. 85300), DNA extraction was performed using a custom nucleic acid extraction protocol (details explained in Bozzi et al.),[23] and sequencing libraries were prepared using the adapter ligation-based BEST protocol.[24] Paired-end 150 bp-long reads were generated on a MGISEQ-2000 sequencing platform over multiple sequencing lanes. Sequencing effort was decided based on the desired depth of the metagenomic fraction of the samples, which was the primary objective of the data generation. A preliminary screening revealed that cecum contents contained a large fraction of microbial DNA (>80–95%), and a limited relative amount of host DNA (<5–15%) (**Figure 1**A). Aiming at ≈15 GB (gigabases, ≈50 million reads) of bacterial DNA per sample, cecum samples yielded between 0.5 and 4 GB of host DNA, which was equivalent to 0.5–4× depth of coverage of the chicken genome (≈1.05 GB). Raw data will be available from European Nucleotide Archive (ENA), with BioProject accession no. PRJEB43192 (https://www.ebi.ac.uk/ena/browser/view/PRJEB43192?show = component-projects). Until the release date, data will be made available upon request.

2.2.2. Reference Samples

Internal and external high-quality genome sequence data were used to create the four reference panels tested in the study. The internal reference data were generated from ileum content samples of 12 randomly selected individuals included in the target population (5 Cobb and 7 Ross), following the same procedures as explained above. In contrast to cecum samples, ileum contents contain a very large fraction (>90–95%) of host DNA, and a small representation of microbial DNA. Hence, in order to generate a comparable amount of microbial data to that of the cecum, ileum samples were sequenced aiming 100 GB/sample. This sequencing effort yielded ≈90 GB of host DNA (≈80–90× depth of chicken genome), which enabled generating a high-quality internal panel from a subset of the studied population. In addition, chicken DNA sequence data of 40 broilers (meat producers), 20 layers (egg producers), and 20 red junglefowls (RJF, wild chickens) generated by Qanbari et al. from blood samples were used as external reference data (Figure 1A).[25]

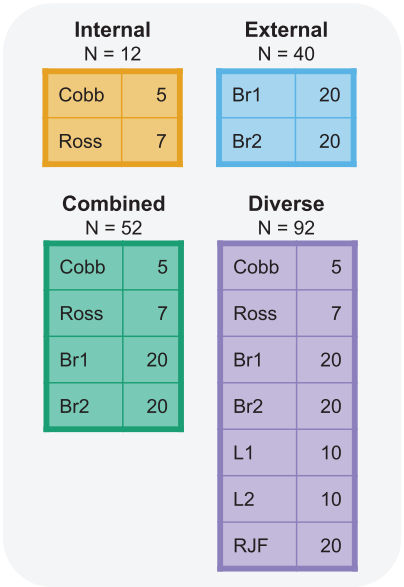
2.2.3. Composition of Reference Panels

Different combinations of the internal and external reference samples were used to create the four reference panels: i) the internal panel comprised 12 animals from the target population, ii) the external panel comprised 40 animals from two broiler lines (different from the target population), iii) the combined panel combined the previous two panels, and iv) the diverse panel also contained more distant populations (Figure 1B). The

(a) Populations

Populations	Target population	Internal reference	External reference
Tissue type	Caecum gut content	Ileum gut content	Blood
Host DNA (%)	~ 5	~ 95	-
Breeds	Cobb (N=47) Ross (N=53)	Cobb (N=5) Ross (N=7)	Br1 (N=20) Br2 (N=20) L1 (N=10) L2 (N=10) RJF (N=20)
Total N	100	12	80

(b) Reference panels



(c) Pipeline

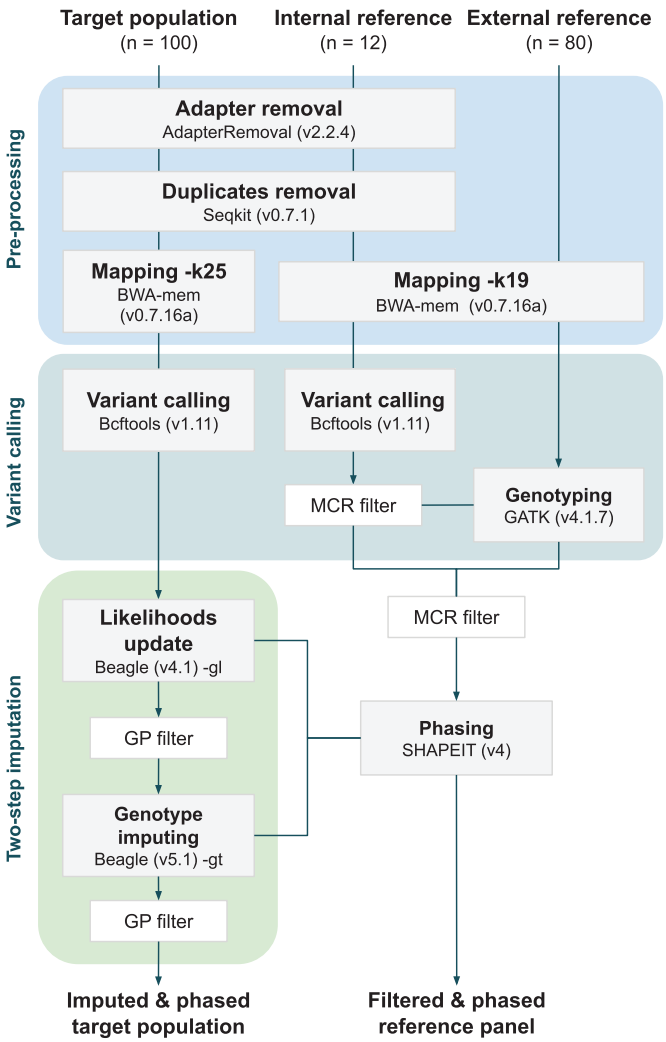


Figure 1. Study design and imputation pipeline for recovering host DNA. A) The characteristics of the three datasets. B) Composition and number of samples of the four reference panels used for imputation. Breeds are coded as Br1 = broiler line A, Br2 = broiler line B, L1 = white layer, L2 = brown layer, RJF = red junglefowl. C) The study design has three datasets: the target population, and internal and external reference samples. The bioinformatic procedure is divided into three steps: pre-processing, variant calling, and imputation. The input format of the starting step is a FASTQ file. After mapping we obtain a BAM file and from variant calling to the final step, procedures are performed using VCF file. The green box represents the steps proposed by Hui et al. (2020). Genotype probability (GP) filters are used during imputation and missing call rate (MCR) filters during panel design.

four panels varied in size and genomic diversity in order to see whether the composition of the reference panels affected imputation accuracy. With the internal panel, it was tested if a small subset of the target population was enough for a proper imputation in low-quality host sequence data derived from metagenomic samples. The use of just an external panel was considered to test whether it was a viable option for studies with a shortage of samples or a limited budget for high-depth host sequencing. The combined panel on the other hand, permits combining both resources. Last, the diverse panel enabled to test whether including distantly related individuals would be more effective than the three previously mentioned strategies.

2.3. Pipeline for Recovering Host Genotypes from Metagenomic Data

2.3.1. Data Pre-Processing

All the metagenomic sequence data that contained both host and microbial DNA, were pre-processed using identical bioinformatic procedures. In short, sequencing adapters were removed using AdapterRemoval (v2.2.4)[26] and exact duplicates using seqkit "rmdup" (v0.7.1)[27] prior to the read-mapping. Read-alignment to the chicken reference genome (galGal6; NCBI Assembly accession GCF_000002315.6) was

conducted with BWA-MEM (v0.7.16a).[28] Default parameters, except for the minimum seed length (-k) were employed, which was increased to 25 in order to reduce the number of incorrectly aligned read pairs as recommended by Robinson et al.[29] This strategy was employed because the standard alignment (-k 19) presented an unconventional distribution of reads across the genome, that is, unspecified read mapping leading to regions being stacked with 80+ reads (Table S1, Supporting information). The "-M" flag, which was used to mark shorter split hits as secondary mappings were added. Aligned reads were sorted and converted into sample-specific BAM files before filtering out the metagenomic fraction (unmapped) using SAMtools view (v1.11)[30] with "-b" and "-F12" flags. Mapping statistics including depth and breadth of coverage as well as percentage of mapped reads were calculated using SAMtools' depth and flagstat functions. DNA damage of 10 cecum samples was assessed by DamageProfiler (v1.1).[31]

Pure genomic data (with no microbial fraction) generated by others[25] were downloaded from the EMBL-EBI ENA database, and mapped to the same chicken reference genome using BWA-MEM with "-k" default value and "-M" flag.

2.3.2. Variant Calling and Genotyping

Variants in the target population were called by chromosome with the mpileup utility of SAMtools using standard parameters (-C 50 -q 30 -Q 20). Variant calling was performed with "-m" and "-v" flags to allow variants to be called on all samples simultaneously. Raw variants were filtered using BCFtools (v1.11)[32] commands "-m2," "-M2," and "-v snps" to keep only bi-allelic SNPs.

Variants of the internal reference samples were called the same way, but additionally, low quality variants with a lower base quality than 30 (QUAL<30) and variants with a base depth higher than three times the average (DP<(AVG(DP)*3) were removed to ensure only highly reliable variants were retained.

Since the interest was solely in imputing variants present in the target population, the external reference samples were genotyped by defining variant sites detected in the internal reference samples. Genotyping was performed for all autosomal chromosomes with GATK (v4.1.7.0)[33] HaplotypeCaller using the "–min-base-quality-score 20," "–standard-min-confidence-threshold-for-calling 30," "–alleles," and "-L" parameters to obtain calls at all given positions, followed by GATK SelectVariants "–"select-type-to-include SNP" to only include SNPs.

In preliminary analyses, variants in the external panel were also called to examine the overlap with the variants present in the internal reference samples. The same procedures explained above were used for chromosome one (GGA1). Genotyping based on the positions of the internal panel and variant calling from scratch were compared by using the 40 broilers from the external panel (Figure 1B). A similar number of variants were obtained for the genotyped (2.5 M) and the variant called VCF files (2.7 M). Moreover, 28% of the variants from the 40 broilers were not present in the internal reference samples (Figure S2, Supporting information). Thus, it was decided to genotype the rest of the samples to reduce possible bias through the high number of variants specific to the external reference for the imputation of the target population.

2.3.3. Two-Step Imputation via Genotype Likelihood Updates

Genotypes were imputed from the four aforementioned reference panels to the target population using a two-step strategy. Prior to imputation, the reference panels were filtered by excluding variants with missing genotypes to remove any potential noise caused by inference errors, and subsequently phased using SHAPEIT (v4).[34]

Imputation was performed in two steps following Homberg et al. (2019) and Hui et al. (2020). First, genotype likelihoods were updated based on one of the reference panels using Beagle 4.1.[35] Beagle 4.1 accepted a probabilistic genotype input with "-gl" mode, and it only updated sites that were present in the input file. Second, missing genotypes in the input file were imputed using Beagle 5.1 with "-gt" mode using the same reference panel. Beagle 5.1 only accepts files with a genotype format field, like later versions than Beagle 4.1. Therefore, the latest version cannot be used for both steps. Format field genotype probabilities (GP) were generated in both steps in order to enrich confident genotypes. The highest GP was required to exceed a threshold of 0.99 after both steps using BCFtools "+setGT" plugin. The rest of the parameters were set to default. Both steps' input and output files were in VCF format. The schematic steps detailed in methods can be found in Figure 1C. Bioinformatic resources, including scripts, sample ENA accession codes and data files have been archived in the following link (10.5281/zenodo.6473506).

2.4. Imputation Accuracy Using 12 Validation Samples

The accuracy of the imputation using the four reference panels was tested using the 12 individuals for which both low-depth (target population) and high-depth (internal reference samples) sequence data were generated from cecum and ileum contents, respectively, hereafter referred to as validation samples. The low-depth samples of the 12 individuals had a depth of coverage spanning 0.05× to 3.73×, and breadth of coverage from 10% to 80%. For an unbiased evaluation, a leave-one-out cross-validation (LOOCV) approach was employed by excluding each of the 12 validation samples once from the reference panel in each of the different imputation scenarios. Considering the large size-variation of avian chromosomes, a macrochromosome (GGA1, 197.6 MB), a mid-size chromosome (GGA7, 36.7 MB), and a microchromosome (GGA20, 13.9 MB) were selected for the test to optimize runtime and computational resources. Concordance between the internal reference samples and imputed genotypes was calculated for each individual chicken using VCFtools, with the "–diff-discordance-matrix" option. Precision of heterozygous (het.) sites was also calculated, since these alleles were the most difficult to impute correctly. Last, imputation accuracy was estimated for variants in different minor allele frequency (MAF) bins to evaluate whether rare variants were correctly imputed. The variant frequencies were thus extracted from the internal panel by analyzing the precision of het. sites for the GGA1 in bins of 0–0.05, 0.05–0.1, 0.1–0.3, and >0.3.

2.5. Impact of Reference Panel on Population Genetics Inference

The implications of using different reference panels in downstream population genetics analyses were explored, including

inferences of population structure, estimates of genetic diversity, and genome scans for signatures of selection.

These analyses were run in all but two outlier samples with depths of coverage of 0.07 and 0.05×, which were below the threshold of 0.28× corresponding to the lowest successfully imputed sample in the validation set (genotype concordance of >0.90 and het. sites precision of >0.75, see results below). 100 samples were thus used (47 Cobb and 53 Ross) for which the host DNA recovery pipeline for all the autosomal chromosomes was run, and the commonly used population genetics statistics were analyzed, including observed heterozygosity (Ho), nucleotide diversity (π), pairwise distance as estimated through identity-by-state (1-IBS), and kinship estimates. The same analyses were also conducted for 10 validation samples (for the low-depth and high-depth samples) after excluding two of them, whose respective counterparts in the target populations (with 0.05 and 0.07× depth) were filtered out as mentioned above. The imputed datasets with each of the panels were filtered for missingness 0 with PLINK (v1.9).[36]

For measuring population genetics parameters, the VCF files were filtered for MAF >0.05. Ho, the percentage of het. sites over the number of variant sites, was calculated for each individual using"–het" in PLINK. π, the average pairwise sequence difference per nucleotide site, was calculated in 40 kb windows with 20 kb step size across autosomal chromosomes using VCFtools.[37] For the validation samples whole-genome windowed values were averaged to generate a genome-wide π for each individual. For the target population, π was calculated for each broiler line. Pairwise distance was calculated using "–distance square 1-ibs" in PLINK. Kinship was calculated with the command "–make-king square" using Manichaikul et al.'s estimator in PLINK (v2).[38]

It was further tested whether genome scans for selection between the Cobb and Ross population with each of the imputed datasets yielded consistent results. To this end, population differentiation along the genome was calculated using Weir and Cockerham's fixation index (F_{ST}) estimate for each panel.[39] F_{ST} was calculated in sliding windows of 40 kb with 20 kb overlap across autosomal chromosomes. Window-based F_{ST} values were then normalized, and regions with values above the 99th and 99.9th empirical percentile were considered as candidates for selective sweep regions.[40] The overlap of these regions across the datasets using the different reference panels was used as an estimate of consistency.

2.6. Statistical Analysis

Kruskal-Wallis test was performed to test for differences in average concordance across chromosomes in the 12 validation samples.[41] A paired sample *T*-test and *F*-test were performed for concordance and precision of het. sites to verify if the difference in means and variances were significant between reference panels.[42] *T*-test *p*-values were adjusted using Bonferroni's correction method.[43] Paired sample *T*-tests were performed for Ho and π estimates in the 100 chicken population. Mantel test was performed with the R package ade4 to test the correlation between the resulting matrices from the pairwise distance and kinship analyses.[44]

3. Results

3.1. Alignment and Coverage

The mapping statistics of the 100 samples used to characterize the target population (cecum content), and the 12 internal reference samples (ileum content) were drastically different. Cecum samples showed an average of 1.84 ± 2.35× (mean ± SD) depth of coverage and 52.41 ± 24.20% breadth of coverage. Ileum samples had 92.70 ± 7.64% of host DNA and an average depth of 93.16 ± 9.07×, practically covering the entire reference genome (98.89 ± 0.01%).

3.2. Imputation Accuracy of 12 Validation Samples

The internal (I, n = 12), external (E, n = 40), combined (C, n = 52), and diverse (D, n = 92) reference panels were used to study (i) the effect of panel size and diversity, and (ii) sample depth of coverage threshold on imputation accuracy in three chromosomes with contrasting dimensions. Variant calling in the internal reference samples detected 2.4 M, 470 K, and 182 K putative SNPs in chromosomes GGA1, GGA7, and GGA20, respectively. After genotyping the external reference samples and combining them to create the external, combined, and diverse panels, each panel was filtered before being phased. As a consequence, the filtering step decreased the number of SNPs by 13.83 ± 1.36% for the external and combined, and by 23.80 ± 0.99% for the diverse, which yielded panels with different numbers of SNPs (**Figure 2**A). Regarding the percentage of imputed variants in the 12 validation samples, more than 96% of the total SNPs in each panel successfully passed the multiple filters of the pipeline, even for samples with less than 1× coverage (Figure 2B). The proportion of imputed SNPs increased and gained uniformity across samples when the panel was larger but had fewer SNPs. The mean number of imputed SNPs across samples differed between all the panels: I versus E (T = 14.58, p<0.001), E versus C (T = 13.56, p<0.001), and C versus D (T = 11.63, p<0.001). The *F*-test was significant only between the diverse and the rest of the panels: I versus D (F = 30.54, p<0.001), E versus D (F = 24.24, p<0.001), and C versus D (F = 11.31, p<0.001). Results indicate that the variance across samples for the diverse panel greatly decreased compared to the rest (Figure 2B).

For each imputation scenario, genotype concordance, and precision of het. sites were assessed in the validation samples by comparing imputed and true genotypes per individual. After performing LOOCV with the four reference panels, average values of genotype concordance exceeded 0.90 for every panel (**Figure 3**A) and precision of het. sites ranged from 0.78 to 0.91 (Figure 3B). According to Kruskal Wallis tests, the values of concordance (p_I>0.85, p_E>0.85, p_C>0.95, and p_D>0.95), and precision of het. sites (p_I>0.95, p_E>0.85, p_C>0.85, and p_D>0.85) did not differ across chromosomes. However, mean values differed between panels for each chromosome (Figure 3). Concordance values significantly differed when comparing the internal, external, and combined panels (Figure 3A). However, no significant differences were detected between the combined and the diverse panels, indicating that imputation accuracy in terms of overall concordance does not increase by adding more distant individuals. For precision of het. sites, differences were detected for all

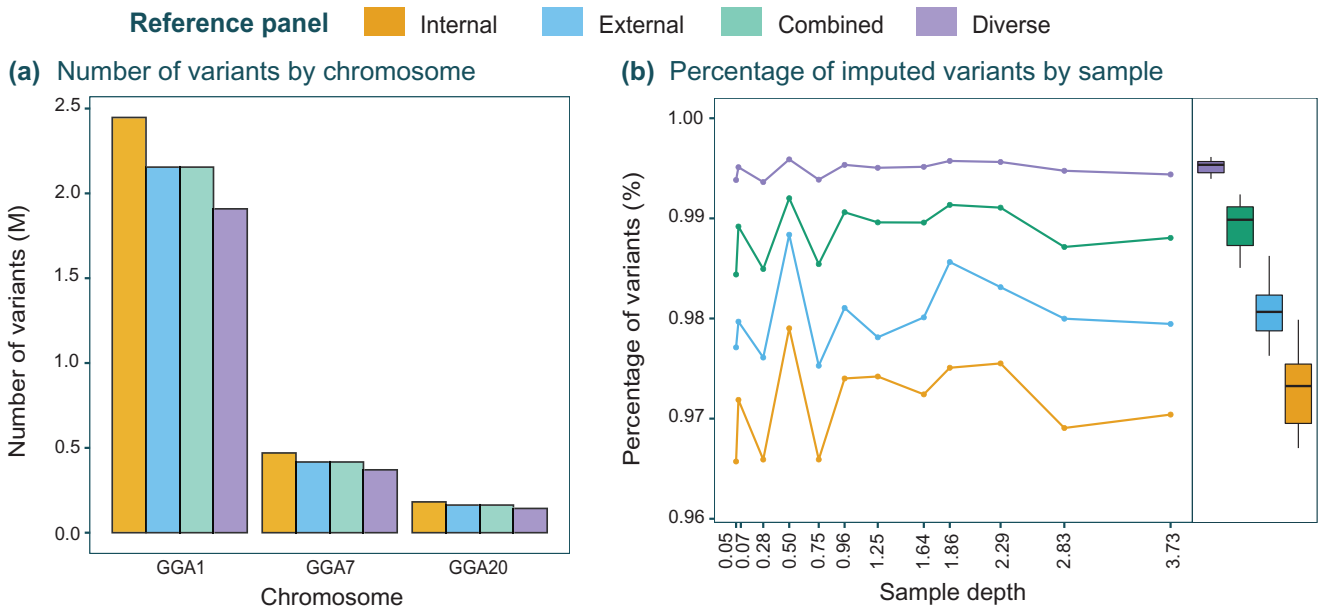


Figure 2. Imputation statistics. A) Number of SNPs in each reference panel for chromosomes GGA1, GGA7, GGA20. B) Depth of coverage and proportion of successfully imputed variants of the 12 validation samples for the three chromosomes tested. Capitalized letters refer to panel names: I = internal, E = external, C = combined, and D = diverse.

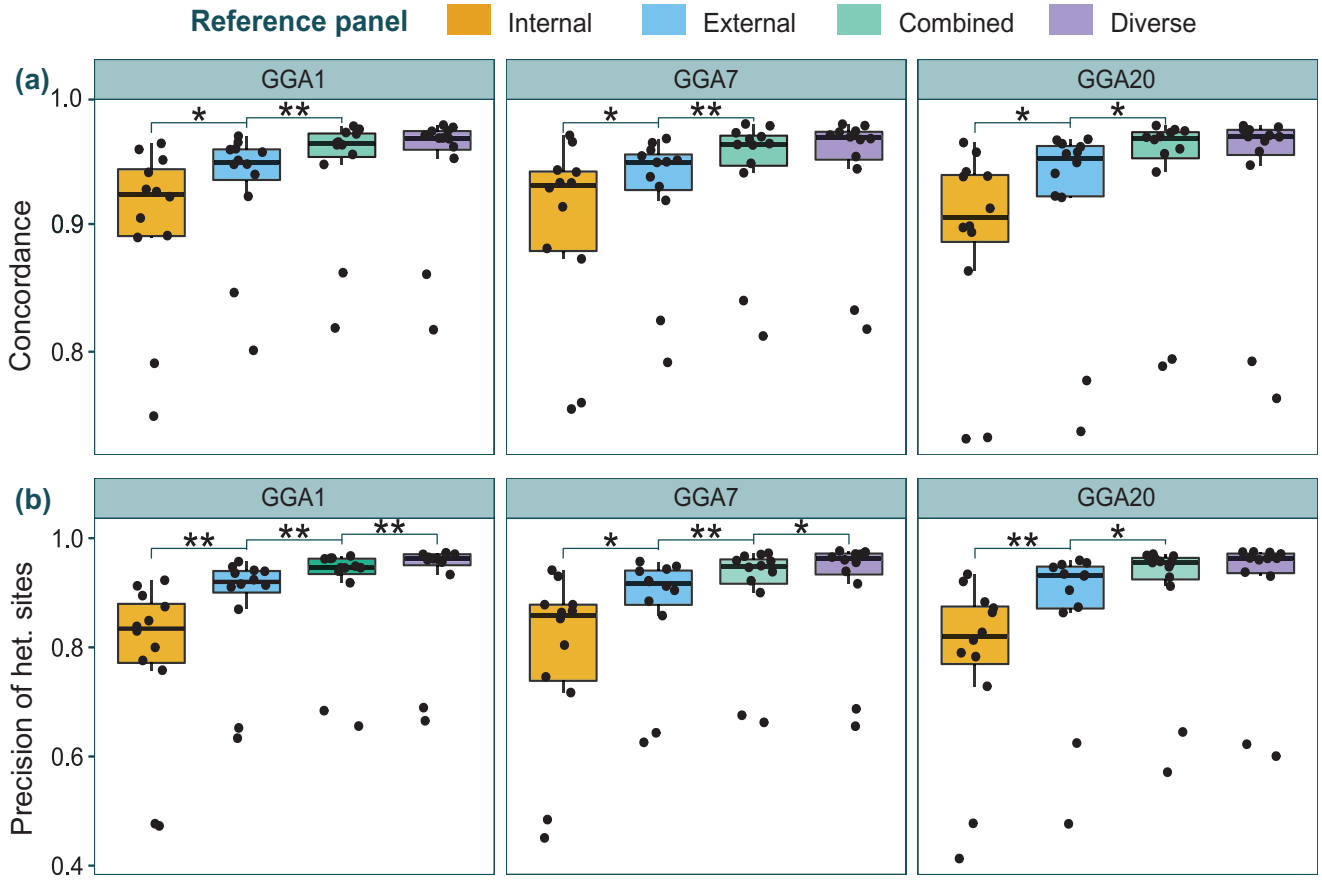


Figure 3. LOOCV test results and comparison of imputation reference panels. A) Genotype concordance, and B) precision of heterozygous (het.) sites between imputed (low-depth 12 validation samples) and true (internal reference samples) genotypes on chromosomes GGA1, GGA7, and GAA20. Paired *T*-tests were performed to identify significant differences in means: the following symbols ("**," "*") indicate different *p*-value cut-points (<0.001, 0.05).

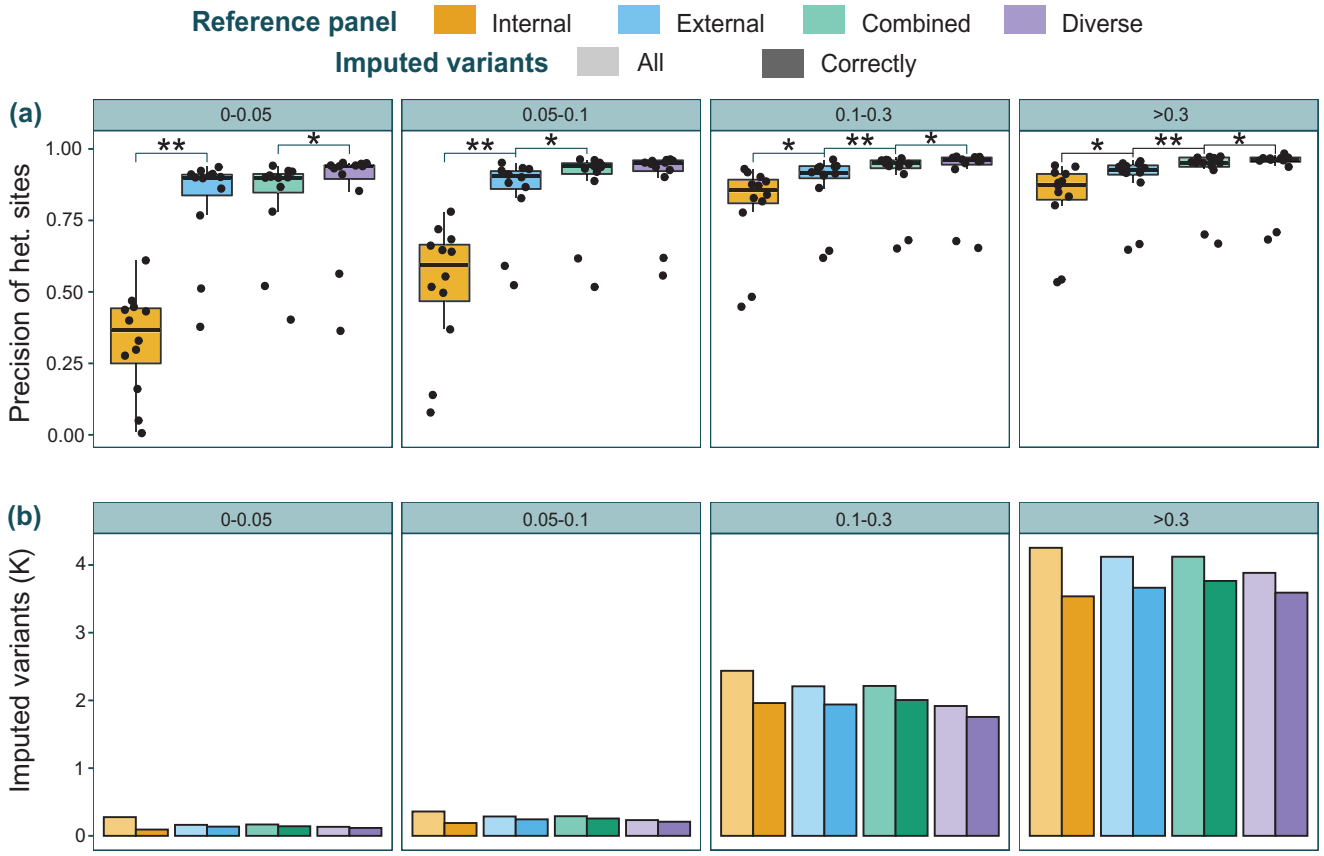


Figure 4. Minor allele frequency variants of LOOCV test. A) Precision of heterozygous sites, and B) number of imputed low-frequency variants for chromosome one (GGA1) divided into four different bins of minor allele frequency ranges: 0–0.05, 0.05–0.1, 0.1–0.3, and >0.3. The lower bars represent correctly imputed variants, while the bars with greater transparency represent the number of all imputed variants within the respective MAF bin. Variants that coincided between imputed (low-depth 12 validation samples) and true (internal reference samples) genotypes were considered correctly imputed variants. Paired *T*-tests were performed to identify significant differences in means across panels: the following symbols ("**," "*") indicate different *p*-value cut-points (<0.001, 0.05).

panels (Figure 3B), including for the combined and the diverse, except for GGA20. This suggests that the het. positions are the most sensitive to the imputation process.

In an attempt to further assess imputation accuracy, we classified variants according to their MAF in four bins (0–0.05, 0.05–0.1, 0.1–0.3, and >0.3) and calculated precision of het. sites, and the number of correctly imputed variants for GGA1 (**Figure 4**). The internal panel, while recovering the largest number of variants, was also the panel with the lowest performance in adequately inferring low-frequency variants, especially for the variants with MAF <0.1 (Figure 4A). Although there was no improvement from the external to the combined panel for the smallest MAF bin, a substantial improvement was seen for the rest of the bins. Some significant differences but not as pronounced were also observed when switching from the combined to the diverse panel. Therefore, the combined panel showed overall the best results with the highest number of correctly imputed variants in all MAF bins (Figure 4B), while maintaining a very similar number of imputed SNPs as the external panel. The diverse panel inferred fewer low-frequency variants, but did so more effectively (Figure 4).

Despite the high overall imputation accuracy, the two samples with depths of 0.05 and 0.07× were outliers that did not achieve a sufficiently high concordance (>0.90) and precision (>0.75) with any of the panels and chromosomes (Figure 3). They were thus excluded from the target population, and we refer from now on to 10 validation samples instead of 12.

3.3. Panel Choice Impact on Population Genetic Inference

3.3.1. Number of Variants and Their Allele Frequency Distribution in the Imputed Target Population

The final number of SNPs recovered from all autosomal chromosomes in the target population with different panels decreased as more distant individuals were included (**Figure 5**A). This was due to the missing call rate (MCR) filter. Using the internal panel, we recovered 11.7 M filtered SNPs in the target population. These were 30% more recovered variants than when using the diverse panel (8.9 M). Most of the excess variants from the internal panel are low-frequency variants that cannot be confidently recovered

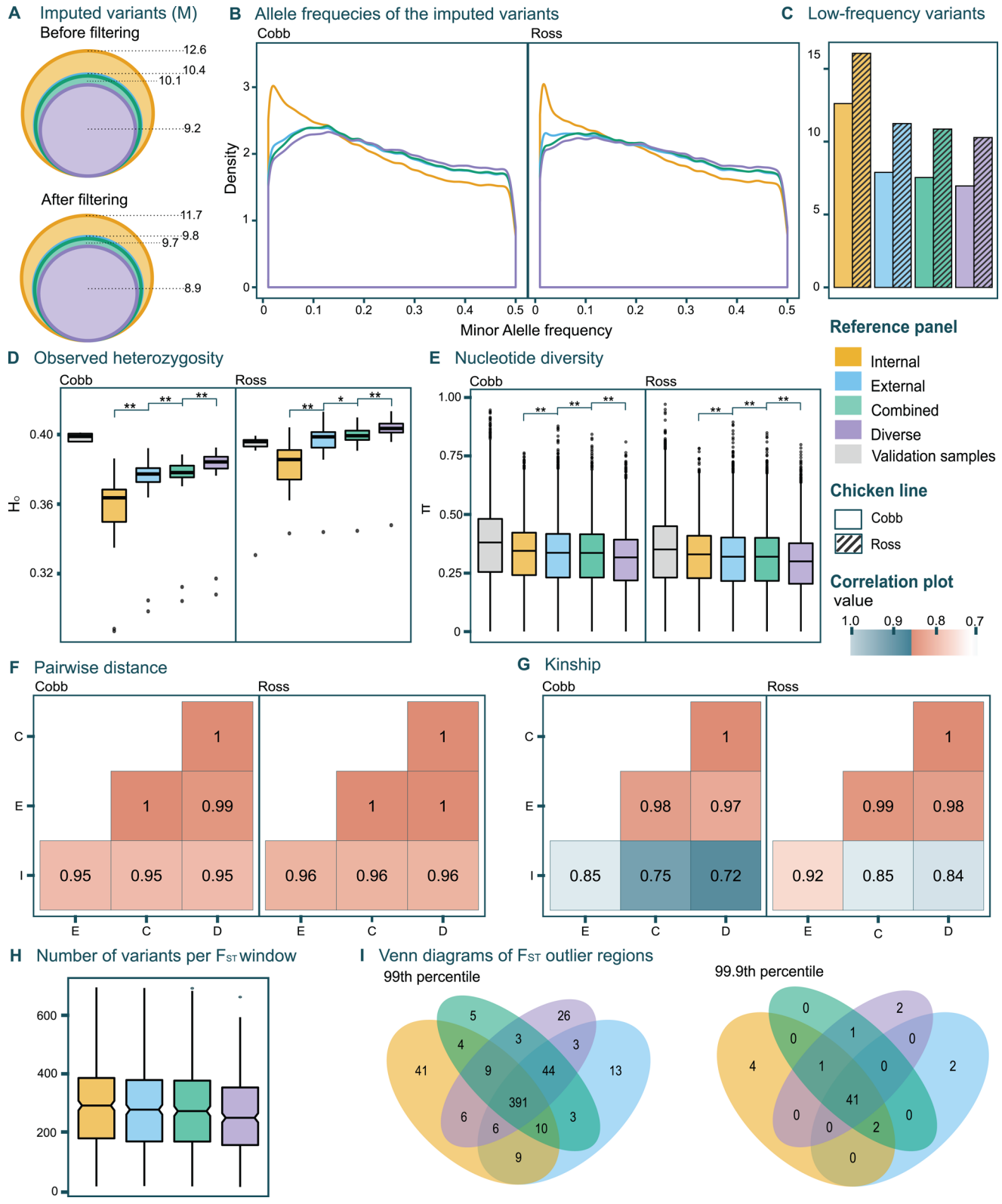


Figure 5. Comparison of the choice of reference panels for imputed target population for all autosomal chromosomes. A) Number of variants in the target population when imputed using the different panels. B) Minor allele frequency spectrum of the imputed variants in Cobb and Ross populations. C) Percentage of variants with a MAF lower than 0.05 by broiler line for all the panels. D) Observed heterozygosity for the 10 validation samples (true genotypes) and for the imputed target population by chicken broiler line (Cobb and Ross). Capitalized letters in the legend refer to the following names: I = internal, E = external, C = combined D = diverse, and V = validation samples. E) Nucleotide diversity of the target population by chicken broiler

(Figure 5B), as seen in the lower effective imputation of low-frequency variants with the internal panel (Figure 4A). Both Cobb and Ross populations showed similar allele frequency profiles, with a high proportion of intermediate MAF (Figure 5B) revealing a substantial loss of rare alleles in the respective populations.

3.3.2. Population Genetic Inference of the Target Population

Mean Ho values differed across all panels for both Cobb and Ross (Figure 5D). The values estimated by imputation tended to increase with panel size and diversity for both broiler lines. Individual Ho values displayed a higher variance when imputed with the internal panel and tended to equalize across samples with the rest of the panels, following the same trend as with the accuracy statistics (Figures 3 and 5D). This high variance displayed by the internal panel might stem from the fewer correctly imputed variants. For the Cobb population, none of the panels reached the Ho values seen with the 4 Cobb individuals (from the high-depth validation samples) (Figure 5D). For Ross, on the contrary, the external, and combined panels showed very similar values to the validation samples, while the diverse panel overestimated them. The very same trend could be seen when comparing imputed and high-depth validation samples (Figure S3, Supporting Information). There were some outlier samples (two from Cobb and one from Ross) that presented lower Ho than the high-depth validation samples (Figure 5D). These samples apparently underwent an incorrect imputation process, but it was not necessarily related to a low mapping depth.

Nucleotide diversity on the other side, decreased with increasing panel size and diversity (Figure 5E), which was directly related to the lower number of variants retained in the rest of the panels compared to the internal. There were significant differences in means except between the external and combined panels for both genetic lines, most likely because of the similar number of variants both panels share (Figure 5A). When comparing the imputed population with the validation samples, π of imputed samples and of the target population were underestimated for all panels (Figure 5E and Figure S3, Supporting Information).

Cobb and Ross populations were very similar, but the imputation tended to accentuate differences between both (Figure 5, Figures S4 and S5, Supporting Information). Within population, pairwise distance and kinship estimates did not vary much according to the panel. For pairwise distance, the diverse panel resulted in larger interindividual distances within genetic lines (Figure S4, Supporting Information). Kinship estimates were lower when computed with the internal panel, since a larger number of SNPs were retained, in particular, low-frequency variants, which are typically unique to one or few individuals, thus decreasing kinship (Figure S5, Supporting Information). Mantel tests did not show any significant differences for pairwise distance and kinship matrices, giving the same result for all panel comparisons (Mantel statistic, p-value < 0.001). Correlation values for pairwise distance were very similar and close to 1 (Figure 5F), even for the validation samples (true genotypes) when compared with any panel (Figure S3, Supporting Information). For kinship instead, it seemed that the internal panel differed more from the rest (Figure 5G and Figure S3, Supporting Information). In both cases, the 10 validation samples were most correlated with samples imputed with the combined and diverse panels (Figure S3, Supporting Information).

Whole-genome mean F_{ST} values between Cobb and Ross populations were very similar (I = 0.071, E = 0.071, C = 0.072 and D = 0.072) indicating overall low differentiation between the genetic lines. When analyzing the putative selective sweep regions using as threshold the 99th percentile, 68.2% of the windows coincided across the four panels, but more interestingly, 75.9% of the windows were shared across the external, combined, and diverse panels. When we raised the threshold to the 99.9th percentile, 77% of windows were identified by the genome-scans regardless of the choice of panel, indicating that the strongest signals are detected with any panel. Yet, there were some regions that only passed the threshold when imputation was performed with a particular panel (Figure 5I). The combined panel did not show specific sweeps when the percentile was set at 99.9, and it was the panel with the lowest panel-specific regions with the 99th percentile as well, potentially indicating the most robust results, that is, without panel-specific biases. Surprisingly, the diverse panel detected the most panel-specific sweeps after the internal panel (Figure 5I). On the other side, in terms of density of variants in the common windows, the mean number of variants reduced significantly from the internal to the diverse panel (Figure 5H). This suggests that although in a broad sense the same outlier F_{ST} regions tend to be recovered by all panels, a reduced number of imputed variants might decrease the probability of detecting true outliers.

4. Discussion

Shotgun metagenomic datasets of host-associated microbial communities often contain host DNA that is usually discarded because the amount of data is too low for accurate host genetic analyses. Here, we introduced an effective and accurate approach to recover high-quality host genomes from gut metagenomic data, which can be used to study host population genetic analyses and ultimately contribute to a better understanding of host-microbiota interactions.

Our analyses yielded drastic differences in mapping statistics between cecum samples used to characterize the target population and ileum samples employed to generate the internal panel. Although both sample types were derived from gut contents, the cecum harbored a very small amount of the host DNA compared to the ileum, because the latter is known to contain fewer bacteria,[45] and a higher permeability and thinner mucus layer of the ileum probably entails higher release of epithelial cells to

line. Paired *T*-tests were performed to identify significant differences in means: the following symbols ("**," "*") indicate different p-value cut-points (<0.001, 0.05). F) Kinship and G) pairwise distance correlation matrices for the target population. Capitalized letters in the *x*– and *y*–axes refer to panel names: I = internal, E = external, C = combined and D = diverse. H) Boxplot showing number of variants in the common windows of the 99th percentile from the F_{ST} genome scan. I) Venn diagram depicting overlap of significantly differentiated windows as estimated by F_{ST} genome scans between Cobb and Ross populations using the different panels for imputation. Significance thresholds were set at the 99th and 99.9th percentiles.

the lumen.[46] Moreover, the low, yet variable, proportion of host DNA retrieved from cecum samples renders sequencing depth adjustment highly unpredictable, as previously reported.[6] Nevertheless, we showed that if a proper reference panel is designed, the low and variable fractions of host DNA recovered from such suboptimal samples can be used for accurately inferring host genetic features.

4.1. Adjustment of the Imputation Strategy for Metagenomic Samples

The two-step imputation strategy performed efficiently despite the structural (e.g., study design, animal taxa, reference panel size) differences between our study system and the ones the strategy was originally designed for.[21,22] First, the data pre-processing steps were adapted to the characteristics of the metagenomic samples. As metagenomic data contain multi-species sequences, mapping seed length was increased in order to increase mapping specificity,[29] although there are also examples where they have set standard parameters for the alignment.[2,47] With standard (19) and increased (25) seed lengths, mapping gaps across the reference genome were unevenly distributed. This is evidenced by the large difference between depth (K19 = 2.78x, K25 = 1.8x) and breadth (K19 = 57%, K25 = 50%) of coverage (Figure S1 & Table S1, Supporting Information), likely hampering accurate computation across the genome. However, in our study, accuracy only dropped significantly in samples below 0.1×, while more than 0.90 of the variants were recovered. Similarly, Hui et al. reported that the proportion of correctly imputed heterozygous sites started decreasing at 0.5× of depth of coverage, reaching 50% of correctly imputed sites at 0.1×.[22] Homburger et al. also reported that imputation performance decreased for samples below 0.5× of coverage, being imputation r2 = 0.90 for samples with 0.5× of coverage.[21] Regarding the generation of genotype likelihoods, results should be minimally affected by the choice of the variant caller.[22] We also verified that the DNA obtained from cecum samples was not partially degraded (i.e., short read length, deamination in last bases of the read). Misincorporations at the 5′ (C to T) and 3′ (G to A) ends due to deamination were less than 0.6% (mean = 0.4 ± 0.03%), and read length was consistent across samples (150 bp) (Tables S2 and S3, Supporting Information), thus, we did not apply a deamination filter.

Second, we used custom reference panels with less than one hundred individuals, while the two-step strategy was originally tested with a large human reference panel (i.e., 1000G).[22] Nevertheless, the accuracy of imputed low-frequency variants for all panels was comparable to Hui et al., most likely because the individuals in our target population were closely related as evidenced by the high kinship estimates, and the stringent genotype filtering (MCR = 0) we used for generating the custom panels in order to reduce the error rate.

Finally, unlike humans, avian genomes present macro- and micro-chromosomes, and the latter show higher interchromosomal interactions and recombination rates.[48] However, it seems that the possible crossovers did not affect imputation in contrast to previous studies,[49,50] since we did not find any significant differences in imputation accuracies between chromosomes. This suggests that the strategy worked equally well for large, mid-sized, and small chromosomes with potentially different linkage patterns.

4.2. Effect of Reference Panel on Accuracy Statistics

Reference panel design depends on data availability as well as computational capacity. It is a common strategy for imputation of inbred populations to resequence a subset of samples with higher resolution in order to optimize imputation performance.[35] Based on previous works, we estimated that 12 individuals out of 100 would be sufficient to represent the genetic diversity of the population. For instance, previous chicken studies deep-sequenced 25 individuals to impute ≈450 chickens genotyped with 600 K SNP arrays (≈5% of sample size).[18,51]

In terms of the panels' SNP density, we decided to genotype variants that did appear in our target population rather than calling specific variants in the rest of the breeds that composed the reference panels. Thereby, we aimed at reducing the noise that the excess variant density could cause in the imputation process. Nevertheless, as the genetic distance between the reference chicken populations and our broilers is very small,[52] we expected them to share many variants, as we evidenced in preliminary analyses (Figure S2, Supporting Information).

The internal panel resulted in a larger variance across samples for overall accuracy statistics. In addition, SNPs with low MAF had the lowest accuracy when imputed with the internal panel, but were also the poorest imputed across panels. However, incorrectly imputed low-frequency variants can be easily removed if a strict MAF filter is applied for downstream analysis. Another possible option is to sequence more individuals of the target population to increase the reference panel size. Hence, despite the internal panel only representing a small subset of the target population, and showing lower imputation values than the rest of the panels, for scientists without access to external reference samples, this approach is equally useful as overall imputation accuracy was higher than 0.90. In this sense, host resequencing of a small subset of the target population might represent a cost-efficient option, especially for researchers working with non-model organisms and inbred populations.

Our results showed that the combined panel performed better in terms of overall accuracy, and - specifically for MAF variants - than the internal and external panels alone. Despite the fact that the external and combined panels had the same number of SNPs, including a subset of individuals from the target population was beneficial. Many studies already mentioned an improvement for the combined option.[53,54] Last, the diverse panel showed the highest values of concordance and precision of het. sites, most probably because of the lower number of SNPs recovered, especially low-frequency variants, which generally yielded lower imputation accuracies. In terms of imputation of low-frequency variants, the combined panel outperformed the diverse one, that is, it correctly imputed a larger number of variants and tended to improve the precision of het. sites in some MAF bins. A recent large-scale study performed in a Han Chinese population showed that a Chinese-specific reference panel worked better than frequently used reference panels such as 1000G.[19] Imputation was greatly improved when the reference panel contained a fraction

Table 1. Metagenomic datasets and available reference panels. Percentage of host DNA was calculated for the studies that did not mention host mapping percentage. Alignment with standard parameters was performed with 10 samples (marked with "*").

System	Pop. characteristics	N	Sample types	Host DNA	Metagenomic dataset	Ref. panel	Downstream analysis
Buffalo	River, swamp, and hybrid buffaloes	695	Gut, intestine, and rectum	<20%	[65]	[66]	Selection signatures
Cattle	Three crossbreeds and one pureline	282	Gut	*3%	[67]	[11]	Selection signatures
Pig	Various breeds	287	Fecal	2%	[68]	[69]	Selection signatures
Pig	Various breeds	470	Fecal	*2%	[70]	[69]	Selection signatures
Chicken	Lohmann Brown and Silkie hens	90	Fecal	8%	[71]	Custom[57]	Selection signatures
Chicken	Red Junglefowl	51	Fecal	49%	[72]	Custom[57]	Implication on domestication
Rat	Sprague Dawley	49	Fecal	11%	[73]	Custom[74]	Host-microbiota association
Rat	SpragueDawley	84	Cecal	*51%	[75]	Custom[74]	Host-microbiota interactions
Mouse	Various breeds	184	Fecal	9%	[76]	Custom[77]	Differences between populations
Mouse	C57BL/6J	88	Fecal	<5%	[78]	Custom[77]	Differences between populations
Zebrafish	Single cohort	29	Fecal	*9%	[79]	Custom[80]	Population genetic inference
Honey bee	Eastern and Western honey bees	40	Gut	<10%	[81]	[82]	Differences between species

of an extra diverse sample, but they obtained a different pattern when the panel size was fixed.[19] Thus, taking into consideration our and previous results on selection of imputation panels, it can be concluded that increasing panel size and diversity improves imputation, but a balance has to be found in the composition of the panel. The distance between the panel and the target population has to be taken into account.

4.3. Effect of Reference Panel on Population Genetic Inference

Besides crude imputation accuracy statistics, we evaluated the impact of the panels on downstream population genetic statistics and their biological interpretation. As imputation accuracies were generally high with our applied pipeline and the stringent filtering approach, we expected population genetic inferences to follow similarly.

Although overall results were in agreement, all the tested parameters showed slight trends according to the used reference panel. Observed heterozygosity, pairwise distance, and kinship values increased while mean F_{ST} and π values decreased with panel size and diversity (Figure 5, Figures S4and S5, Supporting Information). Such biases were related to the composition of the panels and the associated number and distribution of recovered SNPs.

Imputation performance was slightly different for the two broiler lines, as Ross population estimations were closer to the true values than for the Cobb population. Thus, accentuating the distance between both genetic lines. This is most likely due to a smaller representation of Cobb individuals in the reference panels, that is, 5 Cobb and 7 Ross samples constituted the internal panel. Second, there were some samples that were incorrectly imputed because of their low Ho values (Figure 5D). We do not know if there are individuals with lower Ho in our Cobb and Ross populations. For instance, a Ross individual from the high-depth validation samples had considerably lower Ho than the rest of Ross individuals. Each broiler line came from two different hatcheries, which might be the reason why some individuals might have slightly different genetic features. We may have under-represented one of the origins in the internal reference samples. Thus, it is necessary to be more cautious for the interpretation of individual genomes. Nevertheless, results appeared to be robust and similar across panels at the population level. The genome scans yielded overall very consistent results with major differentiation signals identified by any of the imputed datasets, likely indicative of a true selection signature between both lines. However, downstream analyses, such genome scans, and GWAS must be performed with caution since imputation is sensitive to low-frequency variant quality.

Both broiler lines exhibited high density of intermediate-frequency variants, with similar allele frequency distributions to previously described commercial breed populations.[25,55] A high density of intermediate alleles is indicative of genetic drift due to selection in a closed breeding population.[55] Domestication and breeding history are the two major processes that shape haplotype structure.[25,56] Cobb and Ross, together with other commercial lines, have much smaller effective population sizes than other chicken populations.[57] Broiler breeding methods

are described as a pyramid strategy, in which pure, inbred lines are crossed, then F_1 individuals are crossed between each other. In some cases, even a second or a third cross is performed in F_2 and F_3 generations before raising them for meat.[58,59] Therefore, broilers are highly related populations, but at the same time present high Ho values. Ho of our studied broiler lines was much higher than of local populations,[60] but similar to other commercial lines.[56] Similarly, nucleotide diversity and mean fixation index values were comparable to those previously reported.[25]

4.4. Potential Applications

The possibility to retrieve genomic data from metagenomic samples can help reanalyze already published data, and reduce resources spent on population genetic studies. For instance, our approach could be useful to study genome features of endangered populations relying on fecal samples recovered from the environment.[61,62] In an attempt to elucidate the possible applications of the here presented imputation strategy, examples of published metagenomic datasets with potential haplotype reference panels are provided (**Table 1**). As mentioned above, host DNA percentage varies depending on the study system, sample type, and sequencing effort.[63] Some studies collect samples from multiple body sites from each individual,[2,64] and if combined, the amount of host DNA can be greatly increased.[3] Or as in our case, samples with high host DNA can serve as individuals for the reference panel to impute low-coverage samples. Imputation can also be useful for making preliminary explorations of host genome characteristics, while higher quality samples are being processed or sequenced.

The implemented strategy is rather dependent on the availability of a reference genome and high-quality genomic data. Global efforts such as The Vertebrate Genome Project,[83] European Reference Genome Atlas,[84,85] initiatives are contributing to study an increasing number of species, and thereby providing useful reference resources for the scientific community. Likewise, phased haplotype panels are being generated for various species.[69,82] When generating a custom panel in the absence of a publicly available option, the genetic diversity and sample size of the study population have to be taken into account. Larger reference panels are needed with more diverse and heterogeneous populations,[16] while in more isolated populations, reference panels that include population-specific individuals can improve imputation of rare alleles.[19,86,87]

5. Conclusion

Our results show that the two-step imputation implemented in this study can be used to successfully reconstruct genotypes and study population genetic properties of hosts from suboptimal metagenomic samples. The comparison among reference panels also demonstrated that this method is versatile and flexible. This approach could be used in many contexts and exploit different data sources to address a variety of research questions. This includes the possibility of mining published metagenomic datasets to recover discarded host DNA sequences. In our particular case, the reconstructed genotypes will be employed in the H2020 project HoloFood to detect interactions with microbial metagenomic features, and thus implement a hologenomic approach to improve animal production.[88]

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

This research was funded by the European Union's Horizon Research and Innovation Programme under grant agreement No. 817729 (HoloFood, Holistic solution to improve animal food production through deconstructing the biomolecular interactions between feed, gut microorganisms, and animals in relation to performance parameters). The work of S.M. was supported by the Basque Government doctoral fellowship. The authors would like to thank the rest of the HoloFood partners that were involved in the design and execution of the animal trials, specially colleagues Joan Tarradas, Nuria Tous, and Enric Esteve from IRTA.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

A.E. and A.A. conceived and designed the analyses. S.M. collected the data. S.M. and M.P. performed the analysis. S.M. wrote the original draft. All authors contributed significantly to the review and editing of the original draft.

Data Availability Statement

Data used in this study is part of the European H2020 reserach project Holofood No. 817729 and will be released upon completion at the end of 2022. Data will be available upon request from the authors until such time.

Peer Review

The peer review history for this article is available in the Supporting Information for this article.

Keywords

host, imputation, metagenomic data, population genetics

Received: November 27, 2021
Revised: March 5, 2022
Published online:

[1] S. Yang, X. Gao, J. Meng, A. Zhang, Y. Zhou, M. Long, B. Li, W. Deng, L. Jin, S. Zhao, D. Wu, Y. He, C. Li, S. Liu, Y. Huang, H. Zhang, L. Zou, *Front. Microbiol.* **2018**, *9*, 1717.
[2] R. Blekhman, J. K. Goodrich, K. Huang, Q. Sun, R. Bukowski, J. T. Bell, T. D. Spector, A. Keinan, R. E. Ley, D. Gevers, A. G. Clark, *Genome Biol.* **2015**, *16*, 191.

[3] L. Nyholm, A. Koziol, S. Marcos, A. B. Botnen, O. Aizpurua, S. Gopalakrishnan, M. T. Limborg, M. T. P. Gilbert, A. Alberdi, *iScience* **2020**, *23*, 101414.
[4] M. T. Limborg, A. Alberdi, M. Kodama, M. Roggenbuck, K. Kristiansen, M. T. P. Gilbert, *Trends Biotechnol.* **2018**, *36*, 252.
[5] J. A. Rasmussen, K. R. Villumsen, D. A. Duchêne, L. C. Puetz, T. O. Delmont, H. Sveier, L. von Gersdorff Jørgensen, K. Præbel, M. D. Martin, A. M. Bojesen, M. T. P. Gilbert, K. Kristiansen, M. T. Limborg, *Commun. Biol.* **2021**, *4*, 579.
[6] A. Alberdi, O. Aizpurua, M. T. P. Gilbert, K. Bohmann, *Methods Ecol. Evol.* **2018**, *9*, 134.
[7] J. Marchini, B. Howie, *Nat. Rev. Genet.* **2010**, *11*, 499.
[8] T. Iso-Touru, G. Sahana, B. Guldbrandtsen, M. S. Lund, J. Vilkki, *BMC Genet.* **2016**, *17*, 55.
[9] F. Pértille, G. C. M. Moreira, R. Zanella, J. de R da, S. Nunes, C. Boschiero, G. A. Rovadoscki, G. B. Mourão, M. C. Ledur, L. L. Coutinho, *Sci. Rep.* **2017**, *7*, 41748.
[10] 1000 Genomes Project Consortium, G. R. Abecasis, D. Altshuler, A. Auton, L. D. Brooks, R. M. Durbin, R. A. Gibbs, M. E. Hurles, G. A. McVean, *Nature* **2010**, *467*, 1061.
[11] H. D. Daetwyler, A. Capitan, H. Pausch, P. Stothard, R. van Binsbergen, R. F. Brøndum, X. Liao, A. Djari, S. C. Rodriguez, C. Grohs, D. Esquerré, O. Bouchez, M.-N. Rossignol, C. Klopp, D. Rocha, S. Fritz, A. Eggen, P. J. Bowman, D. Coote, A. J. Chamberlain, C. Anderson, C. P. VanTassell, I. Hulsegge, M. E. Goddard, B. Guldbrandtsen, M. S. Lund, R. F. Veerkamp, D. A. Boichard, R. Fries, B. J. Hayes, *Nat. Genet.* **2014**, *46*, 858.
[12] M. S. Artigas, L. V. Wain, S. Miller, A. K. Kheirallah, J. E. Huffman, I. Ntalla, N. Shrine, M. Obeidat, H. Trochet, W. L. McArdle, A. C. Alves, J. Hui, J. H. Zhao, P. K. Joshi, A. Teumer, E. Albrecht, M. Imboden, R. Rawal, L. M. Lopez, J. Marten, S. Enroth, I. Surakka, O. Polasek, L.-P. Lyytikäinen, R. Granell, P. G. Hysi, C. Flexeder, A. Mahajan, J. Beilby, Y. Bossé, et al., *Nat. Commun.* **2015**, *6*, 8658.
[13] H. Pausch, I. M. MacLeod, R. Fries, R. Emmerling, P. J. Bowman, H. D. Daetwyler, M. E. Goddard, *Genet., Sel., Evol.* **2017**, *49*, 24.
[14] G. Pistis, E. Porcu, S. I. Vrieze, C. Sidore, M. Steri, F. Danjou, F. Busonero, A. Mulas, M. Zoledziewska, A. Maschio, C. Brennan, S. Lai, M. B. Miller, M. Marcelli, M. F. Urru, M. Pitzalis, R. H. Lyons, H. M. Kang, C. M. Jones, A. Angius, W. G. Iacono, D. Schlessinger, M. McGue, F. Cucca, G. R. Abecasis, S. Sanna, *Eur. J. Hum. Genet.* **2015**, *23*, 975.
[15] Q. Duan, E. Y. Liu, P. L. Auer, G. Zhang, E. M. Lange, G. Jun, C. Bizon, S. Jiao, S. Buyske, N. Franceschini, C. S. Carlson, L. Hsu, A. P. Reiner, U. Peters, J. Haessler, K. Curtis, C. L. Wassel, J. G. Robinson, L. W. Martin, C. A. Haiman, L. Le Marchand, T. C. Matise, L. A. Hindorff, D. C. Crawford, T. L. Assimes, H. M. Kang, G. Heiss, R. D. Jackson, C. Kooperberg, J. G. Wilson, et al., *Bioinformatics* **2013**, *29*, 2744.
[16] M. Al Kalaldeh, J. Gibson, N. Duijvesteijn, H. D. Daetwyler, I. MacLeod, N. Moghaddar, S. H. Lee, J. H. J. van der Werf, *Genet., Sel., Evol.* **2019**, *51*, 32.
[17] S. van den Berg, J. Vandenplas, F. A. van Eeuwijk, A. C. Bouwman, M. S. Lopes, R. F. Veerkamp, *Genet., Sel., Evol.* **2019**, *51*, 2.
[18] S. Huang, Y. He, S. Ye, J. Wang, X. Yuan, H. Zhang, J. Li, X. Zhang, Z. Zhang, *J. Appl. Genet.* **2018**, *59*, 335.
[19] W.-Y. Bai, X.-W. Zhu, P.-K. Cong, X.-J. Zhang, J. B. Richards, H.-F. Zheng, *Briefings Bioinf.* **2020**, *21*, 1806.
[20] P. Korkuć, D. Arends, G. A. Brockmann, *Front. Genet.* **2019**, *10*, 52.
[21] J. R. Homburger, C. L. Neben, G. Mishne, A. Y. Zhou, S. Kathiresan, A. V. Khera, *Genome Med.* **2019**, *11*, 74.
[22] R. Hui, E. D'Atanasio, L. M. Cassidy, C. L. Scheib, T. Kivisild, *Sci. Rep.* **2020**, *10*, 18542.
[23] D. Bozzi, J. A. Rasmussen, C. Carøe, H. Sveier, K. Nordøy, M. T. P. Gilbert, M. T. Limborg, *Anim. Microbiome* **2021**, *3*, 30.
[24] C. Carøe, S. Gopalakrishnan, L. Vinner, S. S. T. Mak, M. H. S. Sinding, J. A. Samaniego, N. Wales, T. Sicheritz-Pontén, M. T. P. Gilbert, *Methods Ecol. Evol.* **2018**, *9*, 410.
[25] S. Qanbari, C.-J. Rubin, K. Maqbool, S. Weigend, A. Weigend, J. Geibel, S. Kerje, C. Wurmser, A. T. Peterson, I. L. Brisbin Jr, R. Preisinger, R. Fries, H. Simianer, L. Andersson, *PLoS Genet.* **2019**, *15*, e1007989.
[26] M. Schubert, S. Lindgreen, L. Orlando, *BMC Res. Notes* **2016**, *9*, 88.
[27] W. Shen, S. Le, Y. Li, F. Hu, *PLoS One* **2016**, *11*, e0163962.
[28] H. Li, R. Durbin, *Bioinformatics* **2009**, *25*, 1754.
[29] K. M. Robinson, A. S. Hawkins, I. Santana-Cruz, R. S. Adkins, A. C. Shetty, S. Nagaraj, L. Sadzewicz, L. J. Tallon, D. A. Rasko, C. M. Fraser, A. Mahurkar, J. C. Silva, J. C. Dunning Hotopp, *Microb. Genomics* **2017**, *3*, e000122.
[30] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, 1000 Genome Project Data Processing Subgroup, *Bioinformatics* **2009**, *25*, 2078.
[31] J. Neukamm, A. Peltzer, K. Nieselt, *Bioinformatics* **2021**, *37*, 3652.
[32] H. Li, *Bioinformatics* **2011**, *27*, 2987.
[33] A. McKenna, M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, M. A. DePristo, *Genome Res.* **2010**, *20*, 1297.
[34] O. Delaneau, J. Marchini, J.-F. Zagury, *Nat. Methods* **2011**, *9*, 179.
[35] B. L. Browning, S. R. Browning, *Am. J. Hum. Genet.* **2016**, *98*, 116.
[36] C. C. Chang, C. C. Chow, L. C. Tellier, S. Vattikuti, S. M. Purcell, J. J. Lee, *GigaScience* **2015**, *4*, 7.
[37] M. Nei, *Prog. Clin. Biol. Res.* **1982**, *103 Pt A*, 167.
[38] A. Manichaikul, J. C. Mychaleckyj, S. S. Rich, K. Daly, M. Sale, W.-M. Chen, *Bioinformatics* **2010**, *26*, 2867.
[39] B. S. Weir, C. C. Cockerham, *Evolution* **1984**, *38*, 1358.
[40] S. Wilkinson, Z. H. Lu, H.-J. Megens, A. L. Archibald, C. Haley, I. J. Jackson, M. A. M. Groenen, R. P. M. A. Crooijmans, R. Ogden, P. Wiener, *PLoS Genet.* **2013**, *9*, e1003453.
[41] W. H. Kruskal, W. A. Wallis, *J. Am. Stat. Assoc.* **1952**, *47*, 583.
[42] Student, *Biometrika* **1908**, *6*, 1.
[43] H. Abdi, *Others, Encyclopedia of measurement and statistics* **2007**, *3*, 103.
[44] S. Dray, A.-B. Dufour, *J. Stat. Software* **2007**, *22*, 1.
[45] I. Rychlik, *Animals* **2020**, *10*, 103.
[46] Y. Duangnumsawang, J. Zentek, F. Goodarzi Boroojeni, *Front. Immunol.* **2021**, *12*, 3924.
[47] T. Regan, M. W. Barnett, D. R. Laetsch, S. J. Bush, D. Wragg, G. E. Budge, F. Highet, B. Dainat, J. R. de Miranda, M. Watson, M. Blaxter, T. C. Freeman, *Nat. Commun.* **2018**, *9*, 4995.
[48] B. W. Perry, D. R. Schield, R. H. Adams, T. A. Castoe, *Mol. Biol. Evol.* **2021**, *38*, 904.
[49] D. B. Hancock, J. L. Levy, N. C. Gaddis, L. J. Bierut, N. L. Saccone, G. P. Page, E. O. Johnson, *PLoS One* **2012**, *7*, e50610.
[50] G. Ni, T. M. Strom, H. Pausch, C. Reimer, R. Preisinger, H. Simianer, M. Erbe, *BMC Genomics* **2015**, *16*, 824.
[51] S. Ye, X. Yuan, X. Lin, N. Gao, Y. Luo, Z. Chen, J. Li, X. Zhang, Z. Zhang, *J. Anim. Sci. Biotechnol.* **2018**, *9*, 30.
[52] S. Qanbari, H. Simianer, *Livest. Sci.* **2014**, *166*, 133.
[53] S. Ye, X. Yuan, S. Huang, H. Zhang, Z. Chen, J. Li, X. Zhang, Z. Zhang, *Animal* **2019**, *13*, 1119.
[54] S. Ye, Z.-T. Chen, R. Zheng, S. Diao, J. Teng, X. Yuan, H. Zhang, Z. Chen, X. Zhang, J. Li, Z. Zhang, *Front. Genet.* **2020**, *11*, 243.
[55] S. Qanbari, *Front. Genet.* **2019**, *10*, 1304.
[56] R. Talebi, T. Szmatoła, G. Mészáros, S. Qanbari, *G3: Genes, Genomes, Genet.* **2020**, *10*, 4615.
[57] M.-S. Wang, J.-J. Zhang, X. Guo, M. Li, R. Meyer, H. Ashari, Z.-Q. Zheng, S. Wang, M.-S. Peng, Y. Jiang, M. Thakur, C. Suwannapoom, A. Esmailizadeh, N. Y. Hirimuthugoda, M. S. A. Zein, S. Kusza, H. Kharrati-Koopaee, L. Zeng, Y.-M. Wang, T.-T. Yin, M.-M. Yang, M.-L.

Li, X.-M. Lu, E. Lasagna, S. Ceccobelli, H. G. T. N. Gunwardana, T. M. Senasig, S.-H. Feng, H. Zhang, A. K. F. H. Bhuiyan, et al., *BMC Biol.* **2021**, *19*, 118.

[58] A. L. Van Eenennaam, K. A. Weigel, A. E. Young, M. A. Cleveland, J. C. M. Dekkers, *Annu. Rev. Anim. Biosci.* **2014**, *2*, 105.

[59] A. K. Thiruvenkadan, R. Prabakaran, S. Panneerselvam, *World's Poult. Sci. J.* **2011**, *67*, 309.

[60] D. K. Malomane, H. Simianer, A. Weigend, C. Reimer, A. O. Schmitt, S. Weigend, *BMC Genomics* **2019**, *20*, 345.

[61] A. Ang, D. I. Roesma, V. Nijman, R. Meier, A. Srivathsan, Rizaldi, *Sci. Rep.* **2020**, *10*, 9396.

[62] J. D. Negrey, M. E. Thompson, K. E. Langergraber, Z. P. Machanda, J. C. Mitani, M. N. Muller, E. Otali, L. A. Owens, R. W. Wrangham, T. L. Goldberg, *Philos. Trans. R. Soc., B* **2020**, *375*, 20190613.

[63] A. Alberdi, O. Aizpurua, K. Bohmann, S. Gopalakrishnan, C. Lynggaard, M. Nielsen, M. T. P. Gilbert, *Mol. Ecol. Resour.* **2019**, *19*, 327.

[64] Integrative HMP (iHMP) Research Network Consortium, *Nature* **2019**, *569*, 641.

[65] F. Tong, T. Wang, N. L. Gao, Z. Liu, K. Cui, Y. Duan, S. Wu, Y. Luo, Z. Li, C. Yang, Y. Xu, B. Lin, L. Yang, A. Pauciullo, D. Shi, G. Hua, W.-H. Chen, Q. Liu, *Nat. Commun.* **2022**, *13*, 823.

[66] X. Luo, Y. Zhou, B. Zhang, Y. Zhang, X. Wang, T. Feng, Z. Li, K. Cui, Z. Wang, C. Luo, H. Li, Y. Deng, F. Lu, J. Han, Y. Miao, H. Mao, X. Yi, C. Ai, S. Wu, A. Li, Z. Wu, Z. Zhuo, D. Da Giang, B. Mitra, M. F. Vahidi, S. Mansoor, S. A. Al-Bayatti, E. M. Sari, N. A. Gorkhali, S. Prastowo, et al.,*Natl. Sci. Rev.* **2020**, *7*, 686.

[67] R. D. Stewart, M. D. Auffret, A. Warr, A. W. Walker, R. Roehe, M. Watson, *Nat. Biotechnol.* **2019**, *37*, 953.

[68] L. Xiao, J. Estellé, P. Kiilerich, Y. Ramayo-Caldas, Z. Xia, Q. Feng, S. Liang, A. Ø. Pedersen, N. J. Kjeldsen, C. Liu, E. Maguin, J. Doré, N. Pons, E. Le Chatelier, E. Prifti, J. Li, H. Jia, X. Liu, X. Xu, S. D. Ehrlich, L. Madsen, K. Kristiansen, C. Rogel-Gaillard, J. Wang, *Nat. Microbiol.* **2016**, *1*, 16161.

[69] Z. Zhang, P. Ma, Z. Zhang, Z. Wang, Q. Wang, Y. Pan, *Genomics* **2022**, *114*, 340.

[70] C. Chen, Y. Zhou, H. Fu, X. Xiong, S. Fang, H. Jiang, J. Wu, H. Yang, J. Gao, L. Huang, *Nat. Commun.* **2021**, *12*, 1106.

[71] R. Gilroy, A. Ravi, M. Getino, I. Pursley, D. L. Horton, N.-F. Alikhan, D. Baker, K. Gharbi, N. Hall, M. Watson, E. M. Adriaenssens, E. Foster-Nyarko, S. Jarju, A. Secka, M. Antonio, A. Oren, R. R. Chaudhuri, R. La Ragione, F. Hildebrand, M. J. Pallen, *PeerJ* **2021**, *9*, e10941.

[72] L. C. Puetz, T. O. Delmont, O. Aizpurua, C. Guo, G. Zhang, R. Katajamaa, P. Jensen, M. T. P. Gilbert, *Adv. Genet.* **2021**, *2*, 2100018.

[73] H. Pan, R. Guo, J. Zhu, Q. Wang, Y. Ju, Y. Xie, Y. Zheng, Z. Wang, T. Li, Z. Liu, L. Lu, F. Li, B. Tong, L. Xiao, X. Xu, R. Li, Z. Yuan, H. Yang, J. Wang, K. Kristiansen, H. Jia, L. Liu, *GigaScience* **2018**, *7*, giy055.

[74] A. F. Gileta, C. J. Fitzpatrick, A. S. Chitre, C. L. St Pierre, E. V. Joyce, R. J. Maguire, A. M. McLeod, N. M. Gonzales, A. E. Williams, J. D. Morrow, T. E. Robinson, S. B. Flagel, A. A. Palmer, *bioRxiv 412924* **2021**.

[75] R. Mesnage, M. Teixeira, D. Mandrioli, L. Falcioni, Q. R. Ducarmon, R. D. Zwittink, F. Mazzacuva, A. Caldwell, J. Halket, C. Amiel, J.-M. Panoff, F. Belpoggi, M. N. Antoniou, *Environ. Health Perspect.* **2021**, *129*, 17005.

[76] L. Xiao, Q. Feng, S. Liang, S. B. Sonne, Z. Xia, X. Qiu, X. Li, H. Long, J. Zhang, D. Zhang, C. Liu, Z. Fang, J. Chou, J. Glanville, Q. Hao, D. Kotowska, C. Colding, T. R. Licht, D. Wu, J. Yu, J. J. Y. Sung, Q. Liang, J. Li, H. Jia, Z. Lan, V. Tremaroli, P. Dworzynski, H. B. Nielsen, F. Bäckhed, J. Doré, et al., *Nat. Biotechnol.* **2015**, *33*, 1103.

[77] K. Fujiwara, Y. Kawai, K. Moriwaki, T. Takada, T. Shiroishi, N. Saitou, H. Suzuki, N. Osada, *bioRxiv 2021.02.05.429881* **2021**.

[78] J. Zhu, H. Ren, H. Zhong, X. Li, Y. Zou, M. Han, M. Li, L. Madsen, K. Kristiansen, L. Xiao, *mSphere* **2021**, *6*, e01119.

[79] C. A. Gaulke, L. M. Beaver, C. R. Armour, I. R. Humphreys, C. L. Barton, R. L. Tanguay, E. Ho, T. J. Sharpton, *bioRxiv 2020.06.15.153320* **2020**.

[80] M. Balik-Meisner, L. Truong, E. H. Scholl, R. L. Tanguay, D. M. Reif, *Mamm. Genome* **2018**, *29*, 90.

[81] K. M. Ellegaard, S. Suenami, R. Miyazaki, P. Engel, *Curr. Biol.* **2020**, *30*, 2520.

[82] D. Wragg, S. E. Eynard, B. Basso, K. Canale-Tabet, E. Labarthe, O. Bouchez, K. Bienefeld, M. Bieńkowska, C. Costa, A. Gregorc, P. Kryger, M. Parejo, M. Alice Pinto, J.-P. Bidanel, B. Servin, Y. Le Conte, A. Vignal, *bioRxiv 2021.09.20.460798* **2021**.

[83] A. Rhie, S. A. McCarthy, O. Fedrigo, J. Damas, G. Formenti, S. Koren, M. Uliano-Silva, W. Chow, A. Fungtammasan, J. Kim, C. Lee, B. J. Ko, M. Chaisson, G. L. Gedman, L. J. Cantin, F. Thibaud-Nissen, L. Haggerty, I. Bista, M. Smith, B. Haase, J. Mountcastle, S. Winkler, S. Paez, J. Howard, S. C. Vernes, T. M. Lama, F. Grutzner, W. C. Warren, C. N. Balakrishnan, D. Burt, et al., *Nature* **2021**, *592*, 737.

[84] G. Formenti, K. Theissinger, C. Fernandes, I. Bista, A. Bombarely, C. Bleidorn, C. Ciofi, A. Crottini, J. A. Godoy, J. Höglund, J. Malukiewicz, A. Mouton, R. A. Oomen, S. Paez, P. J. Palsbøll, C. Pampoulie, M. J. Ruiz-López, H. Svardal, C. Theofanopoulou, J. de Vries, A.-M. Waldvogel, G. Zhang, C. J. Mazzoni, E. D. Jarvis, M. Bálint, European Reference Genome Atlas (ERGA) Consortium, *Trends Ecol. Evol.* **2022**, *37*, 197.

[85] H. A. Lewin, S. Richards, E. Lieberman Aiden, M. L. Allende, J. M. Archibald, M. Bálint, K. B. Barker, B. Baumgartner, K. Belov, G. Bertorelle, M. L. Blaxter, J. Cai, N. D. Caperello, K. Carlson, J. C. Castilla-Rubio, S.-M. Chaw, L. Chen, A. K. Childers, J. A. Coddington, D. A. Conde, M. Corominas, K. A. Crandall, A. J. Crawford, F. DiPalma, R. Durbin, T. E. Ebenezer, S. V. Edwards, O. Fedrigo, P. Flicek, G. Formenti, et al., *Proc. Natl. Acad. Sci. USA* **2022**, *119*, e2115635118.

[86] C. Sidore, F. Busonero, A. Maschio, E. Porcu, S. Naitza, M. Zoledziewska, A. Mulas, G. Pistis, M. Steri, F. Danjou, A. Kwong, V. D. Ortega Del Vecchyo, C. W. K. Chiang, J. Bragg-Gresham, M. Pitzalis, R. Nagaraja, B. Tarrier, C. Brennan, S. Uzzau, C. Fuchsberger, R. Atzeni, F. Reinier, R. Berutti, J. Huang, N. J. Timpson, D. Toniolo, P. Gasparini, G. Malerba, G. Dedoussis, E. Zeggini, et al., *Nat. Genet.* **2015**, *47*, 1272.

[87] M. Mitt, M. Kals, K. Pärn, S. B. Gabriel, E. S. Lander, A. Palotie, S. Ripatti, A. P. Morris, A. Metspalu, T. Esko, R. Mägi, P. Palta, *Eur. J. Hum. Genet.* **2017**, *25*, 869.

[88] A. Alberdi, S. B. Andersen, M. T. Limborg, R. R. Dunn, M. T. P. Gilbert, *Nat. Rev. Genet.* **2021**, *23*, 281.

A pangenome graph reference of 30 chicken genomes allows genotyping of large and complex structural variants

Annex 3

BMC Biology

RESEARCH ARTICLE

Open Access

A pangenome graph reference of 30 chicken genomes allows genotyping of large and complex structural variants

Edward S. Rice[1,2], Antton Alberdi[3], James Alfieri[4], Giridhar Athrey[5], Jennifer R. Balacco[6], Philippe Bardou[7], Heath Blackmon[8], Mathieu Charles[9], Hans H. Cheng[10], Olivier Fedrigo[6], Steven R. Fiddaman[11], Giulio Formenti[6], Laurent A. F. Frantz[2,12], M. Thomas P. Gilbert[3], Cari J. Hearn[10], Erich D. Jarvis[6,13], Christophe Klopp[14], Sofia Marcos[3,15], Andrew S. Mason[16], Deborah Velez-Irizarry[10], Luohao Xu[17] and Wesley C. Warren[18*]

Abstract

Background The red junglefowl, the wild outgroup of domestic chickens, has historically served as a reference for genomic studies of domestic chickens. These studies have provided insight into the etiology of traits of commercial importance. However, the use of a single reference genome does not capture diversity present among modern breeds, many of which have accumulated molecular changes due to drift and selection. While reference-based resequencing is well-suited to cataloging simple variants such as single-nucleotide changes and short insertions and deletions, it is mostly inadequate to discover more complex structural variation in the genome.

Methods We present a pangenome for the domestic chicken consisting of thirty assemblies of chickens from different breeds and research lines.

Results We demonstrate how this pangenome can be used to catalog structural variants present in modern breeds and untangle complex nested variation. We show that alignment of short reads from 100 diverse wild and domestic chickens to this pangenome reduces reference bias by 38%, which affects downstream genotyping results. This approach also allows for the accurate genotyping of a large and complex pair of structural variants at the K feathering locus using short reads, which would not be possible using a linear reference.

Conclusions We expect that this new paradigm of genomic reference will allow better pinpointing of exact mutations responsible for specific phenotypes, which will in turn be necessary for breeding chickens that meet new sustainability criteria and are resilient to quickly evolving pathogen threats.

Keywords Gallus gallus, K locus, IGLL1, ev21

*Correspondence:
Wesley C. Warren
warrenwc@missouri.edu
Full list of author information is available at the end of the article

Background

Accurately detecting sequence variation associated with traits of economic importance in the domestic chicken is a major goal of genetic research into this globally widespread dietary protein source [1]. Many groups are now genotyping chicken genomes to discover the underlying molecular basis of specific traits [2–6], but current methods, both sequence- and array-based, have unquantified limitations in assessing the underlying variation that connects many loci to studied traits. Investigations in other species into the variant sets compiled by techniques relying on existing linear references have revealed large gaps in variation discovery ability [7–10]. For the domestic chicken, improved completeness and accuracy of bioinformatic queries into this variation are of vital importance to the field, as computational experiments are rapidly becoming the venue of choice to assess the potential of artificial selection to improve qualities such as growth, nutrient digestibility, reproduction, and perhaps most importantly, immune resilience.

Current frequently employed methods for genotyping whole genomes mostly share the core strategy of aligning short reads to a reference genome derived from a single individual [11]; these references are usually compressed haploid representations of diploid genomes, with toggling of haplotypes due to haploid compression, or chimeric haploblocks due to allele mixing [12, 13]. Although these methods, given a reference genome of sufficient quality and reads of sufficient coverage, are able to capture most single-nucleotide variants (SNVs) and small insertions and deletions (indels) in populations, they can lead to reference bias [14, 15], and they consistently underestimate all types of structural variants (SVs) [8]. Furthermore, for best performance, the most accurate genotyping software [16] requires preexisting high-quality data about the distribution of polymorphic sites throughout the genome for statistical calibration [17] or model training [18], information that does not exist for most species. Large-scale long-read resequencing can mitigate some of these limitations [19], but the high cost and low accuracy of long reads compared to short reads, and the large amount of existing publicly available short-read sequencing data — for chicken, there are over 40,000 short-read experiments on the SRA at the time of writing but fewer than 500 long read experiments — make a full transition to the use of long reads for resequencing studies unlikely in the near future. Although there have been improvements in algorithms for using inexpensive data such as short reads for SV detection, these methods have high false positive and false negative rates [7], so previous studies of SVs in chicken using these methods [4, 20] are likely both incomplete and inaccurate.

To counter these limitations, several methods have been developed to create and use pangenome graphs as references [21–25]. A pangenome graph is a data structure that encodes the sequence and variation present among the genomes of multiple individuals [26]. Whereas a linear reference usually contains only the compressed sequence of a single individual, a pangenome includes sequence common to all individuals as well as information about the position, alleles, and frequencies of each variant site within the input assemblies. The recent publication of a draft pangenome for human demonstrated that this new paradigm allows recovery of much sequence that appears with nonnegligible frequency in the genomes of individuals across the species but is missing from even the telomere-to-telomere linear reference [27].

Alignment of short reads to a pangenome reference instead of a linear reference has been demonstrated in humans and other species, including birds, to recapitulate and improve downstream genotype calling accuracy for both small variants (i.e., SNPs and small indels) and larger structural variants [9, 28, 29]. Large insertions are nearly uncallable when using short reads aligned to linear references, with the recall of tools such as Delly [30] falling to zero for insertions larger than 400 bp, whereas graph-based tools such as VG [28] and paragraph [22] are mostly unaffected by variant length. The human pangenome's demonstrations of improvements in read mapping, small variant genotyping, novel variant discovery, SV genotyping, and representation of complex variants [27] show the potential of this new paradigm for genome references.

In chicken, multiple alignments of reference-guided short-read assemblies [31] and de novo assemblies of high-error PacBio CLR reads [32] have revealed sequences present among chickens worldwide but missing from current references, as well as other previously unknown SVs. However, although these whole-genome alignments were both described as pangenomes by their respective authors, neither study generated a pangenome graph that can be used by other researchers as a reference for alignment to overcome the limitations presented by reference bias and difficulty in capturing SVs. They are further limited by their reliance on short reads or low-accuracy long reads, respectively, for assembly.

In this study, we generate a pangenome graph of 30 highly continuous genome assemblies of various chicken breeds, including broilers, layers, and research lines. We use this pangenome to catalog variation present in the input assemblies, including variation that was not detectable in studies using other methods, focussing on SVs in an immune system gene and a feathering-related locus as illustrations. We then go on to align short reads from

100 chickens to the graph, showing the improved performance of this method for alignment accuracy and genotyping recall compared to linear reference alignment. We expect that adoption of this new resource will allow better results in genotyping in future studies, with a goal to move toward more effective uses of chicken genome references and in the process significantly improve researchers' ability to discover the molecular mechanisms that determine bird healthiness.

Results

Selection of chromosome-level assemblies

To build assembly-based pangenome references, we used the five most continuous chromosome-level assemblies of the domestic chicken currently available, along with alternate haplotypes as applicable, and new contig-level assemblies of thirteen additional chickens, most of them locally resolved into haplotypes. The five chromosome-level assemblies have contig N50 values ranging from 5.47 to 91.3 Mb (see Table 1). This includes the current species reference assembly on NCBI RefSeq, bGalGal1b, also known as GRCg7b (contig N50 = 18.8 Mb), a fully haplotype-resolved assembly of a commercial broiler line created using the trio-binning method and an F1 cross between a representative commercial broiler and a white leghorn layer [33]. bGalGal1b, as the current RefSeq reference assembly, is fully annotated, so we use it as the source of annotations in this study. Because this assembly was made using trio-binning, its creation also resulted in a fully haplotype-resolved assembly of the genetic contribution of the other parent, a white leghorn layer. We refer to this assembly as bGalGal1w, and it is also known as GRCg7w and we use both assemblies in our pangenome.

We sequenced and assembled to the chromosome level the genomes of two additional broilers from the Ross (Aviagen) and Cobb (Cobb-Vantress) lines, among the most commercially relevant broiler lines worldwide, to capture more of the diversity present among commercial lines of domestic chickens, and to take advantage of advances in sequencing that have occurred since the assembly of bGalGal1b and bGalGal1w, especially base-calling improvements in PacBio's HiFi/Circular Consensus Sequence (CCS) technology. HiFi reads are accurate enough to allow the hifiasm algorithm to assemble phased contigs for two pseudohaplotypes [35], so although we only assembled the contigs from the primary assemblies into chromosomes, we used the alternate contigs during pangenome construction as well to take full advantage of their individual haploid diversity.

We also integrated the first nearly complete assembly of a chicken [34]. This assembly is of a Huxu, a Chinese broiler breed, and we refer to it as "huxu".

Finally, we sequenced and assembled both haplotypes of 13 additional chickens to a contig level using HiFi sequencing (Additional file 1: Table 1). These chickens include research lines bred to study immune function as well as domestic breeds originating in Spain and Egypt. We produced sequencing coverage of at least 25× (mean 35×) for each bird based on a genome size of 1.1 Gb. Using the hifiasm assembler, which is able to take advantage of the high accuracy of HiFi reads to create two locally phased haploid assemblies for each diploid individual sequenced, we successfully generated two haploid contig-level assemblies for each of 10 out of 13 birds. The remaining three birds are all highly inbred research lines, so their haplotypes are mostly indistinguishable and thus not able to be phased. Therefore, we used the primary assembly output of hifiasm for these. As a result, the pangenome graph includes phased haploid assemblies as well compressed diploid assemblies of these three highly inbred birds. In total, this resulted in 23 assemblies with a minimum contig N50 of 11 Mb (mean 15 Mb).

Together, these 30 assemblies represent a diverse set of domestic chickens, including commercial lines, research lines, and broiler and layer breeds originating on three continents. They also were assembled using three different techniques: haplotype-resolved trio-binning of PacBio CLR reads from the F1 offspring of a cross between two breeds (bGalGal1b and bGalGal1w), PacBio HiFi haplotype-resolved assembly (bGalGal4, bGalGal5, and additional chickens), and the current best-practice de novo assembly technique using a combination of PacBio HiFi and Oxford Nanopore Ultralong (ONT UL) reads (huxu) [34]. Although collectively these genomes do not

Table 1 The five chromosome-level assemblies used as a base for creation of pangenome references for the domestic chicken

ID	Assembled bird	Accession	Ref	Contig N50 (Mb)	Contig count
bGalGal1b	Commercial broiler	GCA_016699485.1	[33]	18.8	677
bGalGal1w	White leghorn layer	GCA_016700215.2	[33]	17.7	685
bGalGal4	Ross broiler	GCA_027557775.1	N/A	5.47	812
bGalGal5	Cobb broiler	GCA_027408225.1	N/A	8.33	712
Huxu	Huxu broiler	GCA_024206055.1	[34]	91.3	54

come close to fully capturing the diversity of domestic chickens worldwide, they provide a good working template of a first pangenome reference of the domestic chicken genome.

Creation of pangenome references

We constructed pangenome references of the chicken genome using two different methods, both used by the Human Pangenome Reference Consortium [27]: PanGenome Graph Builder (PGGB) [27] and minigraph-cactus [36]. PGGB and minigraph-cactus both take multiple assemblies as input, perform whole-genome alignments on them, and derive a pangenome graph from these alignments. However, these two pipelines differ in their fundamental approach: PGGB first performs reference-free multiple sequence alignment of all input sequences and then infers a graph using these alignments, whereas minigraph-cactus uses a single reference chosen by the user as a backbone and then progressively adds complexity to the graph by aligning the other sequences. We made a preliminary graph using each method and five chromosome-level assemblies (Table 1). For minigraph-cactus, we then created a 30-assembly graph using these five chromosome-level assemblies as well as the contig-level alternate haplotype assemblies of bGalGal4 and bGalGal5 and assemblies of both haplotypes of thirteen additional chickens from HiFi data (Additional file 1: Supplementary Table 1).

Due to the computational intractability of the PGGB graph as a reference for short-read alignment, as we discuss in subsequent subsections, we did not create a 30-assembly graph with PGGB, and used only the minigraph-cactus graph for most downstream applications. Nonetheless, we describe the PGGB graph in this section and refer to it occasionally thereafter for the sake of comparison. Therefore, the final two graphs we tested were the 5-assembly PGGB graph and the 30-assembly minigraph-cactus graph. We used the minigraph-cactus graph for most downstream analyses.

The minigraph-cactus pangenome graph contains 49 million nodes and 67 million edges, and therefore a mean degree, or the number of edges attached to a node, of 1.4. The total length of sequence represented in the graph — that is, the sum of the lengths of all nodes in the graph — is 1.13 Gb. The combined length of nodes traversed by the most complete assembly, Huxu, is 1.02 Gb. This is smaller than the 1.10 Gb total size of the assembly. This difference is because a path can traverse the same sequence in the graph multiple times. For example, in the case of a very simple graph containing three nodes, A, B, and C, a haplotype containing a duplication of B would have a path length of (A + 2B + C), whereas the total amount of sequence in the graph would be only (A + B + C). Therefore, there is in total 0.11 Gb (9.9%) of additional sequence in the graph compared to the total length of the nodes traversed by the most complete assembly. Of the other assemblies, bGalGal1b contributes the most additional sequence, 55.6 Mb, to the graph, whereas some assemblies contribute as little as 200 kb of additional sequence as a result of their relatedness to others (Additional file 2: Supplementary Fig. S1).

The PGGB pangenome graph contains 33 million nodes and 45 million edges, and therefore also a mean degree of 1.4. We found that parameter choice had a large effect on the numbers of nodes and edges, as well as the maximum degree, although not the mean degree (Additional file 2: Supplementary Fig. S2). By contrast, we used only default parameters for minigraph-cactus other than those pertaining to input and output.

Although the PGGB pangenome was made up of only five assemblies instead of 30, it contains more sequence than the minigraph-cactus pangenome: the total length of sequence represented in the PGGB graph is 1.23 Gb, compared to 1.13 Gb for the minigraph-cactus graph. This represents an additional 147 Mb or 12.0% of sequence compared to the total length of graph nodes in the Huxu genome (1.09 Gb). The 109 Mb of additional sequence is closer to previous estimates of total variation in diverse groups of chickens [37–40] than 147 Mb, suggesting possible overestimation by PGGB. The structures of these respective graphs are visibly different at the chromosome level in some places, such as at the beginning of chr13 (Additional file 2: Supplementary Fig. S3).

With the exception of the two sex chromosomes, only one of which can be present in each haplotype, all haplotypes are represented in all of the chromosome communities or subgraphs of both graphs; however, the presence of gaps in all assemblies except for Huxu means that there are places in all chromosomes where one or more haplotype paths is missing. In the PGGB graph, none of the contigs unassigned to chromosomes were included in the communities used to make the initial alignments, and thus all unassigned sequences were excluded from the pangenome graph. In contrast, in the minigraph-cactus graph, all sequences from all assemblies were included in the initial alignments. For all assemblies except Huxu, for which there is no unassigned sequence, a mean of 20.9 unassigned contigs containing a per-assembly total of 1.17 Mb of sequence were not aligned to chromosome subgraphs in the final graph.

Cataloging of variants present in input assemblies

A pangenome graph contains the variation present in the input assemblies and can thus be used to genotype the input assemblies compared to one chosen as a reference, based on deviations from this reference path. We

chose bGalGal1b for the reference as it is the highest-quality RefSeq-annotated chicken reference genome currently available. In total, we found 15 million variants in the minigraph-cactus graph present in at least one of the other 29 haplotypes compared to bGalGal1b. Twelve million of these variants are SNVs (Fig. 1a). This is a smaller number of total SNVs than has been detected in large panel studies [39, 40], which is likely a result of the smaller sample size of our experiment, with 30 haplotypes compared to 678 in [39]. We found a similar total length of deleted sequence, 19.2 Mb, as a previous study based on long read alignments, 19.7 Mb [38]. However, we were able to recover 18.5 Mb of inserted sequence, whereas the previous study recovered only 6.74 Mb [38] (Fig. 1a). Although distributions of lengths of deletions found previously by read alignment and by our pangenome method were broadly similar, we found more long insertions than was possible with long-read alignment (Fig. 1b).

The B cell receptor gene *IGLL1*, which has been used as a marker for plasma B cells in chicken [41], contains examples of these different kinds of variation. The overall structure of the pangenome graph of *IGLL1* shows that there are many small variants (SNVs and indels < 50 bp), as well as two SVs longer than 50 bp (Fig. 2). By encoding the presence of small variants and their allele frequencies into the reference (Fig. 2a), alignment to pangenomes has been shown to reduce reference bias compared to a linear reference [21], which we confirm below for our chicken pangenome. For example, for the SNV shown in Fig. 2a, short reads containing the non-reference allele are in less danger of mapping incorrectly as the aligner is aware of the 17% (5/30) chance of an A in this position of the genome.

The larger of the two SVs in the pangenome graph of *IGLL1* is a ~ 5 kb deletion relative to bGalGal1b present in only one haplotype of one chicken, UCD312 (Fig. 2b). By recording this low-frequency deletion in the reference, the pangenome method ensures that reads from resequenced chickens containing the deletion are able to map to both flanking sequences through edge e1 without splitting, which would introduce a potential source of error.

Finally, a ~ 300 bp insertion relative to bGalGal1b demonstrates how a pangenome graph is able to losslessly represent nested variation (Fig. 2c). The SNVs and indels

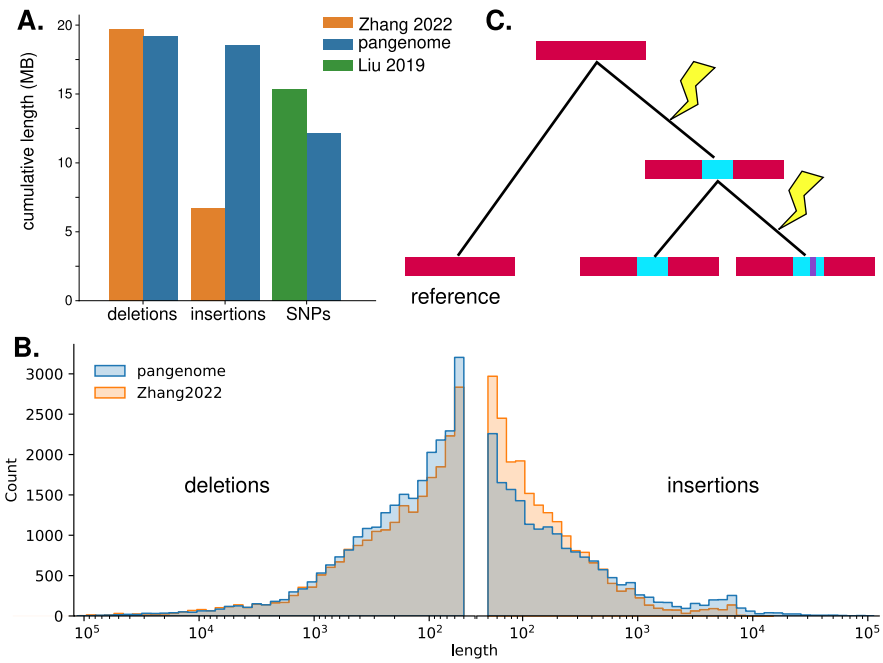


Fig. 1 Cataloging variation in the pangenome graph. **A** Total lengths of sequence contained in insertions (INS), deletions (DEL), and SNVs, compared between this study ("pangenome") and read-alignment methods [38, 39]. **B** Distribution of lengths of insertions and deletions found in this study compared to those found by Zhang et al. [38] using long reads shows that although long-read alignment finds more short insertions (< 1 kb) than the pangenome, the larger cumulative length of insertions found by our pangenome compared to Zhang as shown in **A** is driven by long insertions (> 1 kb), which have a larger effect on cumulative length. **C** A hypothetical schematic of how nested variation can evolve: an insertion mutation is followed by a later single-nucleotide mutation, resulting in an insertion relative to the reference that contains a segregating site. A genotype against a linear reference would represent these as three different alleles, whereas a pangenome conserves the nested structure of this variation

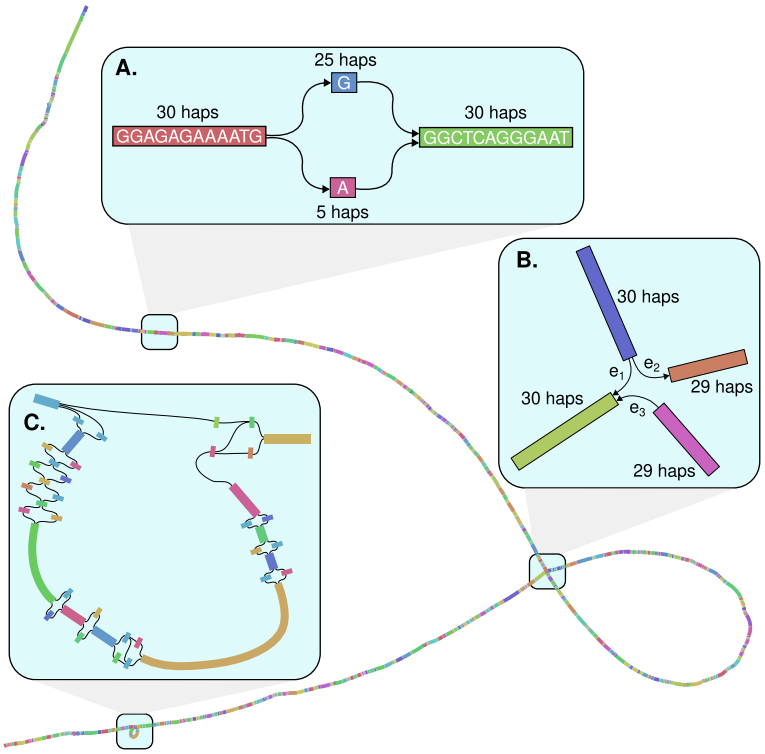


Fig. 2 A visual representation of the pangenome graph for the gene *IGLL1*. **A** *IGLL1* contains many SNVs, including one at bGalGal1b#chr15:7,955,357, in its coding sequence. The graph of this SNV shows that although all 30 haplotypes have the same sequence before and after the SNV, 25 haplotypes have G in this position and 5 have A. **B** The pangenome of *IGLL1* contains a ~5 kb deletion compared to bGalGal1b in one haplotype of a single individual, UCD312. At the juncture in the pangenome graph where the deletion haplotype branches from the rest, this haplotype follows edge e1 to skip the sequence in the loop, whereas the other 29 haplotypes follow edge e2 to include the sequence, and then e3 to join back with the deletion haplotype afterwards. **C** *IGLL1* also contains a ~300 bp insertion compared to bGalGal1b in 22 haplotypes. The inserted sequence contains SNVs, so while a linear representation of this insertion considers each version of the insertion as a different allele, the pangenome graph is able to correctly record it as a biallelic variant (i.e., insertion or no insertion) containing additional variable sites. Furthermore, reads can align to this sequence in the pangenome but would be left unmapped when aligning to bGalGal1b as it does not contain this sequence

within the inserted sequence are encoded in the exact same way as they would be in reference sequence, giving a full picture of the variation present in this region.

Disentangling a tandem repeat and viral insertion at the K locus

The K locus, short for "short wing" (*kürzer Flügel*), is a region of chrZ with an early feathering (EF) allele and a late feathering (LF) allele [42, 43]. The EF allele contains single copies of the genes *PRLR* and *SPEF2*. The LF allele contains a tandem duplication of parts of both genes [44], and often, but not always [45, 46], an insertion of the sequence of the avian leukosis virus ev21. The reference genome bGalGal1b has the EF allele and no ev21 insertion, so genotyping the K locus in other chickens using this reference is difficult because ev21 has a length of over 7kbp [46], an order of magnitude longer than the maximum insertion size that can be genotyped with short reads and a linear reference [28]. As such, it is a region that can be more accurately genotyped with the use of a pangenome graph approach.

We first created a one-dimensional representation of the minigraph-cactus pangenome graph structure of the K locus colored by path coverage, as a node through which the same haplotype path travels more than once indicates a duplication (Fig. 3a). This representation shows that although most of the haplotypes represented in the pangenome graph contain only one copy of this locus, Huxu has a duplicated region and an insertion. The 2× path coverage region in Huxu covers parts of both *PRLR* and *SPEF2*, consistent with the tandem duplication found by Elferink et al. [44]. We also found a misassembly in bGalGal1w, with unassigned scaffolds containing the sequence (see Additional file 2: Supplementary Note 1 [33, 44] and Supplementary Fig. S4). Furthermore, Huxu contains an insertion relative to the reference sequence bGalGal1b. Alignment verified that the inserted sequence is the ev21 viral genome.

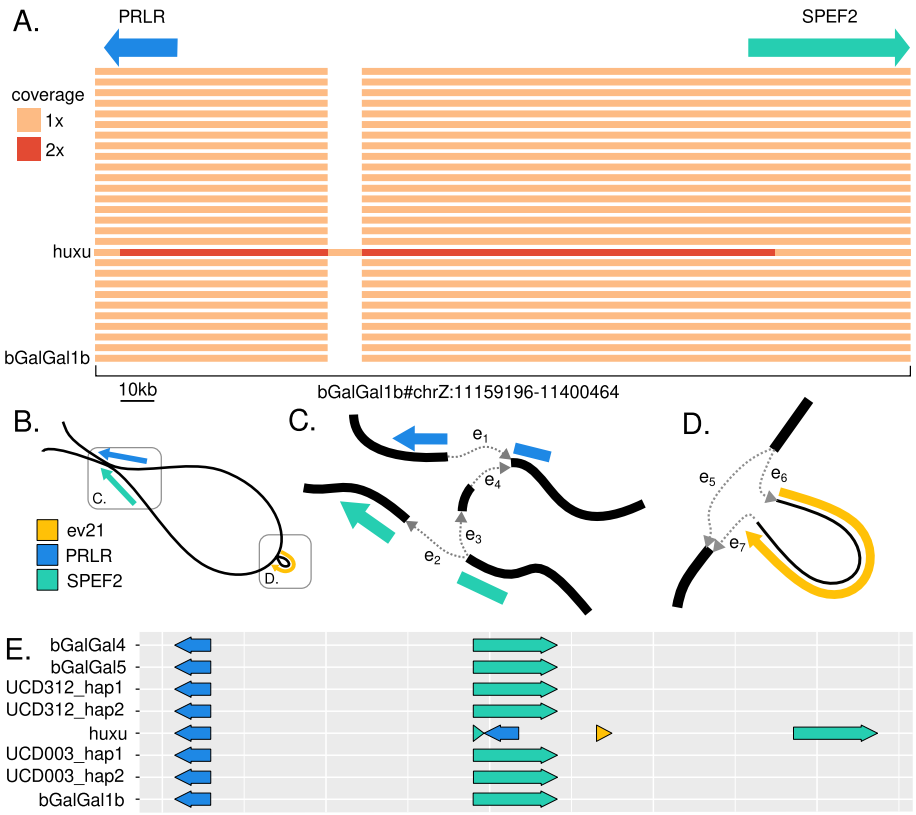


Fig. 3 Disentangling complex variation at the K locus with the pangenome graph. **A** A one-dimensional view of the pangenome subgraph for the K locus, with nodes colored by path coverage (i.e., the number of times a haplotype path passes through them) and the locations of the genes *PRLR* and *SPEF2* denoted. Huxu shows double path coverage of part of the locus, as well as an insertion. Alignment verified that this insertion contains the sequence of the avian leukosis virus ev21. **B** A two-dimensional view of the same graph, showing both the tandem duplication and the ev21 insertion. **C** At the junction where the paths containing the tandem duplication deviate from the paths that do not, all paths begin by traversing edge e1 and moving through most of the sequence of the K locus. However, at the e2/e3 fork, a path can either traverse e2 to leave the K locus, or traverse e3 and e4 to include a tandem duplication of parts of *PRLR* and *SPEF2*. **D** A more detailed view of the ev21 insertion, showing the two possible paths at this juncture: a path can traverse edge e5 to skip the insertion, or it can traverse edge e6, then the ev21 sequence, then e7, to include the insertion. **E** Linear untangled view of the locus, confirming previous studies of the structure of the locus, with a tandem duplication of parts of both genes and an insertion of the ev21 sequence

Next, to better understand the structure of the locus, we created a two-dimensional representation of the graph at this locus (Fig. 3b–d). This representation of the graph shows the tandem duplication as a junction where a path can either leave the K locus or repeat it (Fig. 3c), and the insertion as a loop containing the ev21 genome covered only by Huxu (Fig. 3d).

Finally, to view the alleles linearly, we used the "untangle" function of ODGI [24] to lay out each haplotype of the minigraph-cactus graph (Fig. 3e). The resulting gene layout of the two alleles is consistent with previous knowledge about the structure of the locus [44–46].

Genotyping ALVEs in the pangenome graph

In addition to the ev21 insertion present in some alleles of the K locus, chickens carry other endogenous retroviral insertions of avian leukosis virus subgroup E (ALVE) [47]. Many of these viral insertions remain at least partially functional, retaining their ability to express individual viral proteins or even create full viral particles [48]. The presence of some of these insertions in the chicken genome has been shown to be associated with phenotypic traits such as egg production [49], plumage color [50], and disease susceptibility [51]. As such, these insertions represent structural variants with known phenotypic effects, so we searched for and genotyped them in our pangenome graph.

Including ev21, we found 18 ALVEs common in commercial layers and broilers (Additional file 2: Supplementary Fig. S5). Most (12/18) of these ALVEs are present in only one haplotype, but others are present in up to 20 haplotypes (ALVE1). ALVE-TYR, present in 3 of the 30 haplotypes in the pangenome, disrupts the *Tyrosinase* gene, causing a recessive white phenotype and reductions

in growth rate of muscle mass [52]. Two of the genes in ALVE3, *gag* and *env*, present in seven haplotypes, are known to be highly expressed due to their placement within an intron of the non-viral *HCK* gene. This expression offers some degree of protection from exogenous avian leukosis virus infection through receptor interference [53], but can also lead to immune tolerance, with lower antibody production and higher mortality [54].

Use as a reference for resequencing and genotyping

Given the improvements in accuracy and recall of genotyping shown in other species by using pangenome graph-based methods, we set out to demonstrate the usefulness of our pangenome representations for alignment and genotyping. For this, we used simulated short reads as well as short reads from 100 domestic and wild chickens (Additional file 3: Supplementary Table 2). For comparison between linear and graph-based methods, we called genotypes using both linear alignments to bGalGal1b as well as graph alignments to our pangenomes.

For downstream use by existing short-read genotype callers, alignments must be converted from graph coordinates to linear coordinates; this process is called surjection. Alignment of short reads to the PGGB graph and surjection to bGalGal1b was infeasible, with a throughput of only 1.6 reads per CPU-second on a test set of 10 k paired-end reads, and inability to complete alignment of a larger test set of 1 M paired-end reads without running out of memory with 250 GB allocated to the job. Further investigation revealed that surjection was the bottleneck, as graph alignment without subsequent surjection had a throughput of 147 reads per CPU-second and a maximum memory usage of 31 GB for the 1 M test set. By comparison, alignment of the 1 M test set to the minigraph-cactus graph followed by surjection to bGalGal1b had a throughput of 500 reads per CPU-second and a maximum memory usage of 24 GB, and minimap2 could align 1832 reads per CPU-second to bGalGal1b with 5.4 GB memory (Fig. 4a, b).

To compare accuracy of graph alignment to linear alignment, we simulated one million pairs of paired-end reads through sampling from the graph with random errors added, and aligned them to both the cactus-minigraph pangenome with VG giraffe [9] and the linear bGalGal1b reference with minimap2. We then determined the accuracy of the alignments by comparing the location to which reads were aligned to the location from which they were sampled. Giraffe performed better than minimap at every level of stringency, based on what percentage of all reads were mapped correctly (Fig. 4c).

To test the downstream effects of these differences in mapping accuracy, we genotyped 100 chickens from diverse breeds using both giraffe pangenome alignments

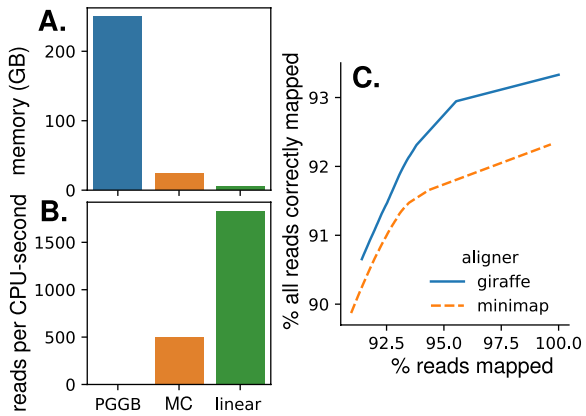


Fig. 4 Comparing pangenome and linear aligner performance for short reads. **A**, **B** Alignment of short reads with VG giraffe is more memory-efficient (**A**) and faster (**B**) when aligning to the minigraph-cactus (MC) pangenome graph compared to the PGGB graph. Linear alignment with minimap2 is the fastest and most memory-efficient. **C** A larger percentage of all simulated reads is correctly aligned with giraffe regardless of how permissive the minimum map quality filter is

and minimap linear alignments of 10–15× coverage short reads, and compared the results between the two methods (Fig. 5). Whereas the two methods found similar sets of SNVs (Fig. 5a) and indels (Fig. 5b), there were substantial differences. Agreement was unsurprisingly higher for SNVs, although the pipeline using giraffe alignments found a larger number with a quality score of at least 10 than the pipeline using minimap (Fig. 5a). For variants found by both methods, per-sample SNV concordance had a mean of 97.9% with a standard deviation of 9.1% (Fig. 5c). Indel concordance was lower, with a mean of 94.0% and a standard deviation of 12.9% (Fig. 5d).

To determine whether reference bias is a factor in the different genotyping results between the two methods, we examined the proportion of mapped reads containing the reference allele at putative heterozygous SNV sites. Reference bias across these sites, which we define as the difference between the mean fraction of reads containing the alternate allele and the expected alternate allele fraction of 0.5, is lower for all of the 100 chickens when using pangenome alignment instead of linear alignment, with a mean reference bias reduction of 38% (Fig. 5e, Additional file 2: Supplementary Fig. S6).

To connect the genotypes of these chickens to the geographic origins of their breeds, we performed principal component analysis (PCA). Although there is not complete separation of geographic origins on the PCA plot, as expected due to admixture, American and Northern European breeds fall into narrow bands on PCs 1 and 3, respectively, while Asian breeds are more diverse (Additional file 2: Supplemental Fig. S7).

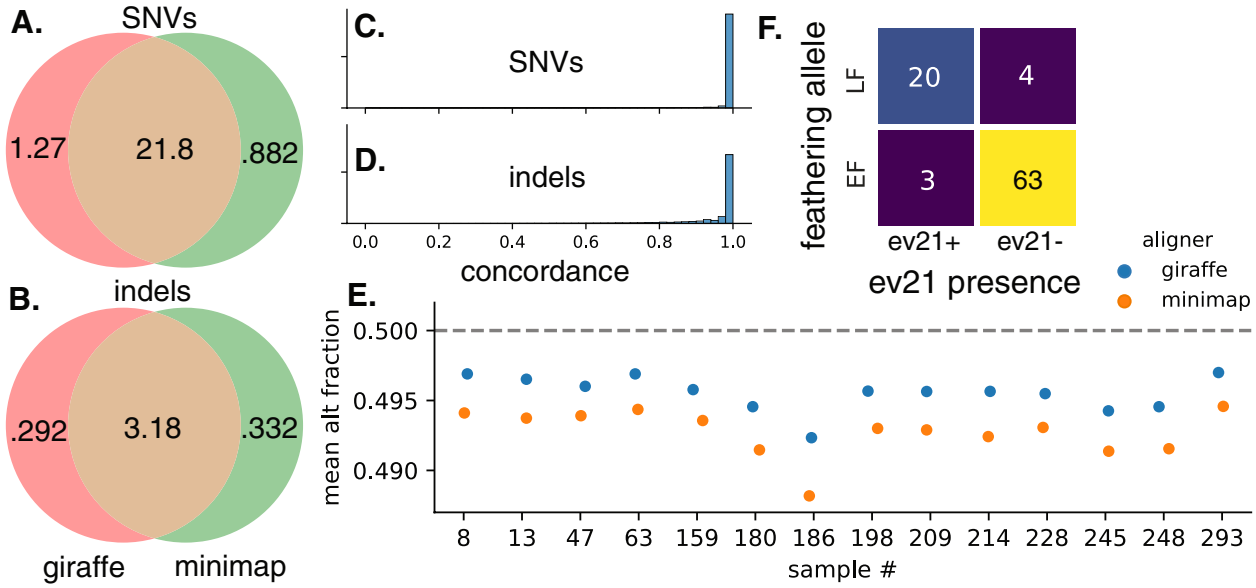


Fig. 5 Genotyping 100 diverse chickens. **A**, **B** Counts in millions of common and different SNVs (**A**) and indels (**B**) found by genotyping pipelines using giraffe vs. minimap as the aligner. Only variants with a quality score of at least 10 are considered. **C**, **D** Concordance distributions for SNVs (**C**) and indels (**D**) detected by both genotyping methods with QUAL ≥ 10. **E** Mean fractions per sample of mapped reads containing the alternate allele at putative heterozygous sites show that giraffe alignments contain less reference bias for every chicken, as they deviate less from the expected value of 0.5. Sample information in Additional file 3: Supplementary Table 2 and full plot for all 100 chickens in Additional file 2: Supplementary Fig. S6. **F** Genotyping 100 chickens at the K locus reproduces previous results finding that although most chickens with the late feathering allele (LF) also have an ev21 insertion at the K locus (ev21+), some chickens have the late feathering allele without an ev21 insertion

Finally, we used the short-read alignments to the pangenome graph to genotype the K locus based on edge coverage (Fig. 5f). All of these chickens are female and thus only have one copy of the Z-linked K locus. Of the 100 chickens, 23 have the ev21 insertion (ev21+) and 24 have the tandem repeat (late feathering/LF). As found in previous studies [45, 46], the ev21 insertion and the tandem duplication are not inextricably linked, although they do usually appear together: three chickens, all standard Rhode Island breeds, have the ev21 insertion but not the tandem repeat, and four chickens, two Silkies and two Cochins, have the tandem repeat but not the ev21 insertion.

Discussion

With the quickly accumulating numbers of haplotype-resolved genomes for many species, the pangenome model of integrated presentation of within-species variation stands to become ubiquitous [26, 27]. Such resources already exist for other livestock such as swine [55] and cattle [56, 57]. One of the greatest advantages of pangenome references in other species has been the capture of sequences not present in linear reference genomes. Compared to the nearly complete assembly of the Huxu chicken genome, our pangenome graph contains 109 Mb of additional sequence. Some of this additional sequence comes from SNVs or small indels that are relatively straightforward to represent in the context of a linear reference, and some of it is made up of nodes whose sequences are similar to nodes traversed by the Huxu assembly, but are represented separately. Thus, the true accessory genome length is likely less than 109 Mb compared to Huxu. Nonetheless, the tripling of total insertion length detectable using this pangenome compared to using long-read alignments as in a previous study [38] shows that much of this additional sequence is made up of variation that cannot be represented in a traditional linear reference genome, and therefore, many reads from these regions of the genome cannot be mapped to a linear reference because it does not contain the parts of the genome the reads came from. By adding additional assembled chicken genomes of more diverse origins this amount of novel sequence will grow.

Other studies have presented multiple alignments of chickens as pangenomes [31, 32], but our graph-based approach, which uses assemblies based on long and highly accurate PacBio HiFi reads as well as one near-complete assembly, allows the pangenome to be used not just as a method for cataloging variation present in the input assemblies, but also as a reference for future resequencing studies. By comparing pipelines using linear versus pangenome alignments of short reads to genotype

100 chickens from diverse breeds, we demonstrated the improved alignment performance of pangenome alignment over linear alignment and showed the downstream effects of these improvements on genotyping. Unfortunately, there does not yet exist a high-confidence truth set of variant calls for chickens as there does for humans [16], so we cannot compare the accuracy of these differing genotype calls. Nonetheless, given the improvements in alignment performance we have shown in chicken with both simulated and real reads, and the improvements in genotyping demonstrated in human and yeast by using the giraffe pangenome aligner [9, 27], we predict that the genotypes we inferred using giraffe pangenome alignment are substantially more accurate than those we inferred using linear alignment.

We tested two approaches for creating pangenome graphs, PGGB and minigraph-cactus, which were both used by the Human Pangenome Reference Consortium to create the first draft of the human pangenome [27]. These two methods each have advantages and disadvantages which we explore here. PGGB uses a reference-free approach, whereas minigraph-cactus makes alignments to a single reference. The single-reference approach of minigraph-cactus greatly increases the efficiency of alignment to the graph, but also results in some regions of non-reference sequence being clipped, as shown in Additional file 2: Supplementary Fig. S3. Furthermore, minigraph-cactus is able to choose most alignment parameters automatically, whereas PGGB results are highly dependent on parameter choice. Due to the absence of a deterministic process for choosing best parameters or even evaluating and comparing graphs made with different parameters, the ability of minigraph-cactus to automatically choose alignment parameters presents an advantage over PGGB. In the end, our choice of the minigraph-cactus graph for most downstream analyses was primarily based on the computational intractability of the PGGB graph for use as an alignment reference; regardless of accuracy or completeness, a graph to which only 1.6 reads per CPU-second can be aligned is not usable for most purposes with the resources currently available to most genomics researchers.

Our determination of the structure of the K locus and subsequent genotyping demonstrates the power of pangenome graphs in the study of loci containing complex structural variants. The initial discovery of the insertion of an endogenous avian leukosis virus in the late feathering allele required cell culture work [58], and a later study establishing the tandem repeat [44] necessitated extensive quantitative PCR experiments targeted at 20 different segments of the locus. Although the latter was performed after a linear reference genome was available, this reference, like all subsequent versions of the reference genome for chicken, contains the early feathering allele and no ev21 insertion at the K locus, and no current method can reliably genotype SVs of this size using short reads and a linear reference [28]. More recent work on the relationship between the ev21 insertion and the late feathering phenotype, though undertaken after improved reference genomes and large amounts of public sequencing data from different breeds of chickens became available, also relied on targeted PCR [45, 46]. In contrast, we were able to replicate these findings using only existing short-read whole-genome sequencing data and pangenome methods. We expect that our pangenome, and future pangenomes using telomere-to-telomere genome assemblies, which exist for increasing numbers of species [59–63] but not yet chickens, will enable discoveries about complex structural variation at important immune loci such as the major histocompatibility complex (MHC) and T cell receptor gene (TCR), providing insight into the genetic diversity necessary to fight evolving pathogen threats in this major worldwide source of protein, which also threaten wildlife with increasing frequency [64].

The tool used by both the minigraph-cactus and PGGB pipelines to produce a VCF of the input assemblies based on the graph, vg deconstruct, does not currently classify SVs based on type, e.g., as inversions or translocations, but instead represents all SVs as either inserted or deleted sequence. Therefore, a complex variant such as a translocation is represented as a deletion of sequence in one location and an insertion of the same sequence in another location. We detected the tandem duplication present at the K locus through manual examination of graph structure. Tools such as vcfwave [65] are able to secondarily reclassify these complex structural variants properly, but due to the rapidly changing nature of software in this field, we report SVs only as insertions and deletions. We expect future versions of this pangenome to use new tools to report inversions and translocations as well.

For the most part, we were able to use best practices established by the human pangenome reference consortium [27] for the creation and use of this pangenome. However, in some cases, such as our inclusion of highly inbred research lines that could not be phased due to the similarity of their haplotypes, there is no available precedent from the human pangenome. As pangenomes are built for more species, we hope to see consensus emerge about best practices for cases such as this that do not apply to humans.

We created this first draft of the chicken pangenome out of a mixture of commercial and research lines and previously published reference assemblies. Despite this somewhat arbitrary sampling process based mostly on

sample availability, using the pangenome as a reference increases accuracy, decreases reference bias, and makes it possible to genotype structural variants that are too large and complex to genotype with a linear reference and short reads. Nonetheless, we expect future versions to improve these measures even further through the inclusion of more chickens, sampled more strategically, to best capture the full diversity and variant frequencies of chickens worldwide.

Conclusions

In this paper, we have presented the first pangenome graph reference for the domestic chicken. We show its utility as a catalog of variation, including structural variation too large or complex to be detected using previous methods, and as a reference for the alignment of short reads. Given the improvements we have demonstrated in this model over a linear reference, we expect this pangenome, and new versions with additional broadly diverse chicken breeds incorporated, to serve as a resource to the community for future resequencing studies as well as investigation of complex loci, especially in immune-related genes.

Methods

Sequencing and assembly of bGalGal4 and bGalGal5

One female Ross 308 (Aviagen) and one female Cobb 550 (Cobb-Vantress), both commercial broiler chickens, were euthanized in the framework of a research experiment at 38 days of age. Cardiac puncture was immediately employed to collect 12 aliquots of 100 µl of blood in tubes with EDTA and 1 ml of ethanol > 99.7% from each animal. Samples were frozen at −20 °C.

For both assemblies (bGalGal4 and bGalGal5), we followed the VGP 2.0 pipeline [12]. We generated 32× Pacbio HiFi data on a Sequel IIe, and then used cutadapt [66] to trim off adapters that were not trimmed in the Pacbio software processing. We assembled contigs using HiFiasm v0.14 [67], generating a semi-haplotyped phased primary contig and alternate contig assembly. From the primary assembly, we removed false haplotype duplication and placed them in the alternate using purge_dups v1.2.5 [68]. We then scaffolded the contigs with Bionano Genomics optical maps (319× and 459× respectively), generated on a Saphyr instrument using DLE label, with Bionano Solve. We then further scaffolded with Arima Genomics Hi-C v2 (65× and 122× respectively), using salsa v2.2 [69]. The primary assembly was then curated using gEVAL [70], structural errors corrected, and chromosomes named according to their numbers in the bGalGal1 GRC7g reference. 10X Genomics data were also generated, and used for orthogonal validation, but not scaffolding. The primary and alternate assemblies were deposited in NCBI under accession numbers GCA_027557775.1 (bGalGal4) and GCA_027408465.1 (bGalGal5), and all data are available in Genome Ark (https://genomeark.github.io/genomeark-all/Gallus_gallus/).

Sequencing and assembly of additional chickens

High molecular weight (HMW) DNA from blood of 13 juvenile male chickens (Additional file 1: Supplementary Table 1), maintained and bled under ADOL IACUC-approved Animal Use Protocol #2019-15 for breeder management, was sequenced on the Pacific BioSciences Sequel IIe. HMW samples were sheared using a Diagenode Megarupter3 shearing device targeting 18–22 kb fragments. Libraries were prepared with the PacBio SMRTbell Prep Kit 3.0. Library size distribution was determined on the Agilent Femto Pulse and a Qubit fluorometer was used to measure concentration. Sequencing polymerase was bound to the SMRTbell libraries with the Binding Kit 3.2 and run on Sequel IIe with the Sequel II Sequencing Kit 2.0 and SMRT Cell 8 M. HiFi data was collected with Instrument Control Software Version 11.0 and Chemistry Bundle 11.0 with a movie time of 30 h. The On Plate Loading Concentration was 130pmolar.

HiFi reads for each of the chickens were assembled into contigs using hifiasm v0.18.9 [35] with default options. Both haplotypes output by hifiasm were used in subsequent analyses.

Creation of PGGB pangenome

We constructed a pangenome reference from the five input assemblies bGalGal1b, bGalGal1w, bGalGal4, bGalGal5, and HuxuT2T (Table "assemblies"). First, we extracted chromosome sequences from the assemblies and gave them names according to the PanSN-spec, in the format of "[assembly name]#[chromosome name]", e.g., "bGalGal4#chr5". The PGGB pipeline recommends first partitioning the assemblies into communities, where each community is a set of sequences that should be aligned to each other, for example, all sequences from each assembly assigned to the same chromosome. We partitioned the assemblies into 41 communities, one for each chromosome based on whole-genome alignments made with mashmap [71] in one-to-one mode and a percent identity cutoff of 90%, and then constructed a pangenome graph for each chromosome separately. Due to disagreements in the naming of microchromosomes among the five assemblies, some of the communities contain chromosomes named differently in the different assemblies (Additional file 4: Supplementary Table 3).

For every chromosome, we constructed its pangenome graph using the Pangenome Graph Builder (PGGB) v0.4.1 [27]. Briefly, this pipeline uses wfmash v0.9.1 [72]

to align the input assemblies, seqwish v0.7.6 [25] to build a graph from the alignments, smoothxg v0.6.5 [73] and gfaffix v0.1.3 [74] to clean up the graph, and odgi v0.7.3 [24] to visualize the graph. We first ran pggb with default parameters, except for parameter "-n" set to the number of assemblies being aligned for the chromosome in question (this number is five for most chromosomes, with the exception of sex chromosomes and some microchromosomes without full representation in all five assemblies) and "-G 3079,3559". For postprocessing and optimal visualization, we redrew the 2D graph visualization using the odgi draw command with parameters "-C -w1000," and we redrew the 1D graph visualization by first resorting the graph based on positions in the bGalGal5 path using the command odgi sort with parameters '-H < (echo "bGalGal5#${chromosome_name}") -Y' and then drawing with the odgi viz command with default parameters.

To find the optimal parameters for each chromosome, we performed a parameter sweep of the segment length (-s), mapping percent identity (-p), and minimum match length (-k) options to the pggb command. We tested every member of the cartesian product set of the parameter values s={5 k, 10 k, 30 k, 50 k, 80 k}, p={85, 90, 94,97}, and k={10, 19, 50, 100, 150}. We evaluated the results as suggested in PGGB documentation, using a combination of examination of graph statistics, especially node count and maximum degree, with the odgi stats command and visual inspection of the graph structure using the odgi viz output. For some microchromosomes, we made more granular adjustments to the parameters to fine-tune their graphs. Additional file 4: Supplementary Table 3 shows the final parameters chosen for each chromosome.

Finally, we created a single pangenome graph containing the respective connected component for each community using the odgi squeeze command with default parameters. This resulted in a single graph file with extension ".og" that is easily convertible to other sequence graph formats such as GFA and VG.

Creation of minigraph-cactus pangenome

We ran the minigraph-cactus pipeline [36] using the cactus v2.4.2 Docker image and a nextflow pipeline built for this purpose [75]. As input, we used the five chromosome-level assemblies in Table 1, the alternate haplotypes of bGalGal4 and bGalGal5, and both haplotype assemblies of an additional 13 chickens listed in Additional file 1: Supplementary Table 1. We specified bGalGal1b as the reference, because although it is not the highest-quality assembly, it is the best RefSeq-annotated assembly on NCBI, so we wanted to call variants against it downstream.

Additional sequence analysis

We determined the amount of additional sequence contributed to the graph by each sample through an iterative process. First, we removed all nodes traversed by the Huxu assembly from the graph as it is the most complete assembly. Then, for each remaining bird, we summed up the length of all nodes traversed by either haplotype of this bird, found the bird with the largest sum, and removed all nodes traversed by this bird's haplotypes from the graph. We repeated this process until there were no samples remaining. The python program we wrote for this purpose is included in the repository cited in the Code Availability statement.

Format conversions and subgraph extraction

To convert GFAv1.1 format as output by minigraph-cactus to OG format for downstream use in ODGI visualization tools, we used the command "vg convert -gfW" to convert to GFAv1.0, and then "odgi build -g -Os" to build an OG graph out of the GFAv1.0 file.

To convert GBZ format to HG format, we used the command "vg convert".

To convert HG format to GFA format, we used the command "vg convert -f".

To convert OG format to GFA format, we used the command "odgi view -a -g".

To extract regions from graphs in HG format, we used the command "vg find -p 'bGalGal1b#[chromosome]:[start]-[end]'".

To extract regions from graphs in OG format, we used the command "odgi extract -d0 -E -r 'bGalGal1b#[chromosome]:[start]-[end]'".

Genotyping input assemblies

Both assembly-based graph construction pipelines, pggb and minigraph-cactus, can output vcf files containing genotypes for the input assemblies relative to the reference, in our case bGalGal1b. Minigraph-cactus does this by default; pggb does with the addition of the option "-V 'bGalGal1b:#:'". Where necessary, we concatenated vcf files for each chromosome into a single genome-wide vcf using the bcftools concat command v1.15.1 [76].

Graph visualization

To visualize specific regions of the pangenome graph, we first looked up coordinates relative to bGalGal1b on RefSeq, extracted them from the graph, output in GFA format, and visualized using bandage v0.8.1 [77].

Commands for extraction and conversion are given under the heading "Format conversions and subgraph extraction".

Genotyping ALVEs

As previously described in [4th chicken report], we identified assembled Avian Leukosis Virus subgroup E (ALVE) integrations by performing a search for ALVE1 (GenBank: AY013303.1) with BLAST v2.10.0 [78] in each of the contributing fully assembled reference sequences using ALVE1 (GenBank: AY013303.1). We used flanking sequence to annotate ALVEs with known integration sites [47]. We then translated all coordinates to bGalGal1b coordinates using odgi position and looked up these insertions or deletions relative to bGalGal1b in the minigraph-cactus vcf output.

Read simulation

We simulated reads using the "vg sim" command with a nucleotide substitution error rate of 0.24% as estimated by Pfeiffer et al. [79] and an indel error rate of 0.029% as in [9]. This command randomly samples reads from the pangenome graph and adds errors based on the specified error rates, keeping information about the location from which the reads were sampled in the read header so that it can be used to test accuracy downstream.

Sequencing of short read chickens

We sampled 236 chickens from 62 breeding farms that specialize in heritage and rare chicken breeds in May and December 2021. In short, we collected 0.5–2 mL of blood from each bird by puncturing the brachial vein with a syringe (gauge size 18.5–28 depending on the size of the bird). The blood was immediately expelled through the syringe into K2EDTA vacutainers and stored on dry ice. Upon arrival at the lab, the blood samples were transferred to a −80 °C freezer. DNA was extracted using the QIAamp Fast DNA Tissue Kit. Library preparation and sequencing were performed at BGI Group. Libraries were prepared using a DNA short-insert protocol for 150 bp paired-end reads and sequenced on the DNBseq platform. Seven samples failed to be sequenced due to low quality, so were excluded from further analyses. We chose a subset of 100 of these samples for the final analysis, selecting breeds that were previously genotyped at the K locus [45, 46] where available and choosing the rest by balancing the conflicting goals of including multiple chickens from the chosen breeds and having many breeds represented.

Short read alignment

To align short reads to the PGGB graph, we first converted the graph to GFA format using the command "odgi view -g" and then converted the GFA format to GBZ format [80] and created giraffe indices from the output with the command "vg autoindex -w giraffe." The minigraph-cactus pipeline outputs all indices necessary to run giraffe by default, so no further processing was necessary to prepare it for alignment of reads with giraffe.

To test timing and memory usage, we arbitrarily chose a publicly available set of short reads from a chicken (SRR9967588) and subsetted the first 1 million pairs. This test failed for alignment to the PGGB graph due to running out of memory, but a smaller subset of 10,000 read pairs was successful. We aligned the test set of reads to the graph using the command "vg giraffe" with arguments "-o BAM." Because the PGGB graph does not contain a reference sequence like the minigraph-cactus graph, we additionally specified the reference chromosomes with the arguments "--ref-paths bGalGal1b_paths.tsv," where bGalGal1b_paths.tsv is a tab-separated file containing a list of all chromosomes in bGalGal1b and their sizes. For comparison to alignment to a linear reference with minimap2 v2.24 [81], we created a short-read minimap index of bGalGal1b with the command "minimap2 -x sr -d" and then aligned reads to it with the command "minimap2 -a" piped to "samtools view -bh" with samtools v1.16.1 [76] to convert to bam format for a fair comparison, since we ran giraffe with bam output.

For alignment of short reads from 100 chickens, we ran vg giraffe with default options, outputting the results in GAM format. We surjected the GAM files to BAM format with bGalGal1b as the reference genome using the command "vg surject" with default options.

Comparison of linear and graph alignments with simulated reads

To compare the accuracy of alignments of simulated reads between linear and graph aligners, we aligned the simulated reads both to the bGalGal1b linear reference using minimap2 and to the pangenome graph reference using giraffe, as described above. We converted the minimap2 output to GAM format using the command "vg inject," and then compared the minimap2 and giraffe GAMs to the truth set using "vg gamcompare," all as in [9].

Genotyping

We genotyped the 100 chickens based on these alignments using elprep [82] v5.1.2, a multithreaded reimplementation of GATK. Briefly, we generated an elfasta sequence reference (an indexed binary form of the reference fasta for downstream use) for bGalGal1b using the command "elprep fasta-to-elfasta," created a list of sites from the minigraph-cactus vcf output with SVs larger than 1000 bp filtered out using the command "elprep

vcf-to-elsites," and ran the "sfm" command with settings as recommended in the manual to generate a gvcf for each bird, which we then combined into a single gvcf with GATK CombineGVCFs and joint genotyped with GATK GenotypeGVCFs [17]. The location of our scripts for genotyping, as well as all other analyses in this paper, is given in the "Availability of data and materials" section.

Genotyping method comparison

To compare the respective outputs of the giraffe- and minimap-based genotyping pipelines, we used bcftools v1.17 [76] command "isec -c some" to create four vcf files: variants only detected by the giraffe pipeline, variants only detected by the minimap pipeline, giraffe pipeline calls of variants detected by both pipelines, and minimap pipeline calls of variants detected by both pipelines. We counted variants with QUAL ≥ 10 in all of these files, subsetting by variant type with "bcftools view -v [snp|indel]." To compare the per-sample calls made by the respective methods for variants detected by both, we used "bcftools merge --force-samples" to create a single vcf containing calls made by both methods, and then used a custom python script (included in code availability) to calculate the percent agreement for each variant.

Reference bias estimation

We estimated the amount of reference bias by calculating the mean fraction of reads mapping to putative heterozygous sites containing the alternate allele, and comparing to the expected value of 0.5. We define putative heterozygous sites as positions with coverage of at least 10× where the portion of reads containing the minor allele is at least 25%, as in [15]. Briefly, we filtered low-quality mappings and multimapping reads with "samtools view -F2304 -q10," created pileups with "samtools mpileup -d100 –no-BAQ," and piped the results to a custom C program to find putative heterozygous sites and calculate alternate allele frequencies at these sites. All code used to perform this analysis is in the project's code repository.

Principal components analysis

To visualize the shared genetic ancestry across chicken breeds, we performed a PCA using Plink 2.0 [83]. We filtered for linkage disequilibrium using the parameters "*indep 50 5 0.5*" following Dementieva et al. [84]. We grouped the samples by the geographic origin of the breed.

K locus genotyping

To genotype the K locus, we converted each GAM file to GAF format using the command "vg convert -G" and counted reads covering the edges e1 through e7 as shown in Figure "K locus." We used binomial tests with p-value cutoffs of 0.05 to assign genotypes to each chicken for both the ev21 insertion and the tandem duplication; chickens with both $p(\text{insertion}) > 0.05$ and $p(\text{no insertion}) > 0.05$ were marked as inconclusive.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12915-023-01758-0.

Additional file 1: Supplementary Table 1. Additional chickens sequenced with HiFi reads and assembled for inclusion in the minigraph-cactus pangenome.

Additional file 2: Figure S1. Additional sequence per sample. **Figure S2.** Effects of parameters on PGGB graph. **Figure S3.** Comparison of chr13 between PGGB and minigraph-cactus. **Figure S4.** Unplaced contigs at K locus. **Figure S5.** Genotyping ALVEs in the pangenome. **Figure S6.** Reference bias in giraffe vs. minimap alignments. **Figure S7.** PCA of short read chickens.

Additional file 3: Supplementary Table 2. Chickens sequenced with short reads and genotyped using pangenome graph alignments.

Additional file 4: Supplementary Table 3. Final parameters for PGGB for each chromosome.

Acknowledgements

We thank Adam Novak and Jordan Eizenga (UC Santa Cruz) for discussion and bug fixes related to our usage of the VG toolkit for this project and Laurie Molitor and Melanie Flesberg (USDA-ARS) for assistance with animal care and DNA isolation.

Authors' contributions

WCW and ESR conceived and designed the project. AA, JA, GA, HB, HHC, MTPG, CJH, SM, and DV generated sequence data used in this project. ESR, JRB, OF, GF, EDJ, and LX assembled genomes used to create the pangenome. ESR, PB, MC, SRF, LAFF, CK, and ASM genotyped chickens used in this project. ESR constructed the pangenome. ESR and WCW wrote the manuscript. All authors edited and approved the manuscript.

Funding

This work was supported by USDA NIFA grants 2020-67015-31574 and 2022-67015-36218 and the European Union's Horizon Research and Innovation Programme under grant agreement No. 817729 (Project HoloFood). Computation for this work was performed on the high performance computing infrastructure provided by Research Computing Support Services and in part by the National Science Foundation under grant number CNS-1429294 at the University of Missouri, Columbia MO. The authors gratefully acknowledge the Gauss Centre for Supercomputing e.V. (www.gauss-centre.eu) for funding this project by providing computing time on the GCS Supercomputer SuperMUC-NG at Leibniz Supercomputing Centre (www.lrz.de).

Availability of data and materials

The datasets generated and/or analyzed in the current study are available in NCBI repositories under BioProject accessions PRJNA838369 [85], PRJNA838370 [86], PRJNA971225 [87], and PRJNA1031205 [88]. The pangenome graph, a vcf of variants present in the graph, and vcfs of the resequenced chickens genotyped using both linear and pangenome methods are available in a Zenodo repository at https://doi.org/10.5281/zenodo.10018222 [89]. The code used to perform the analysis in the current study is available on GitHub at https://github.com/WarrenLab/chicken-pangenome-paper [90].

Declarations

Ethics approval and consent to participate

Chickens used for the bGalGal4 and bGalGal5 assemblies were euthanized according to the procedures regulated in the Spanish Royal Decree RD

53/2013. Experimentation procedures were approved by the Ethical Committee of Generalitat de Catalunya, Spain (Proceeding number 10226).
Chickens used for additional assemblies were maintained and bled under ADOL IACUC-approved Animal Use Protocol #2019-15 for breeder management. SPF birds from each line were grown in colony cages and provided food and water ad libitum.
For chickens used for short-read sequencing, all handling and sample collection of animals were performed in accordance with TAMU AUP 2022-0091.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Bond Life Sciences Center, University of Missouri, Columbia, MO, USA. [2]Faculty of Veterinary Medicine, Ludwig-Maximilians-Universität, Munich, Germany. [3]Center for Evolutionary Hologenomics, Globe Institute, University of Copenhagen (UCPH), Copenhagen, Denmark. [4]Department of Ecology & Evolutionary Biology, Texas A&M University, College Station, TX, USA. [5]Department of Poultry Science, Texas A&M University, College Station, TX, USA. [6]Vertebrate Genome Laboratory, The Rockefeller University, New York, NY, USA. [7]Sigenae, GenPhySE, Université de Toulouse, INRAE, ENVT, Castanet Tolosan 31326, France. [8]Department of Biology, Texas A&M University, College Station, TX, USA. [9]University Paris-Saclay, INRAE, AgroParisTech, GABI, Sigenae, Jouy-en-Josas, France. [10]Avian Disease and Oncology Laboratory, USDA, ARS, USNPRC, East Lansing, MI, USA. [11]Department of Biology, University of Oxford, Oxford OX1 3SZ, UK. [12]School of Biological and Behavioural Sciences, Queen Mary University of London, London E1 4DQ, UK. [13]The Howard Hughes Medical Institute, Chevy Chase, MD, USA. [14]Sigenae, Genotoul Bioinfo, MIAT UR875, INRAE, Castanet Tolosan, France. [15]Applied Genomics and Bioinformatics, University of the Basque Country (UPV/EHU), Leioa, Bilbao, Spain. [16]Department of Biology, The University of York, York, UK. [17]Key Laboratory of Freshwater Fish Reproduction and Development (Ministry of Education), Key Laboratory of Aquatic Science of Chongqing, School of Life Sciences, Southwest University, Chongqing 400715, China. [18]Department of Animal Sciences, University of Missouri, Columbia, MO, USA.

Received: 14 July 2023 Accepted: 2 November 2023
Published online: 22 November 2023

References

1. Athrey G. Chapter 18 - Poultry genetics and breeding. In: Bazer FW, Lamb GC, Wu G, editors. Animal agriculture. Cambridge: Academic Press; 2020. p. 317–30.
2. Drobik-Czwarno W, Wolc A, Fulton JE, Arango J, Jankowski T, O'Sullivan NP, et al. Identifying the genetic basis for resistance to avian influenza in commercial egg layer chickens. Animal. 2018;12:1363–71.
3. Xu L, He Y, Ding Y, Liu GE, Zhang H, Cheng HH, et al. Genetic assessment of inbred chicken lines indicates genomic signatures of resistance to Marek's disease. J Anim Sci Biotechnol. 2018;9:65.
4. Wang Q, Li D, Guo A, Li M, Li L, Zhou J, et al. Whole-genome resequencing of Dulong Chicken reveal signatures of selection. Br Poult Sci. 2020;61:624–31.
5. Seifi Moroudi R, Ansari Mahyari S, Vaez Torshizi R, Lanjanian H, Masoudi-Nejad A. Identification of new genes and quantitative trait locis associated with growth curve parameters in F2 chicken population using genome-wide association study. Anim Genet. 2021;52:171–84.
6. Perlas A, Argilaguet J, Bertran K, Sánchez-González R, Nofrarías M, Valle R, et al. Dual host and pathogen RNA-Seq analysis unravels chicken genes potentially involved in resistance to highly pathogenic avian influenza virus infection. Front Immunol. 2021;12:800188.
7. Zhao X, Collins RL, Lee W-P, Weber AM, Jun Y, Zhu Q, et al. Expectations and blind spots for structural variation detection from long-read assemblies and short-read genome sequencing technologies. Am J Hum Genet. 2021;108:919–28.
8. Mahmoud M, Gobet N, Cruz-Dávalos DI, Mounier N, Dessimoz C, Sedlazeck FJ. Structural variant calling: the long and the short of it. Genome Biol. 2019;20:246.
9. Sirén J, Monlong J, Chang X, Novak AM, Eizenga JM, Markello C, et al. Pangenomics enables genotyping of known structural variants in 5202 diverse genomes. Science. 2021;374:abg8871.
10. Audano PA, Sulovari A, Graves-Lindsay TA, Cantsilieris S, Sorensen M, Welch AE, et al. Characterizing the major structural variant alleles of the human genome. Cell. 2019;176:663-75.e19.
11. Nielsen R, Paul JS, Albrechtsen A, Song YS. Genotype and SNP calling from next-generation sequencing data. Nat Rev Genet. 2011;12:443–51.
12. Rhie A, McCarthy SA, Fedrigo O, Damas J, Formenti G, Koren S, et al. Towards complete and error-free genome assemblies of all vertebrate species. Nature. 2021;592:737–46.
13. Ko BJ, Lee C, Kim J, Rhie A, Yoo DA, Howe K, et al. Widespread false gene gains caused by duplication errors in genome assemblies. Genome Biol. 2022;23:205.
14. Brandt DYC, Aguiar VRC, Bitarello BD, Nunes K, Goudet J, Meyer D. Mapping bias overestimates reference allele frequencies at the HLA genes in the 1000 genomes project phase I data. G3. 2015;5:931–41.
15. Günther T, Nettelblad C. The presence and impact of reference bias on population genomic studies of prehistoric human populations. PLoS Genet. 2019;15:e1008302.
16. Barbitoff YA, Abasov R, Tvorogova VE, Glotov AS, Predeus AV. Systematic benchmark of state-of-the-art variant calling pipelines identifies major factors affecting accuracy of coding sequence variant discovery. BMC Genomics. 2022;23:155.
17. Van der Auwera GA, O'Connor BD. Genomics in the cloud: using Docker, GATK, and WDL in Terra. Sebastopol: O'Reilly Media, Inc.; 2020.
18. Poplin R, Chang P-C, Alexander D, Schwartz S, Colthurst T, Ku A, et al. A universal SNP and small-indel variant caller using deep neural networks. Nat Biotechnol. 2018;36:983–7.
19. De Coster W, Weissensteiner MH, Sedlazeck FJ. Towards population-scale long-read sequencing. Nat Rev Genet. 2021;22:572–87.
20. Rao YS, Li J, Zhang R, Lin XR, Xu JG, Xie L, et al. Copy number variation identification and analysis of the chicken genome using a 60K SNP BeadChip. Poult Sci. 2016;95:1750–6.
21. Garrison E, Sirén J, Novak AM, Hickey G, Eizenga JM, Dawson ET, et al. Variation graph toolkit improves read mapping by representing genetic variation in the reference. Nat Biotechnol. 2018;36:875–9.
22. Chen S, Krusche P, Dolzhenko E, Sherman RM, Petrovski R, Schlesinger F, et al. Paragraph: a graph-based structural variant genotyper for short-read sequence data. Genome Biol. 2019;20:1–13.
23. Li H, Feng X, Chu C. The design and construction of reference pangenome graphs with minigraph. Genome Biol. 2020;21:265.
24. Guarracino A, Heumos S, Nahnsen S, Prins P, Garrison E. ODGI: understanding pangenome graphs. Bioinformatics. 2022;38:3319–26.
25. Garrison E, Guarracino A. Unbiased pangenome graphs. Bioinformatics. 2022;39:btac743.
26. Eizenga JM, Novak AM, Sibbesen JA, Heumos S, Ghaffaari A, Hickey G, et al. Pangenome graphs. Annu Rev Genomics Hum Genet. 2020;21:139–62.
27. Liao W-W, Asri M, Ebler J, Doerr D, Haukness M, Hickey G, et al. A draft human pangenome reference. Nature. 2023;617:312–24.
28. Hickey G, Heller D, Monlong J, Sibbesen JA, Sirén J, Eizenga J, et al. Genotyping structural variants in pangenome graphs using the vg toolkit. Genome Biol. 2020;21:35.
29. Secomandi S, Gallo GR, Sozzoni M, Iannucci A, Galati E, Abueg L, et al. A chromosome-level reference genome and pangenome for barn swallow population genomics. Cell Rep. 2023;42:111992.
30. Rausch T, Zichner T, Schlattl A, Stütz AM, Benes V, Korbel JO. DELLY: structural variant discovery by integrated paired-end and split-read analysis. Bioinformatics. 2012;28:i333–9.
31. Wang K, Hu H, Tian Y, Li J, Scheben A, Zhang C, et al. The chicken pangenome reveals gene content variation and a promoter region deletion in IGF2BP1 affecting body size. Mol Biol Evol. 2021;38:5066–81.
32. Li M, Sun C, Xu N, Bian P, Tian X, Wang X, et al. De novo assembly of 20 chicken genomes reveals the undetectable phenomenon for thousands of core genes on microchromosomes and subtelomeric regions. Mol Biol Evol. 2022;39(4):msac066.

33. Smith J, Alfieri JM, Anthony N, Arensburger P, Athrey GN, Balacco J, et al. Fourth report on chicken genes and chromosomes 2022. Cytogenet Genome Res. 2023. https://doi.org/10.1159/000529376.
34. Huang Z, Xu Z, Bai H, Huang Y, Kang N, Ding X, et al. Evolutionary analysis of a complete chicken genome. Proc Natl Acad Sci U S A. 2023;120:e2216641120.
35. Cheng H, Jarvis ED, Fedrigo O, Koepfli K-P, Urban L, Gemmell NJ, et al. Haplotype-resolved assembly of diploid genomes without parental data. Nat Biotechnol. 2022;40:1332–5.
36. Hickey G, Monlong J, Ebler J, Novak AM, Eizenga JM, Gao Y, et al. Pangenome graph construction from genome alignments with Minigraph-Cactus. Nat Biotechnol. 2023. https://doi.org/10.1038/s41587-023-01793-w.
37. Li D, Li Y, Li M, Che T, Tian S, Chen B, et al. Population genomics identifies patterns of genetic diversity and selection in chicken. BMC Genomics. 2019;20:263.
38. Zhang J, Nie C, Li X, Zhao X, Jia Y, Han J, et al. Comprehensive analysis of structural variants in chickens using PacBio sequencing. Front Genet. 2022;13:971588.
39. Liu R, Xing S, Wang J, Zheng M, Cui H, Crooijmans RPMA, et al. A new chicken 55K SNP genotyping array. BMC Genomics. 2019;20:410.
40. Malomane DK, Simianer H, Weigend A, Reimer C, Schmitt AO, Weigend S. The SYNBREED chicken diversity panel: a global resource to assess chicken diversity at high genomic resolution. BMC Genomics. 2019;20:345.
41. Warren WC, Rice ES, Meyer A, Hearn CJ, Steep A, Hunt HD, et al. The immune cell landscape and response of Marek's disease resistant and susceptible chickens infected with Marek's disease virus. Sci Rep. 2023;13:5355.
42. Hertwig P, Rittershaus T. Die Erbfaktoren der Haushühner. Z Indukt Abstamm Vererbungsl. 1929;51:354–72.
43. Siegel PB, Mueller CD, Craig JV. Some phenotypic differences among homozygous, heterozygous, and hemizygous late feathering chicks 1,2. Poult Sci. 1957;36:232–9.
44. Elferink MG, Vallée AAA, Jungerius AP, Crooijmans RPMA, Groenen MAM. Partial duplication of the PRLR and SPEF2 genes at the late feathering locus in chicken. BMC Genomics. 2008;9:391.
45. Takenouchi A, Toshishige M, Ito N, Tsudzuki M. Endogenous viral gene ev21 is not responsible for the expression of late feathering in chickens. Poult Sci. 2018;97:403–11.
46. Zhang X, Wang H, Zhang L, Wang Q, Du X, Ge L, et al. Analysis of a genetic factors contributing to feathering phenotype in chickens. Poult Sci. 2018;97:3405–13.
47. Mason AS, Miedzinska K, Kebede A, Bamidele O, Al-Jumaili AS, Dessie T, et al. Diversity of endogenous avian leukosis virus subgroup E (ALVE) insertions in indigenous chickens. Genet Sel Evol. 2020;52:29.
48. Weiss RA. The discovery of endogenous retroviruses. Retrovirology. 2006;3:67.
49. Fulton JE, Mason AS, Wolc A, Arango J, Settar P, Lund AR, et al. The impact of endogenous Avian Leukosis Viruses (ALVE) on production traits in elite layer lines. Poult Sci. 2021;100:101121.
50. Chang C-M, Coville J-L, Coquerelle G, Gourichon D, Oulmouden A, Tixier-Boichard M. Complete association between a retroviral insertion in the tyrosinase gene and the recessive white mutation in chickens. BMC Genomics. 2006;7:19.
51. Hu X, Zhu W, Chen S, Liu Y, Sun Z, Geng T, et al. Expression patterns of endogenous avian retrovirus ALVE1 and its response to infection with exogenous avian tumour viruses. Arch Virol. 2017;162:89–101.
52. Fox W, Smyth JR Jr. The effects of recessive white and dominant white genotypes on early growth rate. Poult Sci. 1985;64:429–33.
53. Robinson HL, Astrin SM, Senior AM, Salazar FH. Host Susceptibility to endogenous viruses: defective, glycoprotein-expressing proviruses interfere with infections. J Virol. 1981;40:745–51.
54. Crittenden LB, Smith EJ, Fadly AM. Influence of endogenous viral (ev) gene expression and strain of exogenous avian leukosis virus (ALV) on mortality and ALV infection and shedding in chickens. Avian Dis. 1984;28:1037–56.
55. Jiang Y-F, Wang S, Wang C-L, Xu R-H, Wang W-W, Jiang Y, et al. Pangenome obtained by long-read sequencing of 11 genomes reveal hidden functional structural variants in pigs. iScience. 2023;26:106119.
56. Zhou Y, Yang L, Han X, Han J, Hu Y, Li F, et al. Assembly of a pangenome for global cattle reveals missing sequences and novel structural variations, providing new insights into their diversity and evolutionary history. Genome Res. 2022;32:1585–601.
57. Leonard AS, Crysnanto D, Mapel XM, Bhati M, Pausch H. Graph construction method impacts variation representation and analyses in a bovine super-pangenome. Genome Biol. 2023;24:124.
58. Bacon LD, Smith E, Crittenden LB, Havenstein GB. Association of the slow feathering (K) and an endogenous viral (ev21) gene on the Z chromosome of chickens. Poult Sci. 1988;67:191–7.
59. Nurk S, Koren S, Rhie A, Rautiainen M, Bzikadze AV, Mikheenko A, et al. The complete sequence of a human genome. Science. 2022;376:44–53.
60. Rautiainen M, Nurk S, Walenz BP, Logsdon GA, Porubsky D, Rhie A, et al. Telomere-to-telomere assembly of diploid chromosomes with Verkko. Nat Biotechnol. 2023. https://doi.org/10.1038/s41587-023-01662-6.
61. Xue L, Gao Y, Wu M, Tian T, Fan H, Huang Y, et al. Telomere-to-telomere assembly of a fish Y chromosome reveals the origin of a young sex chromosome pair. Genome Biol. 2021;22:203.
62. Belser C, Baurens F-C, Noel B, Martin G, Cruaud C, Istace B, et al. Telomere-to-telomere gapless chromosomes of banana using nanopore sequencing. Commun Biol. 2021;4:1047.
63. Bliznina A, Masunaga A, Mansfield MJ, Tan Y, Liu AW, West C, et al. Telomere-to-telomere assembly of the genome of an individual Oikopleura dioica from Okinawa using Nanopore-based sequencing. BMC Genomics. 2021;22:222.
64. Stokstad E. Deadly bird flu establishes a foothold in North America. Science. 2022;377:912.
65. Garrison E, Kronenberg ZN, Dawson ET, Pedersen BS, Prins P. A spectrum of free software tools for processing the VCF variant call format: vcflib, bio-vcf, cyvcf2, hts-nim and slivar. PLoS Comput Biol. 2022;18:e1009123.
66. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. 2011;17:10–2.
67. Cheng H, Concepcion GT, Feng X, Zhang H, Li H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nat Methods. 2021;18:170–5.
68. Guan D, McCarthy SA, Wood J, Howe K, Wang Y, Durbin R. Identifying and removing haplotypic duplication in primary genome assemblies. Bioinformatics. 2020;36:2896–8.
69. Ghurye J, Pop M, Koren S, Bickhart D, Chin C-S. Scaffolding of long read assemblies using long range contact information. BMC Genomics. 2017;18:527.
70. Chow W, Brugger K, Caccamo M, Sealy I, Torrance J, Howe K. gEVAL - a web-based browser for evaluating genome assemblies. Bioinformatics. 2016;32:2508–10.
71. Jain C, Koren S, Dilthey A, Phillippy AM, Aluru S. A fast adaptive algorithm for computing whole-genome homology maps. Bioinformatics. 2018;34:i748–56.
72. Marco-Sola S, Moure JC, Moreto M, Espinosa A. Fast gap-affine pairwise alignment using the wavefront algorithm. Bioinformatics. 2021;37(4):456–63.
73. smoothxg: linearize and simplify variation graphs using blocked partial order alignment. GitHub; 2023. https://github.com/pangenome/smoothxg.
74. GFAffix: GFAffix identifies walk-preserving shared affixes in variation graphs and collapses them into a non-redundant graph structure. GitHub; 2023. https://github.com/marschall-lab/GFAffix.
75. minigraph-cactus-nf: a nextflow pipeline for creating a pangenome with minigraph-cactus. GitHub; 2023. https://github.com/WarrenLab/minigraph-cactus-nf.
76. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. Gigascience. 2021;10(2):giab008.
77. Wick RR, Schultz MB, Zobel J, Holt KE. Bandage: interactive visualization of de novo genome assemblies. Bioinformatics. 2015;31:3350–2.
78. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215:403–10.
79. Pfeiffer F, Gröber C, Blank M, Händler K, Beyer M, Schultze JL, et al. Systematic evaluation of error rates and causes in short samples in next-generation sequencing. Sci Rep. 2018;8:10950.
80. Sirén J, Paten B. GBZ file format for pangenome graphs. Bioinformatics. 2022;38:5012–8.
81. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018;34:3094–100.
82. Herzeel C, Costanza P, Decap D, Fostier J, Wuyts R, Verachtert W. Multithreaded variant calling in elPrep 5. PLoS One. 2021;16:e0244471.

83. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 2015;4:s13742-015-0047–8.
84. Dementieva NV, Mitrofanova OV, Dysin AP, Kudinov AA, Stanishevskaya OI, Larkina TA, et al. Assessing the effects of rare alleles and linkage disequilibrium on estimates of genetic diversity in the chicken populations. Animal. 2021;15:100171.
85. Gallus gallus isolate:bGalGal4 | breed:Ross (chicken). BioProject; 2023. http://identifiers.org/bioproject:PRJNA838369.
86. Gallus gallus isolate:bGalGal5 | breed:Cobb (chicken). BioProject; 2023. http://identifiers.org/bioproject:PRJNA838370.
87. HiFi sequencing of chicken research lines. BioProject; 2023. http://identifiers.org/bioproject:PRJNA971225.
88. Gallus gallus Raw sequence reads. BioProject; 2023. http://identifiers.org/bioproject:PRJNA1031205.
89. Rice E, Alberdi A, Alfieri J, Athrey G, Balacco J, Bardou P, et al. A pangenome graph reference of 30 chicken genomes. Zenodo; 2023. https://doi.org/10.5281/zenodo.10018222.
90. chicken-pangenome-paper: scripts used to perform analyses in Rice et al. Github; 2023. https://github.com/WarrenLab/chicken-pangenome-paper.

Publisher's Note

Supplementary tables from Chapter 3

Annex 4

Table S1: Ingredient and nutrient composition of the basal diets (as fed basis).

	1st Experiment (A)			2nd Experiment (B)			3rd Experiment (C)		
Period (days)	**0-7**	**7-21**	**21-35**	**0-7**	**7-21**	**21-35**	**0-7**	**7-21**	**21-35**
Ingredient (g/kg)									
Wheat	52.75	61.24	61.97	52.31	60.52	61.05	54.3	62.19	62.15
Soybean 48%	39.4	30.5	15.8	39.88	31.24	16.79	38.02	29.66	15.75
Extruded Soyabean	-	-	15	-	-	15	-	-	15
Soyabean oil	4.16	4.8	-	4.18	4.85	-	4.01	4.71	-
Animal fat (5 Sysfeed)[1]	-	-	4.01	-	-	4.08	-	-	3.98
Dicalcium phosphate	1.84	1.66	1.5	1.84	1.66	1.5	1.86	1.67	1.5
Calcium carbonate	0.53	0.48	0.44	0.48	0.44	0.39	0.49	0.44	0.39
Vitamin-Mineral premix[2]	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4
Sodium chloride	0.37	0.37	0.53	0.37	0.38	0.36	0.37	0.35	0.35
DL-methionine	0.27	0.23	0.19	0.28	0.24	0.2	0.28	0.23	0.19
L-lysine HCL	0.16	0.19	0.15	0.15	0.17	0.14	0.16	0.18	0.15
Choline Chloride	0.03	0.05	0.05	0.03	0.05	0.05	0.05	0.06	0.06
L-threonine	0.05	0.05	0.04	0.05	0.05	0.04	0.05	0.05	0.04
Sodium bicarbonate	-	0.01	-	-	-	-	-	0.04	-
Antioxidant (Noxyfeed 56P)[3]	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
Nutrients (g/kg)									
Metabolizable Energy (Kcal/kg)	2900	3000	3100	2900	3000	3100	2900	3000	3100
Dry Matter	879	881	882	879	881	882	871	874	878
Crude Protein	255	223	206	241	209	194	241	210	195
Ether Extract	57.0	63.4	81.5	57.0	64.0	82.0	56.8	63.8	82.6

[1]Product of Sysfeed SLU (Granollers, Spain). It contains myristic acid (C14:0) 1.50%, palmitic acid (C16:0) 18.0%, palmitoleic acid (C16:1 n-7) 2.00%, stearic acid (C18:0) 14.0%, oleic acid (C18:1 n-9 cis) 28.0%, linoleic acid (C18:2 n-6 cis) 12.0%, α-linolenic acid (C18:3 n-3 cis) 6.00%, saturated–unsaturated 0.7%.

[2]Vitamin-Mineral premix: Product of TecnoVit S.L. (Alforja, Spain). Supplied per kilogram of feed: Vitamin A: 10 000 IU; Vitamin D3: 4 800 IU; Vitamin E: 45 mg; Vitamin K3: 3 mg; Vitamin B1: 3 mg; Vitamin B_2: 9 mg; Vitamin B_6: 4.5 mg: Vitamin B_{12}: 40 ug; Folic acid: 1.8 mg; Biotin: 150 ug; Calcium pantothenate: 16.5 mg; Niacin: 65 mg; Mn (as $MnSO_4.H_2O$): 90 mg; Zn (as ZnO): 66 mg; I (as KI): 1.2 mg; Fe (as $FeSO_4.H_2O$): 54 mg; Cu (as $CuSO_4.5H_20$): 12 mg; Se (as $NaSeO_3$): 0.18 mg; BHT: 25 mg; Calcium formiate, 5 mg; Silicic acid, dry and precipitated, 25 mg; Calcium stearate , 25 mg; Calcium carbonate to 4 g.

[3]Product of Itpsa (Barcelona, Spain). It contains 56% of antioxidant substances (butylated hydroxytoluene + propyl gallate) and synergistic (Citric acid 14% + authorised support).

Table S2: Performance 0-35 days[1].

			N	Initial BW (g)		Final BW (g)		ADG (g)		ADFI (g)		FCR (g/g)		Mortality (%)		EPEF	
				LSmeans	SE	LSmeans	SE	LSmeans	SE	LSmeans	SE	LSmeans	SE	LSmeans	SE	LSmeans	SE
Treatment (TRT)																	
BD			24	42.9	0.31	2063.3	16.86	57.7	0.48	87.9	0.73	1.523	0.013	2.64	0.469	369.1	5.37
PR			24	42.5	0.31	2042.0	16.84	57.1	0.48	87.0	0.73	1.523	0.013	1.75	0.468	368.6	5.37
PH			24	42.6	0.31	2060.0	16.80	57.6	0.48	87.1	0.73	1.512	0.013	2.82	0.467	371.2	5.36
p-value				*0.6932*		*0.6326*		*0.6323*		*0.6143*		*0.7210*		*0.2343*		*0.9356*	
Genetic line (GL)																	
X			36	44.5	0.26	2232.0	18.84	62.6	0.54	93.6	0.80	1.497	0.015	2.53	0.524	407.7	6.01
Y			36	40.8	0.26	1878.2	18.84	52.4	0.54	81.1	0.80	1.542	0.015	2.27	0.524	331.6	6.01
p-value				*<0.0001*		*<0.0001*		*<0.0001*		*<0.0001*		*0.0467*		*0.7754*		*<0.0001*	
Sex																	
F			36	42.4	0.26	1953.1	13.82	54.6	0.39	83.9	0.61	1.535	0.012	2.05	0.384	348.2	4.41
M			36	42.9	0.26	2157.1	13.82	60.4	0.39	90.7	0.61	1.503	0.012	2.76	0.384	391.2	4.41
p-value				*0.1786*		*<0.0001*		*<0.0001*		*<0.0001*		*0.0207*		*0.2003*		*<0.0001*	
Experiment (EXP)																	
A			24	41.3^{b}	0.31	2114.3^{a}	19.33	59.2^{a}	0.55	93.1^{a}	0.89	1.577^{a}	0.018	1.55	0.538	371.5	6.16
B			24	42.1^{b}	0.31	2064.6^{b}	17.20	57.8^{b}	0.49	88.1^{b}	0.81	1.524^{b}	0.017	2.41	0.478	368.3	5.49
C			24	44.5^{a}	0.31	1986.4^{c}	21.41	55.5^{c}	0.61	80.8^{c}	0.97	1.457^{c}	0.019	3.25	0.596	369.3	6.83
p-value				*<0.0001*		*0.0013*		*0.0013*		*<0.0001*		*0.0004*		*0.1753*		*0.9173*	
TRT	**GL**																
BD	X		12	45.0	0.44	2238.2	28.94	62.7	0.83	93.7	1.20	1.493	0.021	2.90	0.805	407.9	9.23
	Y		12	40.7	0.44	1888.4	27.41	52.7	0.78	82.1	1.14	1.552	0.020	2.39	0.762	330.3	8.74
PR	X		12	44.2	0.44	2217.7	26.02	62.1	0.74	93.6	1.09	1.508	0.019	1.08	0.724	407.9	8.30
	Y		12	40.8	0.44	1866.2	27.11	52.1	0.77	80.4	1.13	1.539	0.020	2.42	0.754	329.4	8.64
PH	X		12	44.3	0.44	2240.0	26.43	62.8	0.76	93.5	1.10	1.489	0.019	3.61	0.735	407.3	8.43
	Y		12	40.9	0.44	1880.0	26.61	52.5	0.76	80.7	1.11	1.534	0.019	2.02	0.740	335.2	8.49
p-value				*0.4816*		*0.9741*		*0.9735*		*0.6812*		*0.6819*		*0.0911*		*0.9009*	
TRT	**Sex**																
BD	F		12	42.6	0.44	1969.6	23.75	55.1	0.68	84.3	1.00	1.530	0.018	2.08	0.661	351.6	7.57
	M		12	43.1	0.44	2157.0	23.93	60.4	0.68	91.4	1.00	1.516	0.018	3.21	0.666	386.7	7.63
PR	F		12	42.2	0.44	1936.2	23.99	54.1	0.69	84.1	1.01	1.553	0.018	1.81	0.667	342.3	7.65
	M		12	42.8	0.44	2147.7	23.77	60.1	0.68	89.9	1.00	1.494	0.018	1.69	0.661	395.0	7.58
PH	F		12	42.4	0.44	1953.6	23.81	54.6	0.68	83.3	1.00	1.524	0.018	2.25	0.662	350.6	7.59
	M		12	42.8	0.44	2166.5	23.79	60.7	0.68	90.9	1.00	1.500	0.018	3.38	0.662	391.9	7.59
p-value				*0.9718*		*0.8345*		*0.8347*		*0.6253*		*0.3558*		*0.5551*		*0.5074*	
GL	**Sex**																
X	F		18	44.5	0.36	2119.4	23.21	59.3	0.66	90.4	0.98	1.523	0.017	2.80^{a}	0.646	379.3^{b}	7.40
	M		18	44.5	0.36	2344.6	23.39	65.8	0.67	96.8	0.98	1.471	0.017	2.26^{a}	0.651	436.2^{a}	7.46
Y	F		18	40.3	0.36	1786.8	25.28	49.8	0.72	77.5	1.06	1.547	0.019	1.30^{b}	0.703	317.0^{d}	8.06
	M		18	41.3	0.36	1969.6	21.64	55.1	0.62	84.7	0.91	1.536	0.016	3.25^{a}	0.602	346.2^{c}	6.90
p-value				*0.2232*		*0.2852*		*0.2850*		*0.6351*		*0.1323*		*0.0264*		*0.0313*	
GL	**Sex**	**TRT**															
X	F	BD	6	45.2	0.63	2134.4	38.02	59.8	1.09	90.3	1.57	1.510	0.0268	3.34	1.057	383.0	12.12
		PR	6	44.2	0.63	2097.1	35.21	58.7	1.01	90.5	1.46	1.543	0.0249	1.92	0.979	373.7	11.23
		PH	6	44.1	0.63	2126.8	35.01	59.5	1.00	90.3	1.45	1.517	0.0248	3.15	0.974	381.1	11.16
	M	BD	6	44.8	0.63	2342.0	36.90	65.7	1.05	97.0	1.52	1.477	0.0260	2.46	1.026	432.9	11.76
		PR	6	44.2	0.63	2338.4	35.25	65.6	1.01	96.7	1.46	1.474	0.0250	0.25	0.980	442.1	11.24
		PH	6	44.5	0.63	2353.3	36.13	66.0	1.03	96.6	1.49	1.462	0.0255	4.07	1.005	433.5	11.52
Y	F	BD	6	40.1	0.63	1804.7	38.10	50.3	1.09	78.4	1.57	1.549	0.0268	0.83	1.060	320.1	12.15
		PR	6	40.2	0.63	1775.4	37.80	49.5	1.08	77.7	1.56	1.562	0.0266	1.71	1.051	310.9	12.02
		PH	6	40.8	0.63	1780.4	36.14	49.6	1.03	76.3	1.49	1.531	0.0256	1.35	1.005	320.1	11.53
	M	BD	6	41.3	0.63	1972.0	34.87	55.1	1.00	85.9	1.44	1.555	0.0247	3.95	0.970	340.6	11.12
		PR	6	41.4	0.63	1957.1	34.72	54.7	0.99	83.1	1.44	1.515	0.0246	3.13	0.966	347.8	11.07
		PH	6	41.1	0.63	1979.6	35.24	55.3	1.01	85.1	1.46	1.538	0.0250	2.69	0.980	350.3	11.24
p-value				*0.6113*		*0.9439*		*0.9439*		*0.6821*		*0.8117*		*0.3836*		*0.9477*	

[1]Values are presented as least squares means (LSmeans). [a–d]Within a column, values without a common superscript differ, P < 0.05. X and Y correspond to two fast growing commercial poultry genetic lines.
SE: Sta

Table S3: Performance 0-7 days[1].

			N	Initial BW (g)		Final BW (g)		ADG (g)		ADFI (g)		FCR (g/g)		Mortality (%)		EPEF	
				LSmeans	SE	LSmeans	SE	LSmeans	SE	LSmeans	SE	LSmeans	SE	LSmeans	SE	LSmeans	SE
Treatment (TRT)																	
BD			24	42.9	0.31	178.4	0.91	19.4	0.13	20.4	0.18	1.056	0.008	0.93	0.329	182.9	1.95
PR			24	42.5	0.31	179.4	0.91	19.5	0.13	20.3	0.18	1.043	0.008	0.84	0.329	186.5	1.95
PH			24	42.6	0.31	178.9	0.90	19.5	0.13	20.3	0.18	1.050	0.008	1.36	0.328	183.9	1.94
p-value				*0.6932*		*0.7366*		*0.7368*		*0.8873*		*0.4955*		*0.4894*		*0.4025*	
Genetic line (GL)																	
X			36	44.5	0.26	182.0	1.01	19.9	0.14	20.5	0.20	1.035	0.009	0.76	0.368	192.3	2.18
Y			36	40.8	0.26	175.9	1.01	19.0	0.14	20.2	0.20	1.064	0.009	1.33	0.368	176.6	2.18
p-value				*<0.0001*		*0.0007*		*0.0007*		*0.4283*		*0.0652*		*0.3691*		*<0.0001*	
Sex																	
F			36	42.4	0.26	177.5	0.75	19.3	0.11	20.5	0.15	1.068	0.007	0.91	0.270	179.9	1.6
M			36	42.9	0.26	180.3	0.75	19.7	0.11	20.2	0.15	1.031	0.007	1.18	0.270	189.0	1.6
p-value				*0.1786*		*0.0096*		*0.0096*		*0.1834*		*0.0003*		*0.4999*		*0.0002*	
Experiment (EXP)																	
A			24	41.3[b]	0.31	166.0[c]	1.07	17.6[c]	0.15	19.4[c]	0.22	1.100[a]	0.009	0.58	0.377	158.8[c]	2.24
B			24	42.1[b]	0.31	175.2[b]	0.97	18.9[b]	0.14	20.4[b]	0.20	1.078[a]	0.008	0.96	0.336	174.0[b]	1.99
C			24	44.5[a]	0.31	195.5[a]	1.18	21.8[a]	0.17	21.2[a]	0.24	0.971[b]	0.01	1.6	0.418	220.6[a]	2.48
p-value				*<0.0001*		*<0.0001*		*<0.0001*		*<0.0001*		*<0.0001*		*0.2903*		*<0.0001*	
TRT		**GL**															
BD		X	12	45.0	0.44	181.8	1.53	19.9	0.22	20.7	0.30	1.046	0.014	0.74	0.565	190.4	3.35
		Y	12	40.7	0.44	175.0	1.45	18.9	0.21	20.2	0.29	1.066	0.013	1.12	0.535	175.3	3.17
PR		X	12	44.2	0.44	181.3	1.38	19.8	0.2	20.4	0.27	1.036	0.012	0.15	0.508	192.5	3.01
		Y	12	40.8	0.44	177.5	1.44	19.3	0.21	20.2	0.28	1.05	0.013	1.54	0.529	180.6	3.14
PH		X	12	44.3	0.44	182.7	1.40	20.0	0.2	20.4	0.28	1.023	0.013	1.40	0.516	194.0	3.06
		Y	12	40.9	0.44	175.0	1.41	18.9	0.2	20.3	0.28	1.076	0.013	1.33	0.519	173.7	3.08
p-value				*0.4816*		*0.2687*		*0.2677*		*0.6184*		*0.1925*		*0.2820*		*0.3125*	
TRT		**Sex**															
BD		F	12	42.6	0.44	178.2	1.26	19.4	0.18	20.8	0.25	1.077	0.011	1.25	0.464	178.6	2.75
		M	12	43.1	0.44	178.7	1.27	19.4	0.18	20.1	0.25	1.035	0.011	0.61	0.467	187.2	2.77
PR		F	12	42.2	0.44	177.1	1.27	19.2	0.18	20.2	0.25	1.053	0.011	0.65	0.468	182.3	2.78
		M	12	42.8	0.44	181.7	1.26	19.9	0.18	20.4	0.25	1.032	0.011	1.04	0.464	190.7	2.75
PH		F	12	42.4	0.44	177.2	1.26	19.2	0.18	20.5	0.25	1.073	0.011	0.84	0.465	178.7	2.76
		M	12	42.8	0.44	180.5	1.26	19.7	0.18	20.1	0.25	1.026	0.011	1.88	0.464	189.0	2.75
p-value				*0.9718*		*0.2511*		*0.2516*		*0.1355*		*0.4740*		*0.1970*		*0.9276*	
GL		**Sex**															
X		F	18	44.5	0.36	180.9	1.23	19.7	0.18	20.6	0.24	1.050	0.011	0.90	0.453	188.5	2.69
		M	18	44.5	0.36	183.0	1.24	20.1	0.18	20.4	0.25	1.020	0.011	0.62	0.457	196.1	2.71
Y		F	18	40.3	0.36	174.2	1.34	18.8	0.19	20.4	0.26	1.085	0.012	0.93	0.493	171.2	2.92
		M	18	41.3	0.36	177.6	1.15	19.3	0.16	20.1	0.23	1.042	0.010	1.73	0.422	181.9	2.50
p-value				*0.2232*		*0.5397*		*0.5377*		*0.7905*		*0.4562*		*0.1666*		*0.5135*	
GL	**Sex**	**TRT**															
		BD	6	45.2	0.63	182.1	2.00	19.9	0.29	20.9	0.39	1.059	0.0179	1.56	0.742	187.7	4.40
	F	PR	6	44.2	0.63	179.0	1.86	19.5	0.27	20.2	0.36	1.044	0.0166	0.36	0.687	187.8	4.07
		PH	6	44.1	0.63	181.6	1.85	19.8	0.26	20.7	0.36	1.046	0.0165	0.78	0.683	189.9	4.05
X		BD	6	44.8	0.63	181.6	1.94	19.8	0.28	20.4	0.38	1.033	0.0174	-0.09	0.720	193.2	4.27
	M	PR	6	44.2	0.63	183.6	1.86	20.1	0.27	20.6	0.36	1.027	0.0166	-0.06	0.688	197.1	4.08
		PH	6	44.5	0.63	183.9	1.90	20.2	0.27	20.1	0.37	1.001	0.0170	2.02	0.705	198.1	4.18
		BD	6	40.1	0.63	174.3	2.01	18.8	0.29	20.6	0.39	1.095	0.0180	0.94	0.744	169.4	4.41
	F	PR	6	40.2	0.63	175.3	1.99	18.9	0.28	20.1	0.39	1.062	0.0178	0.94	0.738	176.8	4.37
		PH	6	40.8	0.63	172.9	1.91	18.6	0.27	20.4	0.37	1.100	0.0170	0.91	0.705	167.5	4.18
Y		BD	6	41.3	0.63	175.7	1.84	19.0	0.26	19.7	0.36	1.037	0.0164	1.30	0.681	181.2	4.03
	M	PR	6	41.4	0.63	179.8	1.83	19.6	0.26	20.3	0.36	1.038	0.0164	2.13	0.678	184.3	4.02
		PH	6	41.1	0.63	177.2	1.86	19.2	0.27	20.2	0.36	1.052	0.0166	1.75	0.688	180.0	4.08
p-value				*0.6113*		*0.8955*		*0.8942*		*0.7976*		*0.7934*		*0.3898*		*0.7461*	

Table S4: Performance 7-21 days[1].

			N	Initial BW(g)		Final BW (g)		ADG (g)		ADFI (g)		FCR (g/g)		Mortality (%)		EPEF	
				LSmeans	SE	LSmeans	SE	LSmeans	SE	LSmeans	SE	LSmeans	SE	LSmeans	SE	LSmeans	SE
Treatment (TRT)																	
BD			24	178.4	0.91	898.1	5.01	51.4	0.32	70.2	1.06	1.367	0.0205	0.79	0.291	374.9	5.79
PR			24	179.4	0.91	890.6	5.01	50.8	0.32	70.1	1.06	1.380	0.0205	0.38	0.291	368.7	5.79
PH			24	178.9	0.90	897.6	4.99	51.3	0.32	69.4	1.06	1.353	0.0204	0.74	0.291	378.0	5.78
p-value				*0.7366*		*0.5002*		*0.3502*		*0.7723*		*0.4920*		*0.5408*		*0.3937*	
Genetic line (GL)																	
X			36	182.0	1.01	970.2	5.60	56.3	0.36	75.8	1.14	1.341	0.0220	0.96	0.323	415.3	6.29
Y			36	175.9	1.01	820.7	5.60	46.1	0.36	64.0	1.14	1.392	0.0220	0.32	0.323	332.4	6.29
p-value				*0.0007*		*<0.0001*		*<0.0001*		*<0.0001*		*0.1035*		*0.2359*		*<0.0001*	
Sex																	
F			36	177.5	0.75	869.5	4.11	49.4	0.26	67.3	0.95	1.363	0.0183	0.51	0.244	362.7	5.07
M			36	180.3	0.75	921.3	4.11	52.9	0.26	72.4	0.95	1.371	0.0183	0.77	0.244	385.0	5.07
p-value				*0.0096*		*<0.0001*		*<0.0001*		*<0.0001*		*0.6784*		*0.4287*		*0.0003*	
Experiment (EXP)																	
A			24	166.0^{c}	1.07	868.1^{c}	5.75	50.1^{b}	0.37	65.5^{b}	1.49	1.309	0.0287	0.08	0.352	384.7	7.78
B			24	175.2^{b}	0.97	898.4^{b}	5.11	51.7^{a}	0.33	71.8^{a}	1.42	1.390	0.0274	1.02	0.320	368.0	7.35
C			24	195.5^{a}	1.18	919.7^{a}	6.37	51.7^{a}	0.41	72.4^{a}	1.56	1.402	0.0300	0.81	0.384	368.9	8.23
p-value				*<0.0001*		*<0.0001*		*0.0055*		*0.0036*		*0.0651*		*0.1263*		*0.2474*	
TRT	**GL**																
BD	X		12	181.8	1.53	974.9	8.61	56.6	0.55	75.7	1.57	1.330	0.0304	1.01	0.486	419.8	8.95
	Y		12	175.0	1.45	821.3	8.15	46.2	0.52	64.6	1.50	1.404	0.0291	0.56	0.461	330.1	8.54
PR	X		12	181.3	1.38	962.9	7.74	55.8	0.49	76.7	1.44	1.371	0.0279	0.59	0.439	406.0	8.16
	Y		12	177.5	1.44	818.3	8.06	45.8	0.52	63.4	1.49	1.390	0.0288	0.18	0.457	331.3	8.46
PH	X		12	182.7	1.40	972.8	7.86	56.4	0.50	74.9	1.46	1.323	0.0282	1.28	0.445	420.2	8.27
	Y		12	175.0	1.41	822.4	7.91	46.2	0.51	63.9	1.46	1.383	0.0284	0.20	0.448	335.7	8.32
p-value				*0.2687*		*0.8137*		*0.8897*		*0.5361*		*0.4683*		*0.6329*		*0.5525*	
TRT	**Sex**																
BD	F		12	178.2	1.26	874.5	7.06	49.7	0.45	66.8	1.34	1.343	0.0259	0.42	0.402	370.1	7.55
	M		12	178.7	1.27	921.7	7.12	53.1	0.46	73.6	1.35	1.391	0.0261	1.15	0.405	379.7	7.60
PR	F		12	177.1	1.27	863.8	7.13	49.0	0.46	68.6	1.35	1.399	0.0262	0.30	0.406	352.2	7.62
	M		12	181.7	1.26	917.4	7.07	52.6	0.45	71.5	1.34	1.361	0.0260	0.47	0.402	385.1	7.56
PH	F		12	177.2	1.26	870.3	7.08	49.5	0.45	66.6	1.34	1.346	0.0260	0.81	0.403	365.7	7.57
	M		12	180.5	1.26	924.8	7.07	53.2	0.45	72.2	1.34	1.360	0.0260	0.68	0.403	390.2	7.56
p-value				*0.2511*		*0.8547*		*0.9421*		*0.2427*		*0.1737*		*0.5425*		*0.2381*	
GL	**Sex**																
X	F		18	180.9	1.23	939.6	6.90	54.2	0.44	72.9	1.32	1.340	0.0255	1.17^{a}	0.393	400.5	7.42
	M		18	183.0	1.24	1000.7	6.96	58.4	0.44	78.7	1.33	1.343	0.0257	0.74^{ab}	0.396	430.2	7.47
Y	F		18	174.2	1.34	799.4	7.52	44.7	0.48	61.8	1.41	1.386	0.0273	-0.16^{b}	0.427	324.9	7.97
	M		18	177.6	1.15	841.9	6.44	47.5	0.41	66.2	1.25	1.399	0.0242	0.79^{a}	0.368	339.9	7.01
p-value				*0.5397*		*0.1163*		*0.0617*		*0.4591*		*0.7983*		*0.0390*		*0.2013*	
GL	**Sex**	**TRT**															
X	F	BD	6	182.1	2.00	947.1	11.30	54.6	0.72	72.1	1.98	1.314	0.0384	1.05	0.635	410.5	11.46
		PR	6	179.0	1.86	925.8	10.47	53.3	0.67	74.3	1.85	1.388	0.0359	0.79	0.589	383.6	10.67
		PH	6	181.6	1.85	945.9	10.41	54.6	0.67	72.1	1.84	1.318	0.0357	1.67	0.586	407.3	10.62
	M	BD	6	181.6	1.94	1002.7	10.97	58.7	0.70	79.3	1.93	1.347	0.0374	0.96	0.616	429.0	11.14
		PR	6	183.6	1.86	999.9	10.48	58.3	0.67	79.2	1.85	1.353	0.0359	0.38	0.589	428.4	10.68
		PH	6	183.9	1.90	999.6	10.74	58.3	0.69	77.7	1.89	1.328	0.0367	0.89	0.604	433.1	10.93
Y	F	BD	6	174.3	2.01	801.8	11.33	44.8	0.72	61.4	1.98	1.373	0.0385	-0.21	0.636	329.8	11.48
		PR	6	175.3	1.99	801.7	11.24	44.7	0.72	62.9	1.97	1.411	0.0382	-0.20	0.631	320.8	11.39
		PH	6	172.9	1.91	794.7	10.75	44.4	0.69	61.0	1.89	1.374	0.0367	-0.06	0.604	324.1	10.93
	M	BD	6	175.7	1.84	840.7	10.37	47.5	0.66	67.9	1.83	1.434	0.0356	1.34	0.583	330.4	10.58
		PR	6	179.8	1.83	834.9	10.32	46.8	0.66	63.9	1.83	1.369	0.0354	0.56	0.581	341.8	10.53
		PH	6	177.2	1.86	850.0	10.48	48.1	0.67	66.8	1.85	1.392	0.0359	0.46	0.589	347.3	10.68
p-value				*0.8955*		*0.3292*		*0.2888*		*0.6690*		*0.9266*		*0.9532*		*0.7353*	

[1]Values are presented as least squares means (LSmeans). [a–d]Within a column, values without a common superscript differ, P < 0.05. X and Y correspond to two fast growing commercial poultry genetic lines.
SE: Standard error. BD: Basal diet; PR: Basal diet + probiotic; PH: Basal diet + phytobiotic. F: female; M: male. BW: Body weight; ADG: Average daily gain; ADFI: Average daily feed intake; FCR: Feed conversion ratio; EPEF: European production efficiency factor.

Table S5: Performance 21-35 days[1].

			N	Initial BW (g)		Final BW (g)		ADG (g)		ADFI (g)		FCR (g/g)		Mortality (%)		EPEF	
				LSmeans	SE	LSmeans	SE	LSmeans	SE	LSmeans	SE	LSmeans	SE	LSmeans	SE	LSmeans	SE
Treatment (TRT)																	
BD			24	898.1	5.01	2063.3	16.86	83.2	1.06	143.3	1.65	1.719	0.022	1.15	0.508	480.1	10.53
PR			24	890.6	5.01	2042.0	16.84	82.2	1.06	140.9	1.64	1.711	0.022	0.69	0.507	480.1	10.52
PH			24	897.6	4.99	2060.0	16.80	83.0	1.06	142.0	1.64	1.710	0.022	0.90	0.507	482.9	10.50
p-value				*0.5002*		*0.6326*		*0.7787*		*0.6045*		*0.9385*		*0.7090*		*0.9757*	
Genetic line (GL)																	
X			36	970.2	5.60	2232.0	18.84	90.1	1.18	151.9	1.84	1.691	0.024	1.04	0.545	531.4	11.77
Y			36	820.7	5.60	1878.2	18.84	75.5	1.18	132.3	1.84	1.736	0.024	0.78	0.545	430.7	11.77
p-value				*<0.0001*		*<0.0001*		*<0.0001*		*<0.0001*		*0.2331*		*0.7358*		*<0.0001*	
Sex																	
F			36	869.5	4.11	1953.1	13.82	77.4	0.88	136.1	1.35	1.750	0.019	0.80	0.455	439.6	8.64
M			36	921.3	4.11	2157.1	13.82	88.3	0.88	148.0	1.35	1.677	0.019	1.02	0.455	522.5	8.64
p-value				*<0.0001*		*<0.0001*		*<0.0001*		*<0.0001*		*0.0023*		*0.6295*		*<0.0001*	
Experiment (EXP)																	
A			24	868.1^{c}	5.75	2114.3^{a}	19.33	89.0^{a}	1.26	162.4^{a}	1.89	1.834^{a}	0.029	1.04	0.717	486.2	12.08
B			24	898.4^{b}	5.11	2064.6^{b}	17.20	83.3^{b}	1.13	141.7^{b}	1.68	1.700^{b}	0.027	0.53	0.687	487.3	10.75
C			24	919.7^{a}	6.37	1986.4^{c}	21.41	76.2^{c}	1.38	122.2^{c}	2.09	1.607^{c}	0.030	1.17	0.748	469.5	13.38
p-value				*<0.0001*		*0.0013*		*<0.0001*		*<0.0001*		*<0.0001*		*0.7893*		*0.6146*	
TRT	**GL**																
BD	X		12	974.9	8.61	2238.2	28.94	90.2	1.8	152.3	2.83	1.694	0.035	1.44	0.748	529.1	18.08
	Y		12	821.3	8.15	1888.4	27.41	76.2	1.71	134.3	2.68	1.744	0.033	0.86	0.715	431.0	17.13
PR	X		12	962.9	7.74	2217.7	26.02	89.6	1.62	150.7	2.54	1.687	0.032	0.47	0.686	533.1	16.26
	Y		12	818.3	8.06	1866.2	27.11	74.9	1.69	131.2	2.65	1.736	0.033	0.91	0.709	427.1	16.94
PH	X		12	972.8	7.86	2240.0	26.43	90.5	1.65	152.7	2.58	1.693	0.032	1.21	0.695	531.9	16.51
	Y		12	822.4	7.91	1880.0	26.61	75.5	1.66	131.4	2.60	1.728	0.032	0.59	0.699	433.9	16.63
p-value				*0.8137*		*0.9741*		*0.9425*		*0.7739*		*0.9494*		*0.5563*		*0.9534*	
TRT	**Sex**																
BD	F		12	874.5	7.06	1969.6	23.75	78.2	1.48	137.7	2.32	1.751	0.029	0.49	0.640	444.8	14.84
	M		12	921.7	7.12	2157.0	23.93	88.2	1.49	148.9	2.34	1.688	0.029	1.81	0.644	515.4	14.95
PR	F		12	863.8	7.13	1936.2	23.99	76.6	1.50	135.4	2.34	1.757	0.029	1.14	0.645	432.1	14.99
	M		12	917.4	7.07	2147.7	23.77	87.9	1.48	146.5	2.32	1.666	0.029	0.24	0.640	528.1	14.85
PH	F		12	870.3	7.08	1953.6	23.81	77.4	1.49	135.3	2.33	1.742	0.029	0.78	0.641	441.9	14.88
	M		12	924.8	7.07	2166.5	23.79	88.7	1.48	148.8	2.32	1.679	0.029	1.02	0.641	524.0	14.86
p-value				*0.8547*		*0.8345*		*0.8804*		*0.8525*		*0.8411*		*0.1448*		*0.6938*	
GL	**Sex**																
X	F		18	939.6	6.90	2119.4	23.21	84.3	1.45	147.0	2.27	1.745	0.029	0.93	0.630	481.1	14.50
	M		18	1000.7	6.96	2344.6	23.39	96.0	1.46	156.9	2.29	1.638	0.029	1.15	0.633	581.7	14.62
Y	F		18	799.4	7.52	1786.8	25.28	70.5	1.58	125.3	2.47	1.755	0.031	0.67	0.672	398.1	15.80
	M		18	841.9	6.44	1969.6	21.64	80.6	1.35	139.2	2.11	1.717	0.027	0.90	0.598	463.3	13.52
p-value				*0.1163*		*0.2852*		*0.4861*		*0.2958*		*0.1325*		*0.9926*		*0.1548*	
GL	**Sex**	**TRT**															
X	F	BD	6	947.1	11.30	2134.4	38.02	84.8	2.35	147.5	3.71	1.740	0.045	0.86	0.943	485.6	23.76
		PR	6	925.8	10.47	2097.1	35.21	83.7	2.18	145.7	3.44	1.741	0.042	1.00	0.881	478.1	22.00
		PH	6	945.9	10.41	2126.8	35.01	84.3	2.17	147.8	3.42	1.753	0.042	0.93	0.877	479.5	21.88
	M	BD	6	1002.7	10.97	2342.0	36.90	95.7	2.29	157.1	3.60	1.648	0.044	2.02	0.918	572.6	23.05
		PR	6	999.9	10.48	2338.4	35.25	95.6	2.18	155.8	3.44	1.632	0.042	-0.07	0.882	588.1	22.02
		PH	6	999.6	10.74	2353.3	36.13	96.7	2.24	157.7	3.53	1.633	0.043	1.50	0.901	584.4	22.57
Y	F	BD	6	801.8	11.33	1804.7	38.10	71.6	2.36	127.8	3.72	1.761	0.045	0.12	0.944	404.0	23.81
		PR	6	801.7	11.24	1775.4	37.80	69.6	2.34	125.2	3.69	1.773	0.045	1.27	0.938	386.0	23.62
		PH	6	794.7	10.75	1780.4	36.14	70.4	2.24	122.9	3.53	1.731	0.043	0.63	0.902	404.2	22.58
	M	BD	6	840.7	10.37	1972.0	34.87	80.8	2.16	140.7	3.41	1.727	0.041	1.60	0.874	458.1	21.79
		PR	6	834.9	10.32	1957.1	34.72	80.2	2.15	137.2	3.39	1.700	0.041	0.55	0.870	468.1	21.70
		PH	6	850.0	10.48	1979.6	35.24	80.7	2.18	139.8	3.44	1.724	0.042	0.55	0.882	463.6	22.02
p-value				*0.3292*		*0.9439*		*0.9922*		*0.8588*		*0.7714*		*0.8769*		*0.9561*	

Table S6: Concentration of corticosterone in feathers (pg/mg)[1].

			N	Corticosterone 0-35 days LSmeans	SE	N	Corticosterone day 7 LSmeans	SE	Corticosterone day 21 LSmeans	SE	Corticosterone day 35 LSmeans	SE
Treatment (TRT)												
BD			72	12.18	0.962	24	9.80[b]	0.625	11.81	1.129	14.94	1.749
PR			72	13.09	0.962	24	10.59[ab]	0.625	11.43	1.129	17.27	1.749
PH			72	12.89	0.962	24	11.59[a]	0.625	12.77	1.129	14.31	1.749
p-value				*0.4325*			*0.0347*		*0.4555*		*0.2032*	
Genetic line (GL)												
X			108	13.91	0.914	36	9.91	0.562	13.66	1.037	18.15	1.601
Y			108	11.54	0.914	36	11.41	0.562	10.35	1.037	12.86	1.601
p-value				*0.0001*			*0.0082*		*0.0005*		*0.0004*	
Sex												
F			108	11.68	0.914	36	9.73	0.562	11.40	1.037	13.92	1.601
M			108	13.76	0.914	36	11.59	0.562	12.61	1.037	17.09	1.601
p-value				*0.0008*			*0.0013*		*0.1811*		*0.0281*	
Experiment (EXP)												
A			72	13.08	1.496	24	12.62[a]	0.850	11.67	1.621	14.96	2.491
B			72	13.95	1.496	24	11.13[a]	0.850	14.25	1.621	16.46	2.491
C			72	11.14	1.496	24	8.22[b]	0.850	10.08	1.621	15.10	2.491
p-value				*0.3989*			*0.0021*		*0.1950*		*0.8958*	
TRT		**GL**										
BD		X	36	13.41	1.093	12	8.60	0.785	13.46	1.369	18.17	2.131
		Y	36	10.96	1.093	12	11.00	0.785	10.15	1.369	11.71	2.131
PR		X	36	14.22	1.093	12	10.26	0.785	12.22	1.369	20.18	2.131
		Y	36	11.97	1.093	12	10.91	0.785	10.64	1.369	14.36	2.131
PH		X	36	14.09	1.093	12	10.86	0.785	15.29	1.369	16.11	2.131
		Y	36	11.69	1.093	12	12.32	0.785	10.25	1.369	12.50	2.131
P-value				*0.9890*			*0.4363*		*0.2944*		*0.6880*	
TRT		**Sex**										
BD		F	36	10.53	1.093	12	9.00	0.785	9.26[b]	1.369	13.34	2.131
		M	36	13.84	1.093	12	10.61	0.785	14.36[a]	1.369	16.54	2.131
PR		F	36	12.15	1.093	12	9.27	0.785	11.57[ab]	1.369	15.60	2.131
		M	36	14.04	1.093	12	11.91	0.785	11.28[ab]	1.369	18.93	2.131
PH		F	36	12.36	1.093	12	10.92	0.785	13.36[a]	1.369	12.81	2.131
		M	36	13.41	1.093	12	12.26	0.785	12.18[ab]	1.369	15.80	2.131
P-value				*0.3042*			*0.5939*		*0.0120*		*0.9949*	
GL		**Sex**										
X		F	54	12.37	1.008	18	9.85[b]	0.682	12.74	1.215	14.53	1.885
		M	54	15.44	1.008	18	9.96[b]	0.682	14.57	1.215	21.78	1.885
Y		F	54	10.99	1.008	18	9.61[b]	0.682	10.05	1.215	13.31	1.885
		M	54	12.09	1.008	18	13.21[a]	0.682	10.64	1.215	12.40	1.885
P-value				*0.1030*			*0.0023*		*0.4878*		*0.0053*	
GL	**Sex**	**TRT**										
X	F	BD	18	11.14	1.317	6	8.58	1.032	10.03	1.753	14.82	2.740
		PR	18	12.51	1.317	6	8.63	1.032	16.89	1.753	21.53	2.740
		PH	18	13.47	1.317	6	9.42	1.032	8.49	1.753	11.86	2.740
	M	BD	18	15.68	1.317	6	12.58	1.032	11.82	1.753	11.56	2.740
		PR	18	15.93	1.317	6	10.17	1.032	12.50	1.753	14.85	2.740
		PH	18	14.71	1.317	6	10.35	1.032	11.93	1.753	25.51	2.740
Y	F	BD	18	9.92	1.317	6	8.36	1.032	10.64	1.753	16.36	2.740
		PR	18	11.79	1.317	6	13.46	1.032	10.64	1.753	12.36	2.740
		PH	18	11.26	1.317	6	10.81	1.032	15.68	1.753	13.91	2.740
	M	BD	18	11.99	1.317	6	10.91	1.032	14.90	1.753	18.31	2.740

	PR	18	12.15	1.317	6	11.04	1.032	11.03	1.753	11.71	2.740
	PH	18	12.12	1.317	6	13.60	1.032	9.46	1.753	13.29	2.740
P-value				*0.6362*			*0.6353*		*0.6390*		*0.2286*
Day											
7		72	10.66[b]	0.96							
21		72	12.00[b]	0.96							
35		72	15.51[a]	0.96							
P-value				*<0.0001*							
TRT	**Day**										
	7	24	9.80	1.210							
BD	21	24	11.81	1.210							
	35	24	14.94	1.210							
	7	24	10.59	1.210							
PR	21	24	11.43	1.210							
	35	24	17.27	1.210							
	7	24	11.59	1.210							
PH	21	24	12.77	1.210							
	35	24	14.31	1.210							
P-value				*0.1203*							
GL	**Day**										
	7	36	9.91[d]	1.093							
X	21	36	13.66[b]	1.093							
	35	36	18.15[a]	1.093							
	7	36	11.41[cd]	1.093							
Y	21	36	10.35[d]	1.093							
	35	36	12.86[bc]	1.093							
P-value				*<0.0001*							
Sex	**Day**										
	7	36	9.73	1.093							
F	21	36	11.40	1.093							
	35	36	13.92	1.093							
	7	36	11.59	1.093							
M	21	36	12.61	1.093							
	35	36	17.09	1.093							
P-value				*0.3997*							

[1]Values are presented as least squares means (LSmeans). SE: Standard error. BD: Basal diet; PR: Basal diet + probiotic; PH: Basal diet + phytobiotic. F: female; M: male.

X and Y correspond to two fast growing commercial poultry genetic lines.
[a–d]Within a column, values without a common superscript differ, P < 0.05

Table S7a: Concentration of C-reactive protein (CRP), chicken haptoglobin-like protein (PIT54) and Lipopolysaccharides (LPS) in blood 0-35 days[1].

		N	CRP (pg/mL)		PIT54 (ng/mL)		LPS (ng/L)	
			LSmeans	SE	LSmeans	SE	LSmeans	SE
Treatment (TRT)								
BD		72	2.40	0.18	1.01	0.15	25.43	2.72
PR		72	2.57	0.18	1.11	0.15	24.79	2.72
PH		72	2.47	0.18	1.26	0.15	25.20	2.72
p-value			*0.8031*		*0.4608*		*0.9258*	
Genetic line (GL)								
X		108	2.55	0.15	1.31	0.12	25.42	2.64
Y		108	2.41	0.15	0.95	0.12	24.86	2.64
p-value			*0.4871*		*0.0354*		*0.6718*	
Sex								
F		108	2.44	0.15	1.09	0.12	25.18	2.64
M		108	2.51	0.15	1.17	0.12	25.10	2.64
p-value			*0.7346*		*0.6528*		*0.9466*	
Day								
7		72	3.27[a]	0.18	2.10[a]	0.15	29.25[a]	2.72
21		72	2.92[a]	0.18	0.59[b]	0.15	19.23[b]	2.72
35		72	1.24[b]	0.18	0.70[b]	0.15	26.94[a]	2.72
p-value			*<0.0001*		*<0.0001*		*<0.0001*	
Experiment (EXP)								
A		72	1.28[c]	0.19	1.07[b]	0.15	15.06[b]	4.43
B		72	2.06[b]	0.19	1.71[a]	0.15	35.15[a]	4.43
C		72	4.10[a]	0.19	0.61[c]	0.15	25.21[ab]	4.43
p-value			*<0.0001*		*<0.0001*		*0.0073*	
TRT	**GL**							
1	X	36	2.61	0.25	1.11	0.21	26.69	2.96
	Y	36	2.19	0.25	0.90	0.21	24.17	2.96
2	X	36	2.51	0.25	1.39	0.21	24.58	2.96
	Y	36	2.62	0.25	0.84	0.21	25.01	2.96
3	X	36	2.52	0.25	1.42	0.21	25.00	2.96
	Y	36	2.41	0.25	1.10	0.21	25.39	2.96
p-value			*0.5744*		*0.7157*		*0.5851*	
TRT	**Sex**							
BD	F	36	2.26	0.25	1.10	0.21	24.50	2.96
	M	36	2.54	0.25	0.91	0.21	26.35	2.96
PR	F	36	2.50	0.25	0.95	0.21	26.28	2.96
	M	36	2.63	0.25	1.27	0.21	23.31	2.96
PH	F	36	2.56	0.25	1.22	0.21	24.78	2.96
	M	36	2.37	0.25	1.31	0.21	25.62	2.96
p-value			*0.6451*		*0.4737*		*0.3029*	
GL	**Sex**							
X	F	54	2.62	0.21	1.23	0.17	25.41	2.8
	M	54	2.48	0.21	1.38	0.17	25.44	2.8
Y	F	54	2.27	0.21	0.95	0.17	24.96	2.8
	M	54	2.54	0.21	0.95	0.17	24.75	2.8
p-value			*0.3259*		*0.6622*		*0.9315*	
TRT	**Day**							
BD	7	24	3.09	0.31	1.74	0.25	29.73	3.18
	21	24	2.80	0.31	0.58	0.25	18.87	3.18
	35	24	1.32	0.31	0.71	0.25	27.68	3.18
PR	7	24	3.31	0.31	2.20	0.25	28.05	3.18
	21	24	3.17	0.31	0.56	0.25	20.59	3.18
	35	24	1.23	0.31	0.58	0.25	25.74	3.18
PH	7	24	3.41	0.31	2.38	0.25	29.96	3.18
	21	24	2.81	0.31	0.62	0.25	18.22	3.18
	35	24	1.18	0.31	0.80	0.25	27.41	3.18
p-value			*0.8808*		*0.6893*		*0.7924*	
GL	**Day**							
X	7	36	3.54	0.25	2.56	0.21	28.74	2.96
	21	36	2.89	0.25	0.67	0.21	20.65	2.96
	35	36	1.21	0.25	0.70	0.21	26.88	2.96
Y	7	36	2.99	0.25	1.65	0.21	29.76	2.96
	21	36	2.96	0.25	0.50	0.21	17.80	2.96
	35	36	1.27	0.25	0.69	0.21	27.01	2.96
p-value			*0.3632*		*0.0757*		*0.4658*	
Sex	**Day**							
F	7	36	3.17	0.25	2.01	0.21	28.90	2.96
	21	36	2.85	0.25	0.59	0.21	19.45	2.96
	35	36	1.31	0.25	0.68	0.21	27.21	2.96
M	7	36	3.37	0.25	2.20	0.21	29.60	2.96
	21	36	2.99	0.25	0.58	0.21	19.00	2.96
	35	36	1.18	0.25	0.71	0.21	26.68	2.96
p-value			*0.7807*		*0.8717*		*0.915*	
TRT	**EXP**							

		A	24	1.25	0.31	1.09	0.25	15.62	4.72
BD		B	24	2.02	0.31	1.23	0.25	34.87	4.72
		C	24	3.92	0.31	0.70	0.25	25.80	4.72
		A	24	1.37	0.31	1.00	0.25	15.27	4.72
PR		B	24	1.96	0.31	1.84	0.25	34.32	4.72
		C	24	4.36	0.31	0.51	0.25	24.80	4.72
		A	24	1.20	0.31	1.13	0.25	14.28	4.72
PH		B	24	2.18	0.31	2.06	0.25	36.27	4.72
		C	24	4.02	0.31	0.61	0.25	25.05	4.72
	p-value			*0.8927*		*0.3365*		*0.9475*	
GL		**EXP**							
		A	36	1.36	0.26	1.33	0.21	14.44	4.57
X		B	36	2.14	0.26	1.88	0.21	35.88	4.57
		C	36	4.15	0.26	0.72	0.21	25.95	4.57
		A	36	1.20	0.26	0.81	0.21	15.67	4.57
Y		B	36	1.98	0.26	1.54	0.21	34.42	4.57
		C	36	4.05	0.26	0.49	0.21	24.48	4.57
	p-value			*0.9926*		*0.7714*		*0.6357*	
Sex		**EXP**							
		A	36	1.32	0.26	1.14	0.21	15.45	4.57
F		B	36	1.93	0.26	1.51	0.21	35.25	4.57
		C	36	4.07	0.26	0.63	0.21	24.86	4.57
		A	36	1.23	0.26	1.00	0.21	14.66	4.57
M		B	36	2.18	0.26	1.91	0.21	35.05	4.57
		C	36	4.13	0.26	0.58	0.21	25.57	4.57
	p-value			*0.7838*		*0.3832*		*0.898*	
Day		**EXP**							
		A	24	1.47^{c}	0.31	2.09^{b}	0.25	11.72^{d}	4.72
7		B	24	3.90^{b}	0.31	3.48^{a}	0.25	46.70^{a}	4.72
		C	24	4.43^{b}	0.31	0.74^{c}	0.25	29.33^{b}	4.72
		A	24	1.11^{c}	0.31	0.54^{c}	0.25	12.83^{cd}	4.72
21		B	24	1.14^{c}	0.31	0.72^{c}	0.25	25.33^{bc}	4.72
		C	24	6.52^{a}	0.31	0.50^{c}	0.25	19.52^{cd}	4.72
		A	24	1.25	0.31	0.58^{c}	0.25	20.62^{bc}	4.72
35		B	24	1.12^{c}	0.31	0.93^{c}	0.25	33.43^{b}	4.72
		C	24	1.36^{c}	0.31	0.58^{c}	0.25	26.79^{b}	4.72
	p-value			*<0.0001*		*<0.0001*		*<0.0001*	

[1]Values are presented as least squares means (LSmeans). SE: Standard error.
BD: Basal diet; PR: Basal diet + probiotic; PH: Basal diet + phytobiotic. F: female; M: male.
X and Y correspond to two fast growing commercial poultry genetic lines.
[a–d]Within a column, values without a common superscript differ, P < 0.05.

Table S7b: Concentration of C-reactive protein (CRP), chicken haptoglobin-like protein (PIT54) and Lipopolysaccharides (LPS) in blood per day[1].

		N	CRP d7 (pg/mL)		CRP d21 (pg/mL)		CRP d35 (pg/mL)		PIT54 d7 (ng/ml)		PIT54 d21 (ng/ml)		PIT54 d35 (ng/ml)		LPS d7 (ng/L)		LPS d21 (ng/L)		LPS d 35 (ng/L)	
			LS ó	SE	LS ó	SE	LS ó	SE	LS ó	SE	LS ó	SE	LS ó	SE	LS ó	SE	LS ó	SE	LS ó	SE
Treatment (TRT)																				
BD		24	3.09	0.22	2.80	0.57	1.32	0.08	1.74	0.42	0.58	0.04	0.71	0.08	29.73	2.96	18.87	2.38	27.68	5.95
PR		24	3.31	0.22	3.17	0.57	1.23	0.08	2.20	0.42	0.56	0.04	0.58	0.08	28.05	2.96	20.59	2.38	25.74	5.95
PH		24	3.41	0.22	2.81	0.57	1.18	0.08	2.38	0.42	0.62	0.04	0.80	0.08	29.96	2.96	18.22	2.38	27.41	5.95
P-value			*0.5406*		*0.7581*		*0.2094*		*0.5406*		*0.6850*		*0.1543*		*0.1774*		*0.5088*		*0.7140*	
Genetic line (GL)																				
X		36	3.54	0.18	2.89	0.53	1.21	0.08	2.56	0.34	0.67	0.04	0.70	0.06	28.74	2.93	20.65	2.23	26.88	5.86
Y		36	2.99	0.18	2.96	0.53	1.27	0.08	1.65	0.34	0.50	0.04	0.69	0.06	29.76	2.93	17.80	2.23	27.01	5.86
P-value			*0.0687*		*0.8870*		*0.3300*		*0.0687*		*0.0010*		*0.9519*		*0.2657*		*0.1007*		*0.9477*	
Sex																				
F		36	3.17	0.18	2.85	0.53	1.31	0.08	2.01	0.34	0.59	0.04	0.68	0.06	28.9	2.93	19.45	2.23	27.21	5.86
M		36	3.37	0.18	2.99	0.53	1.18	0.08	2.20	0.34	0.58	0.04	0.71	0.06	29.6	2.93	19.00	2.23	26.68	5.86
P-value			*0.6824*		*0.7601*		*0.0454*		*0.6824*		*0.9344*		*0.7171*		*0.4384*		*0.7964*		*0.8000*	
Experiment (EXP)																				
A		24	1.47[b]	0.24	1.11[b]	0.82	1.25	0.12	2.09[b]	0.42	0.54[b]	0.04	0.58[b]	0.08	11.72[c]	5.01	12.83	3.56	20.62	9.99
B		24	3.90[a]	0.24	1.14[b]	0.82	1.12	0.12	3.48[a]	0.42	0.72[a]	0.04	0.93[a]	0.08	46.70[a]	5.01	25.33	3.56	33.43	9.99
C		24	4.43[a]	0.24	6.52[a]	0.82	1.36	0.12	0.74[c]	0.42	0.50[b]	0.04	0.58[b]	0.08	29.33[b]	5.01	19.52	3.56	26.79	9.99
P-value			*0.0001*		*<0.0001*		*0.4057*		*0.0001*		*0.0012*		*0.0032*		*<0.0001*		*0.0535*		*0.6646*	
TRT	**GL**																			
BD	X	12	3.49	0.3	3.11	0.7	1.23	0.1	1.88	0.59	0.65	0.06	0.81	0.11	29.99[a]	3.06	20.22	2.8	29.86	6.22
	Y	12	2.68	0.3	2.49	0.7	1.40	0.1	1.59	0.59	0.50	0.06	0.60	0.11	29.47[a]	3.06	17.52	2.8	25.51	6.22
PR	X	12	3.55	0.3	2.73	0.7	1.25	0.1	2.95	0.59	0.65	0.06	0.56	0.11	25.74[b]	3.06	22.11	2.8	25.88	6.22
	Y	12	3.06	0.3	3.60	0.7	1.20	0.1	1.45	0.59	0.48	0.06	0.60	0.11	30.36[a]	3.06	19.06	2.8	25.61	6.22
PH	X	12	3.59	0.3	2.84	0.7	1.14	0.1	2.84	0.59	0.71	0.06	0.72	0.11	30.49[a]	3.06	19.62	2.8	24.89	6.22
	Y	12	3.23	0.3	2.79	0.7	1.22	0.1	1.92	0.59	0.52	0.06	0.88	0.11	29.44[a]	3.06	16.82	2.8	29.92	6.22
P-value			*0.5998*		*0.4133*		*0.3487*		*0.5998*		*0.9325*		*0.2629*		*0.0235*		*0.9965*		*0.1919*	
TRT	**Sex**																			
BD	F	12	2.71	0.3	2.64	0.7	1.44	0.1	2	0.59	0.62	0.06	0.69	0.11	30.52	3.06	16.79	2.8	26.2	6.22
	M	12	3.46	0.3	2.95	0.7	1.20	0.1	1.48	0.59	0.54	0.06	0.73	0.11	28.95	3.06	20.95	2.8	29.17	6.22
PR	F	12	3.36	0.3	2.89	0.7	1.27	0.1	1.80	0.59	0.54	0.06	0.52	0.11	27.72	3.06	21.53	2.8	29.58	6.22
	M	12	3.25	0.3	3.44	0.7	1.18	0.1	2.60	0.59	0.59	0.06	0.64	0.11	28.39	3.06	19.64	2.8	21.91	6.22
PH	F	12	3.44	0.3	3.03	0.7	1.22	0.1	2.21	0.59	0.61	0.06	0.83	0.11	28.46	3.06	20.01	2.8	25.85	6.22
	M	12	3.38	0.3	2.59	0.7	1.15	0.1	2.54	0.59	0.62	0.06	0.77	0.11	31.47	3.06	16.43	2.8	28.96	6.22
P-value			*0.5366*		*0.6567*		*0.4872*		*0.5366*		*0.5615*		*0.7481*		*0.1260*		*0.1606*		*0.0613*	
GL	**Sex**																			
X	F	18	3.49	0.25	3.10	0.62	1.25	0.09	2.32	0.49	0.68	0.05	0.70	0.09	28.11	3	20.89	2.53	27.23	6.04

		M	18	3.60	0.25	2.68	0.62	1.17	0.09	2.79	0.49	0.66	0.05	0.70	0.09	29.37	3	20.42	2.53	26.52	6.04
	Y	F	18	2.84	0.25	2.60	0.62	1.36	0.09	1.69	0.49	0.49	0.05	0.66	0.09	29.68	3	18.01	2.53	27.19	6.04
		M	18	3.13	0.25	3.31	0.62	1.18	0.09	1.62	0.49	0.50	0.05	0.73	0.09	29.83	3	17.59	2.53	26.84	6.04
	P-value			*0.5776*		*0.2219*		*0.4209*		*0.5776*		*0.8065*		*0.6819*		*0.5440*		*0.9883*		*0.9311*	
GL	**Sex**	**TRT**																			
		BD	6	3.14	0.42	3.44	0.89	1.29	0.13	2.30	0.84	0.72	0.09	0.78	0.16	31.13	3.26	16.20	3.5	29.58	6.72
	F	PR	6	3.77	0.42	2.91	0.89	1.33	0.13	2.59	0.84	0.61	0.09	0.53	0.16	25.38	3.26	22.34	3.5	29.48	6.72
		PH	6	3.56	0.42	2.97	0.89	1.13	0.13	2.07	0.84	0.71	0.09	0.78	0.16	27.83	3.26	24.12	3.5	22.63	6.72
X		BD	6	3.84	0.42	2.78	0.89	1.18	0.13	1.46	0.84	0.58	0.09	0.84	0.16	28.85	3.26	24.25	3.5	30.13	6.72
	M	PR	6	3.33	0.42	2.56	0.89	1.18	0.13	3.30	0.84	0.69	0.09	0.60	0.16	26.11	3.26	21.88	3.5	22.28	6.72
		PH	6	3.62	0.42	2.71	0.89	1.15	0.13	3.61	0.84	0.72	0.09	0.66	0.16	33.15	3.26	15.12	3.5	27.15	6.72
		BD	6	2.28	0.42	1.85	0.89	1.59	0.13	1.69	0.84	0.51	0.09	0.59	0.16	29.90	3.26	17.39	3.5	22.81	6.72
	F	PR	6	2.94	0.42	2.87	0.89	1.20	0.13	1.02	0.84	0.46	0.09	0.51	0.16	30.05	3.26	20.73	3.5	29.67	6.72
		PH	6	3.31	0.42	3.09	0.89	1.30	0.13	2.36	0.84	0.51	0.09	0.87	0.16	29.08	3.26	15.91	3.5	29.08	6.72
Y		BD	6	3.08	0.42	3.12	0.89	1.22	0.13	1.50	0.84	0.50	0.09	0.62	0.16	29.04	3.26	17.64	3.5	28.21	6.72
	M	PR	6	3.18	0.42	4.33	0.89	1.19	0.13	1.89	0.84	0.49	0.09	0.68	0.16	30.67	3.26	17.40	3.5	21.54	6.72
		PH	6	3.15	0.42	2.48	0.89	1.14	0.13	1.47	0.84	0.53	0.09	0.89	0.16	29.79	3.26	17.74	3.5	30.76	6.72
	P-value			*0.3871*		*0.5218*		*0.4079*		*0.3871*		*0.7641*		*0.9213*		*0.3720*		*0.0785*		*0.7356*	

[1]Values are presented as least squares means (LS ó). SE: Standard error. BD: Basal diet; PR: Basal diet + probiotic; PH: Basal diet + phytobiotic. F: female; M: male. X and Y correspond to two fast growing commercial poultry genetic lines.

[a–d]Within a column, values without a common superscript differ, P < 0.05.

Table S8a: Punctuation of footpad dermatitis lesions (%) 0-35 days[1].

		N	Punctuation 0		Punctuation 1		Punctuation 2		Punctuation 3		Punctuation 4	
			LSmeans	SE	LSmeans	SE	LSmeans	SE	LSmeans	SE	LSmeans	SE
Treatment (TRT)												
BD		72	83.29	2.217	9.21	1.046	6.11	1.086	1.31	0.580	0.08	0.088
PR		72	86.34	2.217	9.03	1.046	4.19	1.086	0.40	0.580	0.04	0.088
PH		72	87.70	2.217	7.48	1.046	3.50	1.086	1.13	0.580	0.19	0.088
p-value			*0.0959*		*0.2274*		*0.0508*		*0.3575*		*0.3649*	
Genetic line (GL)												
X		108	81.70	2.051	10.19	0.946	6.25	0.989	1.65	0.512	0.21	0.076
Y		108	89.86	2.051	6.95	0.946	2.94	0.989	0.24	0.512	0.00	0.076
p-value			*<0.0001*		*0.0005*		*0.0003*		*0.0109*		*0.0212*	
Sex												
F		108	85.56	2.051	8.56	0.946	4.91	0.989	0.88	0.512	0.10	0.076
M		108	86.00	2.051	8.58	0.946	4.29	0.989	1.01	0.512	0.12	0.076
p-value			*0.7928*		*0.9769*		*0.4902*		*0.8112*		*0.8320*	
Day												
7		72	99.90^{a}	2.217	0.11^{c}	1.046	0.00^{c}	1.086	0.00^{b}	0.580	0.00	0.088
21		72	85.08^{b}	2.217	9.29^{b}	1.046	4.38^{b}	1.086	1.01^{ab}	0.580	0.25	0.088
35		72	72.36^{c}	2.217	16.32^{a}	1.046	9.42^{a}	1.086	1.84^{a}	0.580	0.06	0.088
p-value			*<0.0001*		*<0.0001*		*<0.0001*		*0.0254*		*0.0626*	
Experiment (EXP)												
A		72	73.82^{b}	3.238	16.08^{a}	1.442	8.84^{a}	1.528	1.01	0.752	0.25	0.106
B		72	94.08^{a}	3.238	4.74^{b}	1.442	1.13^{b}	1.528	0.05	0.752	0.00	0.106
C		72	89.44^{a}	3.238	4.88^{b}	1.442	3.83^{b}	1.528	1.79	0.752	0.06	0.106
p-value			*<0.0001*		*<0.0001*		*0.0021*		*0.2673*		*0.2151*	
TRT	**GL**											
BD	X	36	79.13	2.655	10.94	1.303	7.65	1.334	2.12	0.748	0.16	0.118
	Y	36	87.44	2.655	7.48	1.303	4.58	1.334	0.50	0.748	0.00	0.118
PR	X	36	82.61	2.655	10.61	1.303	6.02	1.334	0.67	0.748	0.08	0.118
	Y	36	90.07	2.655	7.45	1.303	2.35	1.334	0.13	0.748	0.00	0.118
PH	X	36	83.34	2.655	9.02	1.303	5.09	1.334	2.17	0.748	0.39	0.118
	Y	36	92.07	2.655	5.94	1.303	1.90	1.334	0.09	0.748	0.00	0.118
p-value			*0.9517*		*0.9831*		*0.9572*		*0.4995*		*0.3649*	
TRT	**Sex**											
BD	F	36	82.02	2.655	9.85	1.303	7.19	1.334	0.86	0.748	0.16	0.118
	M	36	84.56	2.655	8.56	1.303	5.04	1.334	1.76	0.748	0.00	0.118
PR	F	36	87.90	2.655	8.63	1.303	3.48	1.334	0.00	0.748	0.08	0.118
	M	36	84.78	2.655	9.43	1.303	4.89	1.334	0.80	0.748	0.00	0.118
PH	F	36	86.75	2.655	7.19	1.303	4.06	1.334	1.79	0.748	0.39	0.118
	M	36	88.66	2.655	7.76	1.303	2.93	1.334	0.47	0.748	0.00	0.118
p-value			*0.3279*		*0.5796*		*0.2495*		*0.1757*		*0.8696*	
GL	**Sex**											
X	F	54	81.41	2.372	10.13	1.138	6.79	1.174	1.49	0.641	0.19	0.099
	M	54	81.98	2.372	10.24	1.138	5.72	1.174	1.82	0.641	0.23	0.099
Y	F	54	89.70	2.372	6.99	1.138	3.03	1.174	0.28	0.641	0.00	0.099
	M	54	90.02	2.372	6.92	1.138	2.85	1.174	0.20	0.641	0.00	0.099
p-value			*0.9398*		*0.9208*		*0.6227*		*0.1999*		*0.8320*	
TRT	**Day**											
BD	7	24	99.79	3.030	0.21	1.517	0.00	1.543	0.00	0.884	0.00	0.141
	21	24	81.61	3.030	10.12	1.517	6.65	1.543	1.38	0.884	0.24	0.141
	35	24	68.46	3.030	17.28	1.517	11.70	1.543	2.56	0.884	0.00	0.141
PR	7	24	99.89	3.030	0.11	1.517	0.00	1.543	0.00	0.884	0.00	0.141
	21	24	85.10	3.030	10.49	1.517	3.67	1.543	0.62	0.884	0.13	0.141
	35	24	74.03	3.030	16.50	1.517	8.89	1.543	0.59	0.884	0.00	0.141
PH	7	24	100.00	3.030	0.00	1.517	0.00	1.543	0.00	0.884	0.00	0.141
	21	24	88.52	3.030	7.25	1.517	2.81	1.543	1.02	0.884	0.39	0.141
	35	24	74.59	3.030	15.18	1.517	7.68	1.543	2.37	0.884	0.19	0.141
p-value			*0.6399*		*0.7844*		*0.5467*		*0.7555*		*0.8764*	
GL	**Day**											
X	7	36	99.86^{a}	2.655	0.14^{d}	1.303	0.00^{c}	1.334	0.00^{b}	0.748	0.00	0.118
	21	36	81.75^{c}	2.655	11.13^{b}	1.303	5.19^{b}	1.334	1.42^{b}	0.748	0.51	0.118
	35	36	63.48^{d}	2.655	19.29^{a}	1.303	13.57^{a}	1.334	3.54^{a}	0.748	0.13	0.118
Y	7	36	99.93^{a}	2.655	0.07^{d}	1.303	0.00^{c}	1.334	0.00^{b}	0.748	0.00	0.118
	21	36	88.41^{b}	2.655	7.44^{c}	1.303	3.56^{b}	1.334	0.59^{b}	0.748	0.00	0.118
	35	36	81.25^{c}	2.655	13.35^{b}	1.303	5.27^{b}	1.334	0.13^{b}	0.748	0.00	0.118
p-value			*0.0002*		*0.0297*		*0.0006*		*0.0320*		*0.0626*	
Sex	**Day**											
F	7	36	99.79	2.655	0.21	1.303	0.00	1.334	0.00	0.748	0.00	0.118
	21	36	84.27	2.655	9.97	1.303	4.52	1.334	1.08	0.748	0.16	0.118
	35	36	72.61	2.655	15.49	1.303	10.21	1.334	1.57	0.748	0.13	0.118
M	7	36	100.00	2.655	0.00	1.303	0.00	1.334	0.00	0.748	0.00	0.118
	21	36	85.89	2.655	8.60	1.303	4.23	1.334	0.93	0.748	0.35	0.118
	35	36	72.11	2.655	17.15	1.303	8.64	1.334	2.11	0.748	0.00	0.118
p-value			*0.8722*		*0.3848*		*0.7483*		*0.8612*		*0.3719*	
TRT	**EXP**											

	A	24	68.39	3.840	17.19	1.813	12.80^{a}	1.881	1.38	1.005	0.24	0.153
BD	B	24	93.68	3.840	4.91	1.813	1.25^{c}	1.881	0.15	1.005	0.00	0.153
	C	24	87.80	3.840	5.51	1.813	4.29^{bc}	1.881	2.40	1.005	0.00	0.153
	A	24	72.83	3.840	17.57	1.813	8.86^{b}	1.881	0.62	1.005	0.13	0.153
PR	B	24	93.60	3.840	5.31	1.813	1.10^{c}	1.881	0.00	1.005	0.00	0.153
	C	24	92.60	3.840	4.21	1.813	2.60^{c}	1.881	0.59	1.005	0.00	0.153
	A	24	80.25	3.840	13.49	1.813	4.86^{bc}	1.881	1.02	1.005	0.39	0.153
PH	B	24	94.95	3.840	4.01	1.813	1.04^{c}	1.881	0.00	1.005	0.00	0.153
	C	24	87.92	3.840	4.93	1.813	4.59^{bc}	1.881	2.37	1.005	0.19	0.153
p-value			*0.0696*		*0.4692*		*0.0167*		*0.8069*		*0.8764*	
GL	**EXP**											
	A	36	68.45^{c}	3.552	18.46^{a}	1.638	11.16^{a}	1.713	1.42^{ab}	0.887	0.51	0.131
X	B	36	94.33^{a}	3.552	4.45^{cd}	1.638	1.12^{b}	1.713	0.10^{b}	0.887	0.00	0.131
	C	36	82.31^{b}	3.552	7.65^{c}	1.638	6.48^{a}	1.713	3.44^{a}	0.887	0.13	0.131
	A	36	79.19^{b}	3.552	13.70^{b}	1.638	6.52^{a}	1.713	0.59^{b}	0.887	0.00	0.131
Y	B	36	93.82^{a}	3.552	5.04^{cd}	1.638	1.14^{b}	1.713	0.00^{b}	0.887	0.00	0.131
	C	36	96.57^{a}	3.552	2.12^{d}	1.638	1.17^{b}	1.713	0.13^{b}	0.887	0.00	0.131
p-value			*0.0014*		*0.0121*		*0.0333*		*0.0459*		*0.0626*	
Sex	**EXP**											
	A	36	72.43	3.552	16.74	1.638	9.59	1.713	1.08	0.887	0.16	0.131
F	B	36	94.57	3.552	4.41	1.638	1.01	1.713	0.00	0.887	0.00	0.131
	C	36	89.66	3.552	4.52	1.638	4.12	1.713	1.57	0.887	0.13	0.131
	A	36	75.21	3.552	15.43	1.638	8.09	1.713	0.93	0.887	0.35	0.131
M	B	36	93.58	3.552	5.08	1.638	1.25	1.713	0.10	0.887	0.00	0.131
	C	36	89.22	3.552	5.25	1.638	3.53	1.713	2.01	0.887	0.00	0.131
p-value			*0.6161*		*0.5756*		*0.7336*		*0.9060*		*0.3719*	
Day	**EXP**											
	A	24	99.69^{a}	3.840	0.32^{d}	1.813	0.00^{b}	1.881	0.00^{b}	1.005	0.00^{b}	0.153
7	B	24	100.00^{a}	3.840	0.00^{d}	1.813	0.00^{b}	1.881	0.00^{b}	1.005	0.00^{b}	0.153
	C	24	100.00^{a}	3.840	0.00^{d}	1.813	0.00^{b}	1.881	0.00^{b}	1.005	0.00^{b}	0.153
	A	24	62.93^{cd}	3.840	20.16^{b}	1.813	13.13^{a}	1.881	3.02^{a}	1.005	0.76^{a}	0.153
21	B	24	95.34^{ab}	3.840	4.66^{d}	1.813	0.00^{b}	1.881	0.00^{b}	1.005	0.00^{b}	0.153
	C	24	96.96^{ab}	3.840	3.04^{d}	1.813	0.00^{b}	1.881	0.00^{b}	1.005	0.00^{b}	0.153
	A	24	58.84^{d}	3.840	27.77^{a}	1.813	13.39^{a}	1.881	0.00^{b}	1.005	0.00^{b}	0.153
35	B	24	86.89^{b}	3.840	9.57^{c}	1.813	3.39^{b}	1.881	0.15^{b}	1.005	0.00^{b}	0.153
	C	24	71.35^{c}	3.840	11.62^{c}	1.813	11.48^{a}	1.881	5.36^{a}	1.005	0.19^{b}	0.153
p-value			*<0.0001*		*<0.0001*		*<0.0001*		*<0.0001*		*0.0036*	

[1]Values are presented as least squares means (LSmeans). SE: Standard error. BD: Basal diet; PR: Basal diet + probiotic; PH: Basal diet + phytobiotic. F: female; M: male.
X and Y correspond to two fast growing commercial poultry genetic lines.
[a–d]Within a column, values without a common superscript differ, P < 0.05.

Table S8b: Punctuation of footpad dermatitis (FPD) lesions (%) per day[1].

Punctuation of FPD (%)			Day 7				Day 21										Day 35									
		N	0		1		0		1		2		3		4		0		1		2		3		4	
			LS ó	SE	LS ó	SE	LS ó	SE	LS ó	SE	LS ó	SE	LS ó	SE	LS ó	SE	LS ó	SE	LS ó	SE	LS ó	SE	LS ó	SE	LS ó	SE
Treatment (TRT)																										
BD		24	99.79	0.094	0.21	0.094	81.61	3.824	10.12	1.827	6.65	2.249	1.38	0.587	0.24	0.232	68.46	4.612	17.28	1.856	11.70	2.040	2.56	1.637	0.00	0.110
PR		24	99.89	0.094	0.11	0.094	85.10	3.824	10.49	1.827	3.67	2.249	0.62	0.587	0.13	0.232	74.03	4.612	16.50	1.856	8.89	2.040	0.59	1.637	0.00	0.110
PH		24	100.00	0.094	0.00	0.094	88.52	3.824	7.25	1.827	2.81	2.249	1.02	0.587	0.39	0.232	74.59	4.612	15.18	1.856	7.68	2.040	2.37	1.637	0.19	0.110
P-value			*0.2980*		*0.2980*		*0.2519*		*0.2132*		*0.2040*		*0.5662*		*0.6179*		*0.3837*		*0.7121*		*0.2383*		*0.5123*		*0.3745*	
Genetic line (GL)																										
X		36	99.86	0.076	0.14	0.076	81.75	3.437	11.13	1.637	5.19	2.058	1.42	0.512	0.51	0.204	63.48	4.164	19.29	1.529	13.57	1.789	3.54	1.448	0.13	0.089
Y		36	99.93	0.076	0.07	0.076	88.41	3.437	7.44	1.637	3.56	2.058	0.59	0.512	0.00	0.204	81.25	4.164	13.35	1.529	5.27	1.789	0.13	1.448	0.00	0.089
P-value			*0.5334*		*0.5334*		*0.0523*		*0.0270*		*0.3714*		*0.1574*		*0.0253*		*<0.0001*		*0.0067*		*<0.0001*		*0.0295*		*0.3217*	
Sex																										
F		36	99.79	0.076	0.21	0.076	84.27	3.437	9.97	1.637	4.52	2.058	1.08	0.512	0.16	0.204	72.61	4.164	15.49	1.529	10.21	1.789	1.57	1.448	0.13	0.089
M		36	100.00	0.076	0.00	0.076	85.89	3.437	8.60	1.637	4.23	2.058	0.93	0.512	0.35	0.204	72.11	4.164	17.15	1.529	8.64	1.789	2.11	1.448	0.00	0.089
P-value			*0.0573*		*0.0573*		*0.6307*		*0.4032*		*0.8742*		*0.7934*		*0.4071*		*0.9000*		*0.4351*		*0.4269*		*0.7235*		*0.3217*	
Experiment (EXP)																										
A		24	99.69[b]	0.094	0.32[a]	0.094	62.93[b]	5.195	20.16[a]	2.463	13.13[a]	3.198	3.02[a]	0.733	0.76	0.297	58.84[b]	6.343	27.77[a]	1.920	13.39[a]	2.590	0.00	2.132	0.00	0.110
B		24	100.00[a]	0.094	0.00[b]	0.094	95.34[a]	5.195	4.66[b]	2.463	0.00[b]	3.198	0.00[b]	0.733	0.00	0.297	86.89[a]	6.343	9.57[b]	1.920	3.39[b]	2.590	0.15	2.132	0.00	0.110
C		24	100.00[a]	0.094	0.00[b]	0.094	96.96[a]	5.195	3.04[b]	2.463	0.00[b]	3.198	0.00[b]	0.733	0.00	0.297	71.3[ab]	6.343	11.62[b]	1.920	11.48[a]	2.590	5.36	2.132	0.19	0.110
P-value			*0.0292*		*0.0292*		*<0.0001*		*<0.0001*		*0.0060*		*0.0059*		*0.1227*		*0.0110*		*<0.0001*		*0.0200*		*0.1390*		*0.3745*	
TRT	**GL**																									
1	X	12	99.58	0.132	0.42	0.132	81.77	4.803	10.51	2.305	5.99	2.745	1.25	0.770	0.48	0.300	56.04	5.749	21.89	2.602	16.96	2.655	5.11	2.104	0.00	0.155
	Y	12	100.00	0.132	0.00	0.132	81.45	4.803	9.74	2.305	7.31	2.745	1.50	0.770	0.00	0.300	80.88	5.749	12.68	2.602	6.44	2.655	0.00	2.104	0.00	0.155
2	X	12	100.00	0.132	0.00	0.132	81.70	4.803	12.27	2.305	4.53	2.745	1.24	0.770	0.25	0.300	66.14	5.749	19.54	2.602	13.54	2.655	0.78	2.104	0.00	0.155
	Y	12	99.79	0.132	0.21	0.132	88.49	4.803	8.70	2.305	2.81	2.745	0.00	0.770	0.00	0.300	81.93	5.749	13.45	2.602	4.23	2.655	0.40	2.104	0.00	0.155
3	X	12	100.00	0.132	0.00	0.132	81.78	4.803	10.61	2.305	5.07	2.745	1.76	0.770	0.78	0.300	68.24	5.749	16.44	2.602	10.20	2.655	4.74	2.104	0.38	0.155
	Y	12	100.00	0.132	0.00	0.132	95.27	4.803	3.89	2.305	0.56	2.745	0.28	0.770	0.00	0.300	80.93	5.749	13.92	2.602	5.15	2.655	0.00	2.104	0.00	0.155
P-value			*0.0620*		*0.0620*		*0.2520*		*0.3329*		*0.4291*		*0.4164*		*0.6179*		*0.4346*		*0.4367*		*0.4936*		*0.3777*		*0.3745*	
TRT	**Sex**																									
1	F	12	99.58	0.132	0.42	0.132	78.75	4.803	11.33	2.305	7.95	2.745	1.74	0.770	0.24	0.300	67.73	5.749	17.82	2.602	13.62	2.655	0.84	2.104	0.00	0.155
	M	12	100.00	0.132	0.00	0.132	84.48	4.803	8.92	2.305	5.34	2.745	1.01	0.770	0.25	0.300	69.20	5.749	16.75	2.602	9.78	2.655	4.28	2.104	0.00	0.155
2	F	12	99.79	0.132	0.21	0.132	88.41	4.803	9.24	2.305	2.36	2.745	0.00	0.770	0.00	0.300	75.51	5.749	16.42	2.602	8.07	2.655	0.00	2.104	0.00	0.155
	M	12	100.00	0.132	0.00	0.132	81.79	4.803	11.73	2.305	4.98	2.745	1.24	0.770	0.25	0.300	72.56	5.749	16.57	2.602	9.70	2.655	1.17	2.104	0.00	0.155
3	F	12	100.00	0.132	0.00	0.132	85.65	4.803	9.35	2.305	3.25	2.745	1.50	0.770	0.25	0.300	74.60	5.749	12.23	2.602	8.93	2.655	3.86	2.104	0.38	0.155
	M	12	100.00	0.132	0.00	0.132	91.40	4.803	5.15	2.305	2.37	2.745	0.54	0.770	0.54	0.300	74.57	5.749	18.12	2.602	6.43	2.655	0.88	2.104	0.00	0.155
P-value			*0.2980*		*0.2980*		*0.2305*		*0.2281*		*0.4912*		*0.2374*		*0.8485*		*0.8987*		*0.3608*		*0.4990*		*0.2276*		*0.3745*	
GL	**Sex**																									
X	F	18	99.72	0.108	0.28	0.108	81.19	4.176	12.19	1.999	4.98	2.426	1.32	0.654	0.32	0.256	63.32	5.019	17.92	2.134	15.38	2.264	3.13	1.806	0.25	0.126
	M	18	100.00	0.108	0.00	0.108	82.31	4.176	10.07	1.999	5.41	2.426	1.51	0.654	0.69	0.256	63.63	5.019	20.66	2.134	11.76	2.264	3.95	1.806	0.00	0.126
Y	F	18	99.86	0.108	0.14	0.108	87.35	4.176	7.75	1.999	4.06	2.426	0.84	0.654	0.00	0.256	81.90	5.019	13.06	2.134	5.04	2.264	0.00	1.806	0.00	0.126
	M	18	100.00	0.108	0.00	0.108	89.47	4.176	7.13	1.999	3.05	2.426	0.35	0.654	0.00	0.256	80.59	5.019	13.64	2.134	5.51	2.264	0.26	1.806	0.00	0.126
P-value			*0.5334*		*0.5334*		*0.8836*		*0.6464*		*0.6941*		*0.5573*		*0.4071*		*0.8390*		*0.6085*		*0.3023*		*0.8560*		*0.3217*	
GL	**Sex**	**TRT**																								
X	F	1	6	99.17	0.187	0.83	0.187	77.82	6.322	13.16	3.044	7.02	3.534	1.52	1.044	0.48	0.403	59.44	7.524	21.05	3.663	17.84	3.582	1.67	2.815	0.00	0.219
		2	6	100.00	0.187	0.00	0.187	85.72	6.322	7.86	3.044	4.95	3.534	0.98	1.044	0.49	0.403	52.65	7.524	22.72	3.663	16.08	3.582	8.55	2.815	0.00	0.219
		3	6	100.00	0.187	0.00	0.187	79.67	6.322	9.49	3.044	8.89	3.534	1.96	1.044	0.00	0.403	76.02	7.524	14.58	3.663	9.40	3.582	0.00	2.815	0.00	0.219
	M	1	6	100.00	0.187	0.00	0.187	83.24	6.322	9.99	3.044	5.73	3.534	1.04	1.044	0.00	0.403	85.74	7.524	10.79	3.663	3.47	3.582	0.00	2.815	0.00	0.219
		2	6	100.00	0.187	0.00	0.187	87.58	6.322	9.90	3.044	2.53	3.534	0.00	1.044	0.00	0.403	65.31	7.524	20.64	3.663	14.05	3.582	0.00	2.815	0.00	0.219
		3	6	100.00	0.187	0.00	0.187	75.83	6.322	14.65	3.044	6.54	3.534	2.48	1.044	0.51	0.403	66.97	7.524	18.45	3.663	13.04	3.582	1.55	2.815	0.00	0.219
Y	F	1	6	99.57	0.187	0.43	0.187	89.24	6.322	8.58	3.044	2.19	3.534	0.00	1.044	0.00	0.403	85.70	7.524	12.21	3.663	2.09	3.582	0.00	2.815	0.00	0.219
		2	6	100.00	0.187	0.00	0.187	87.75	6.322	8.82	3.044	3.43	3.534	0.00	1.044	0.00	0.403	78.15	7.524	14.69	3.663	6.37	3.582	0.79	2.815	0.00	0.219
		3	6	100.00	0.187	0.00	0.187	78.16	6.322	13.51	3.044	5.39	3.534	2.45	1.044	0.49	0.403	65.22	7.524	12.06	3.663	14.24	3.582	7.73	2.815	0.76	0.219
	M	1	6	100.00	0.187	0.00	0.187	85.39	6.322	7.72	3.044	4.74	3.534	1.08	1.044	1.08	0.403	71.27	7.524	20.81	3.663	6.17	3.582	1.76	2.815	0.00	0.219
		2	6	100.00	0.187	0.00	0.187	93.14	6.322	5.19	3.044	1.11	3.534	0.56	1.044	0.00	0.403	83.98	7.524	12.40	3.663	3.62	3.582	0.00	2.815	0.00	0.219
		3	6	100.00	0.187	0.00	0.187	97.41	6.322	2.59	3.044	0.00	3.534	0.00	1.044	0.00	0.403	77.88	7.524	15.43	3.663	6.69	3.582	0.00	2.815	0.00	0.219
P-value			*0.0620*		*0.0620*		*0.6213*		*0.4094*		*0.9648*		*0.4995*		*0.8485*		*0.2754*		*0.5204*		*0.2823*		*0.2368*		*0.3745*	

[1]Values presented as least squares means (LS ó). SE: Standard error. BD: Basal diet; PR: BD + probiotic; PH: BD + phytobiotic. F: female; M: male. [a–d]Within a column, values without a common superscript differ, P<0.05. X and Y: two fast growing commercial poultry genetic lines.

Table S9: Litter quality (pH)[1].

			N	Litter pH 0-35 days LSmeans	SE	N	Litter pH day 7 LSmeans	SE	Litter pH day 21 LSmeans	SE	Litter pH day 35 LSmeans	SE
Treatment (TRT)												
BD			72	7.27	0.086	24	6.35	0.188	7.53	0.128	7.93	0.130
PR			72	7.15	0.086	24	6.17	0.188	7.55	0.128	7.72	0.130
PH			72	7.22	0.086	24	6.43	0.188	7.45	0.128	7.79	0.130
p-value				*0.6039*			*0.5461*		*0.8440*		*0.4271*	
Genetic line (GL)												
X			108	7.13	0.070	36	6.11	0.160	7.44	0.104	7.83	0.112
Y			108	7.30	0.070	36	6.52	0.160	7.57	0.104	7.79	0.112
p-value				*0.0896*			*0.0427*		*0.3813*		*0.8098*	
Sex												
F			108	7.18	0.070	36	6.32	0.160	7.44	0.104	7.78	0.112
M			108	7.24	0.070	36	6.31	0.160	7.58	0.104	7.84	0.112
p-value				*0.5088*			*0.9823*		*0.3479*		*0.6394*	
Experiment (EXP)												
A			72	6.98[b]	0.086	24	6.27	0.216	7.38	0.128	7.29[c]	0.156
B			72	7.34[a]	0.086	24	6.69	0.216	7.55	0.128	7.79[b]	0.156
C			72	7.31[a]	0.086	24	5.98	0.216	7.59	0.128	8.35[a]	0.156
p-value				*0.0059*			*0.0736*		*0.4607*		*<0.0001*	
TRT		**GL**										
BD		X	36	7.21	0.122	12	6.29	0.255	7.37	0.180	7.98	0.174
		Y	36	7.32	0.122	12	6.40	0.255	7.68	0.180	7.88	0.174
PR		X	36	7.00	0.122	12	5.97	0.255	7.47	0.180	7.58	0.174
		Y	36	7.29	0.122	12	6.37	0.255	7.63	0.180	7.86	0.174
PH		X	36	7.17	0.122	12	6.08	0.255	7.50	0.180	7.93	0.174
		Y	36	7.28	0.122	12	6.79	0.255	7.40	0.180	7.65	0.174
P-value				*0.7200*			*0.4685*		*0.5246*		*0.2298*	
TRT		**Sex**										
BD		F	36	7.33	0.122	12	6.34	0.255	7.72	0.180	7.95	0.174
		M	36	7.20	0.122	12	6.36	0.255	7.33	0.180	7.91	0.174
PR		F	36	7.10	0.122	12	6.28	0.255	7.38	0.180	7.63	0.174
		M	36	7.20	0.122	12	6.06	0.255	7.73	0.180	7.80	0.174
PH		F	36	7.11	0.122	12	6.34	0.255	7.22	0.180	7.76	0.174
		M	36	7.34	0.122	12	6.53	0.255	7.68	0.180	7.82	0.174
P-value				*0.3093*			*0.6976*		*0.0468**		*0.8095*	
GL		**Sex**										
X		F	54	7.12	0.099	18	6.11	0.213	7.43	0.147	7.83	0.146
		M	54	7.13	0.099	18	6.11	0.213	7.46	0.147	7.83	0.146
Y		F	54	7.24	0.099	18	6.53	0.213	7.45	0.147	7.73	0.146
		M	54	7.36	0.099	18	6.52	0.213	7.70	0.147	7.86	0.146
P-value				*0.5725*			*0.9624*		*0.4383*		*0.6364*	
GL	**Sex**	**TRT**										
X	F	BD	18	7.40	0.172	6	6.26	0.353	7.91[a]	0.255	8.03	0.239
		PR	18	6.98	0.172	6	6.32	0.353	6.82[c]	0.255	7.93	0.239
		PH	18	6.99	0.172	6	6.41	0.353	7.53[abc]	0.255	7.87	0.239
	M	BD	18	7.02	0.172	6	6.39	0.353	7.84[a]	0.255	7.89	0.239
		PR	18	7.03	0.172	6	6.16	0.353	7.30[abc]	0.255	7.49	0.239
		PH	18	7.35	0.172	6	5.77	0.353	7.64[ab]	0.255	7.67	0.239
Y	F	BD	18	7.27	0.172	6	6.40	0.353	7.45[abc]	0.255	7.78	0.239
		PR	18	7.21	0.172	6	6.35	0.353	7.82[a]	0.255	7.94	0.239
		PH	18	7.23	0.172	6	5.90	0.353	7.09[bc]	0.255	7.97	0.239
	M	BD	18	7.37	0.172	6	6.25	0.353	7.91[a]	0.255	7.89	0.239
		PR	18	7.37	0.172	6	6.78	0.353	7.36[abc]	0.255	7.55	0.239
		PH	18	7.34	0.172	6	6.81	0.353	7.45[abc]	0.255	7.75	0.239
P-value				*0.3237*			*0.7948*		*0.0149*		*0.8954*	
Day												
7			72	6.32[c]	0.09							
21			72	7.51[b]	0.09							
35			72	7.81[a]	0.09							
P-value				*<0.0001*								
TRT		**Day**										
BD		7	24	6.35	0.149							
		21	24	7.53	0.149							
		35	24	7.93	0.149							
PR		7	24	6.17	0.149							
		21	24	7.55	0.149							

		35	24	7.72	0.149
		7	24	6.43	0.149
PH		21	24	7.45	0.149
		35	24	7.79	0.149
	P-value			*0.7528*	
GL		**Day**			
		7	36	6.11	0.122
X		21	36	7.44	0.122
		35	36	7.83	0.122
		7	36	6.52	0.122
Y		21	36	7.57	0.122
		35	36	7.79	0.122
	P-value			*0.1851*	
Sex		**Day**			
		7	36	6.32	0.122
F		21	36	7.44	0.122
		35	36	7.78	0.122
		7	36	6.31	0.122
M		21	36	7.58	0.122
		35	36	7.84	0.122
	P-value			*0.8399*	

[1]Values are presented as least squares means (LSmeans). SE: Standard error. BD: Basal diet; PR: Basal diet + probiotic; PH: Basal diet + phytobiotic. F: female; M: male.
X and Y correspond to two fast growing commercial poultry genetic lines.
[a–d]Within a column, values without a common superscript differ, $P < 0.05$.
*Analysis of variance reached significance, but the mean comparisons did not reach statistical significance between groups.

Table S10: Meat oxidative stability.

			N	TBARS LSmeans	SE
Treatment (TRT)					
BD			24	0.06	0.004
PR			24	0.06	0.004
PH			24	0.06	0.004
P-value				*0.9670*	
Genetic line (GL)					
X			36	0.06	0.004
Y			36	0.07	0.004
P-value				*0.0538*	
Sex					
F			36	0.07	0.004
M			36	0.06	0.004
P-value				*0.3732*	
Experiment (EXP)					
A			24	0.05[b]	0.006
B			24	0.10[a]	0.006
C			24	0.05[b]	0.006
P-value				*<0.0001*	
TRT		**GL**			
BD		X	12	0.06	0.006
		Y	12	0.07	0.006
PR		X	12	0.06	0.006
		Y	12	0.07	0.006
PH		X	12	0.06	0.006
		Y	12	0.07	0.006
P-value				*0.3879*	
TRT		**Sex**			
BD		F	12	0.06	0.006
		M	12	0.06	0.006
PR		F	12	0.07	0.006
		M	12	0.06	0.006
PH		F	12	0.07	0.006
		M	12	0.06	0.006
P-value				*0.7154*	
GL		**Sex**			
X		F	18	0.06	0.005
		M	18	0.06	0.005
Y		F	18	0.07	0.005
		M	18	0.07	0.005
P-value				*0.2349*	
GL	**Sex**	**TRT**			
X	F	BD	6	0.05	0.007
		PR	6	0.06	0.007
		PH	6	0.07	0.007
	M	BD	6	0.07	0.007
		PR	6	0.07	0.007
		PH	6	0.06	0.007
Y	F	BD	6	0.07	0.007
		PR	6	0.07	0.007
		PH	6	0.07	0.007
	M	BD	6	0.05	0.007
		PR	6	0.06	0.007
		PH	6	0.07	0.007
P-value				*0.3027*	

[1]Values are presented as least squares means (LSmeans). SE: Standard error.
TBARS: Thiobarbituric acid reactive substances (TBARS units). BD: Basal diet; PR: Basal diet + probiotic; PH: Basal diet + phytobiotic. F: female; M: male.
X and Y correspond to two fast growing commercial poultry genetic lines.
[a–d]Within a column, values without a common superscript differ, P < 0.05.

Table S11: Carcass cuts (%)[1].

			N	Carcass yield %		Breast yield %		Thigh yield %		Abdominal Fat yield %	
				LSmeans	SE	LSmeans	SE	LSmeans	SE	LSmeans	SE
Treatment (TRT)											
BD			24	76.6	0.24	19.0	0.12	20.4	0.09	1.5	0.03
PR			24	76.4	0.24	18.9	0.12	20.5	0.09	1.5	0.03
PH			24	76.3	0.24	18.8	0.12	20.4	0.09	1.4	0.03
P-value				*0.7134*		*0.5932*		*0.5923*		*0.5720*	
Genetic line (GL)											
X			36	76.7	0.24	19.6^{a}	0.12	20.2^{b}	0.09	1.5^{a}	0.03
Y			36	76.2	0.24	18.2^{b}	0.12	20.7^{a}	0.09	1.4^{b}	0.03
P-value				*0.1676*		*<0.0001*		*0.0004*		*0.0066*	
Sex											
F			36	76.7	0.23	19.6^{a}	0.12	20.0^{b}	0.08	1.6^{a}	0.03
M			36	76.2	0.23	18.2^{b}	0.12	20.8^{a}	0.08	1.3^{b}	0.03
P-value				*0.1534*		*<0.0001*		*<0.0001*		*<0.0001*	
Experiment (EXP)											
A			36	76.7	0.29	19.6^{a}	0.15	20.2^{b}	0.11	1.5	0.04
B			36	76.1	0.28	19.5^{a}	0.14	20.1^{b}	0.11	1.4	0.04
C				76.6	0.39	17.7^{b}	0.20	21.0^{a}	0.14	1.4	0.05
P-value				*0.2266*		*<0.0001*		*0.0001*		*0.3580*	
TRT		**GL**									
BD		X	12	77.1	0.35	19.9^{a}	0.18	20.1	0.12	1.5	0.05
		Y	12	76.1	0.35	18.1^{d}	0.18	20.8	0.12	1.4	0.05
PR		X	12	76.5	0.39	19.3^{bc}	0.20	20.3	0.13	1.5	0.05
		Y	12	76.4	0.36	18.5^{dc}	0.19	20.6	0.12	1.4	0.05
PH		X	12	76.5	0.36	19.6^{ab}	0.19	20.1	0.12	1.5	0.05
		Y	12	76.1	0.38	18.1^{d}	0.20	20.6	0.13	1.3	0.05
P-value				*0.3510*		*0.0218*		*0.1784*		*0.7382*	
TRT		**Sex**									
BD		F	12	76.7	0.36	19.5	0.19	20.0	0.12	1.6	0.05
		M	12	76.5	0.36	18.5	0.19	20.8	0.12	1.3	0.05
PR		F	12	76.8	0.35	19.7	0.18	20.0	0.12	1.6	0.05
		M	12	76.0	0.37	18.1	0.19	21.0	0.13	1.3	0.05
PH		F	12	76.6	0.36	19.5	0.19	20.0	0.13	1.6	0.05
		M	12	76.0	0.35	18.2	0.18	20.7	0.12	1.2	0.05
P-value				*0.6956*		*0.2003*		*0.5371*		*0.2347*	
GL		**Sex**									
X		F	18	77.0	0.28	20.1	0.14	19.9^{c}	0.10	1.6	0.04
		M	18	76.4	0.41	19.0	0.22	20.5^{b}	0.14	1.4	0.06
Y		F	18	76.4	0.35	19.0	0.18	20.1^{cb}	0.12	1.5	0.05
		M	18	76.0	0.28	17.5	0.15	21.2^{a}	0.10	1.2	0.04
P-value				*0.7640*		*0.1776*		*0.0318*		*0.5458*	
GL	**Sex**	**TRT**									
X	F	BD	6	77.2	0.47	20.2^{ab}	0.25	19.8	0.16	1.7	0.06
		PR	6	76.7	0.47	19.8^{ab}	0.25	19.8	0.16	1.6	0.06
		PH	6	77.2	0.47	20.5^{a}	0.25	20.0	0.16	1.7	0.06
	M	BD	6	77.1	0.53	19.6^{bc}	0.28	20.4	0.18	1.4	0.07
		PR	6	76.3	0.62	18.7^{de}	0.33	20.8	0.21	1.4	0.09
		PH	6	75.9	0.55	18.7^{de}	0.29	20.3	0.18	1.3	0.07
Y	F	BD	6	76.2	0.52	18.9^{cde}	0.27	20.3	0.18	1.5	0.07
		PR	6	77.0	0.51	19.6^{bcd}	0.27	20.1	0.17	1.6	0.07
		PH	6	76.0	0.54	18.6^{e}	0.28	20.0	0.18	1.5	0.07
	M	BD	6	75.9	0.47	17.4^{f}	0.25	21.3	0.16	1.3	0.06
		PR	6	75.8	0.48	17.4^{f}	0.25	21.1	0.16	1.3	0.07
		PH	6	76.2	0.48	17.6^{f}	0.25	21.2	0.16	1.1	0.07
P-value				*0.1869*		*0.0181*		*0.1813*		*0.5572*	

[1]Values are presented as least squares means (LSmeans). SE: Standard error. BD: Basal diet; PR: Basal diet + probiotic; PH: Basal diet + phytobiotic. F: female; M: male. X and Y correspond to two fast growing commercial poultry genetic lines.

[a–d]Within a column, values without a common superscript differ, $P < 0.05$.

Table S12: Variability of individual body weight of sampled animals.

				Individual BW			
				N	Mean	SD	CV
Genetic line			Day				
			7	216	194	25.6	13.2
	X		21	216	995	110.7	11.1
			35	216	2280	302.6	13.3
			7	216	189	21.5	11.4
	Y		21	216	891	119.2	13.4
			35	216	2029	322.4	15.9
Experiment		Genetic line	Day				
			7	72	194	27.4	14.2
		X	21	72	996	111.2	11.2
			35	72	2403	268.1	11.2
	A		7	72	189	22.7	12.0
		Y	21	72	892	124.9	14.0
			35	72	2142	228.0	10.7
			7	72	201	24.9	12.4
		X	21	72	1004	115.1	11.5
			35	72	2361	285.6	12.1
	B		7	72	195	20.7	10.6
		Y	21	72	942	94.2	10.0
			35	72	2183	280.5	12.9
			7	72	189	23.3	12.3
		X	21	72	985	106.2	10.8
			35	72	2077	245.1	11.8
	C		7	72	182	19.5	10.7
		Y	21	72	841	115.6	13.7
			35	72	1762	274.0	15.6
Experiment	Pen	Genetic line	Day				
			7	6	190	32.1	16.9
	1	X	21	6	1020	72.6	7.1
			35	6	2660	186.4	7.0
			7	6	193	41.2	21.3
	2	Y	21	6	848	116.0	13.7
			35	6	2238	148.3	6.6
			7	6	196	31.3	16.0
	3	X	21	6	994	66.0	6.6
			35	6	2227	250.1	11.2
			7	6	190	20.2	10.6
	4	X	21	6	965	96.4	10.0
			35	6	2244	158.2	7.1
			7	6	183	21.0	11.5
	5	Y	21	6	798	81.9	10.3
			35	6	2081	145.2	7.0
			7	6	183	24.7	13.5
	6	X	21	6	990	159.3	16.1
			35	6	2390	172.0	7.2
			7	6	197	25.0	12.7
	7	X	21	6	1068	98.6	9.2
			35	6	2661	265.4	10.0
			7	6	183	35.0	19.1
	8	Y	21	6	1008	159.2	15.8
A			35	6	2139	234.3	11.0
			7	6	180	34.5	19.1
	9	X	21	6	904	93.1	10.3
			35	6	2229	249.1	11.2
			7	6	186	19.4	10.4
	10	Y	21	6	798	104.4	13.1
			35	6	2143	144.9	6.8
			7	6	185	10.8	5.9
	11	Y	21	6	878	131.9	15.0
			35	6	2161	229.8	10.6
			7	6	183	19.4	10.6
	12	Y	21	6	909	124.0	13.6
			35	6	2156	333.7	15.5
			7	6	190	22.4	11.8
	13	X	21	6	1013	97.9	9.7
			35	6	2562	368.6	14.4
			7	6	184	16.8	9.1
	14	Y	21	6	929	110.0	11.8
			35	6	2127	214.6	10.1
			7	6	202	28.7	14.2
	15	X	21	6	1047	75.8	7.2
			35	6	2309	111.6	4.8
	16	X	7	6	199	20.3	10.2
			21	6	1048	157.4	15.0

			35	6	2585	289.4	11.2
			7	6	206	28.2	13.7
	17	Y	21	6	974	165.2	17.0
			35	6	2221	290.5	13.1
			7	6	189	35.7	18.9
	18	X	21	6	961	114.5	11.9
			35	6	2234	138.3	6.2
			7	6	184	19.7	10.7
	19	Y	21	6	856	61.8	7.2
			35	6	2025	231.0	11.4
			7	6	198	25.4	12.8
	20	X	21	6	956	135.1	14.1
			35	6	2253	108.7	4.8
			7	6	193	17.9	9.3
	21	Y	21	6	889	141.5	15.9
			35	6	2213	328.2	14.8
			7	6	208	36.4	17.5
	22	X	21	6	989	114.6	11.6
			35	6	2481	246.4	9.9
			7	6	198	21.8	11.0
	23	Y	21	6	925	90.1	9.7
			35	6	2232	224.7	10.1
			7	6	185	8.8	4.8
	24	Y	21	6	886	92.0	10.4
			35	6	1963	137.0	7.0
			7	6	197	27.0	13.7
	25	X	21	6	1010	46.2	4.6
			35	6	2162	159.4	7.4
			7	6	187	16.4	8.7
	26	Y	21	6	924	164.2	17.8
			35	6	2296	260.1	11.3
			7	6	190	27.4	14.4
	27	X	21	6	973	178.1	18.3
			35	6	2465	224.8	9.1
			7	6	187	14.1	7.6
	28	Y	21	6	955	102.4	10.7
			35	6	2016	189.4	9.4
			7	6	178	26.6	15.0
	29	Y	21	6	951	87.2	9.2
			35	6	1988	197.4	9.9
			7	6	183	15.4	8.5
	30	X	21	6	1008	79.4	7.9
			35	6	2170	187.5	8.6
			7	6	208	26.5	12.8
	31	X	21	6	1055	91.2	8.6
			35	6	2494	340.8	13.7
			7	6	204	17.1	8.4
	32	X	21	6	1019	108.7	10.7
			35	6	2380	324.4	13.6
			7	6	196	21.2	10.8
B	33	Y	21	6	965	81.1	8.4
			35	6	2075	349.6	16.9
			7	6	190	21.7	11.4
	34	Y	21	6	922	49.6	5.4
			35	6	2041	170.0	8.3
			7	6	207	25.4	12.2
	35	X	21	6	929	70.5	7.6
			35	6	2234	227.5	10.2
			7	6	205	17.3	8.4
	36	Y	21	6	987	125.1	12.7
			35	6	2488	263.0	10.6
			7	6	201	34.7	17.3
	37	X	21	6	1009	102.9	10.2
			35	6	2228	236.1	10.6
			7	6	194	26.3	13.6
	38	X	21	6	1001	34.1	3.4
			35	6	2335	164.5	7.0
			7	6	196	18.1	9.3
	39	Y	21	6	972	65.4	6.7
			35	6	2138	112.1	5.3
			7	6	208	31.6	15.2
	40	X	21	6	953	130.7	13.7
			35	6	2511	383.6	15.3
			7	6	201	21.4	10.6
	41	X	21	6	1033	232.1	22.5
			35	6	2554	193.1	7.6
	42	Y	7	6	200	20.7	10.4

			21	6	896	116.4	13.0
			35	6	2370	276.1	11.7
			7	6	204	24.6	12.0
	43	Y	21	6	968	75.3	7.8
			35	6	2288	217.9	9.5
			7	6	203	11.8	5.8
	44	Y	21	6	964	93.3	9.7
			35	6	2432	231.4	9.5
			7	6	190	19.3	10.1
	45	Y	21	6	903	19.1	2.1
			35	6	2023	235.1	11.6
			7	6	201	29.6	14.8
	46	Y	21	6	895	100.6	11.2
			35	6	2039	278.5	13.7
			7	6	200	20.0	10.0
	47	X	21	6	1034	88.0	8.5
			35	6	2157	261.7	12.1
			7	6	216	25.3	11.7
	48	X	21	6	1023	124.3	12.2
			35	6	2643	232.1	8.8
			7	6	187	25.8	13.8
	49	X	21	6	1078	120.6	11.2
			35	6	2327	282.2	12.1
			7	6	172	28.4	16.5
	50	X	21	6	982	112.3	11.4
			35	6	2031	80.5	4.0
			7	6	181	15.4	8.5
	51	X	21	6	973	100.5	10.3
			35	6	1993	316.6	15.9
			7	6	178	15.2	8.5
	52	Y	21	6	873	85.6	9.8
			35	6	1647	242.5	14.7
			7	6	183	24.4	13.3
	53	X	21	6	954	98.4	10.3
			35	6	2020	184.6	9.1
			7	6	170	12.3	7.2
	54	Y	21	6	764	70.2	9.2
			35	6	1798	334.3	18.6
			7	6	180	18.5	10.3
	55	Y	21	6	838	141.2	16.9
			35	6	1546	159.3	10.3
			7	6	179	25.0	14.0
	56	Y	21	6	859	87.1	10.2
			35	6	1699	285.5	16.8
			7	6	177	19.9	11.2
	57	Y	21	6	811	136.7	16.9
			35	6	1839	215.9	11.7
			7	6	184	20.4	11.1
	58	Y	21	6	808	105.0	13.0
C			35	6	1523	130.4	8.6
			7	6	199	22.2	11.2
	59	X	21	6	998	98.8	9.9
			35	6	1974	105.9	5.4
			7	6	186	22.1	11.9
	60	X	21	6	983	124.6	12.7
			35	6	2097	360.4	17.2
			7	6	184	22.8	12.4
	61	X	21	6	942	58.7	6.2
			35	6	2185	252.1	11.5
			7	6	176	18.5	10.5
	62	X	21	6	917	81.0	8.8
			35	6	1882	133.6	7.1
			7	6	189	12.1	6.4
	63	X	21	6	934	114.6	12.3
			35	6	2157	297.1	13.8
			7	6	181	24.2	13.4
	64	Y	21	6	797	76.7	9.6
			35	6	1653	211.0	12.8
			7	6	211	32.7	15.5
	65	X	21	6	1087	126.9	11.7
			35	6	2017	245.4	12.2
			7	6	177	18.4	10.4
	66	Y	21	6	838	161.6	19.3
			35	6	1855	178.7	9.6
			7	6	188	7.5	4.0
	67	Y	21	6	871	140.7	16.1
			35	6	1960	179.2	9.1
	68	X	7	6	191	16.0	8.4

		21	6	932	53.9	5.8
		35	6	2003	133.1	6.6
		7	6	195	26.9	13.8
69	Y	21	6	1004	70.2	7.0
		35	6	2077	343.2	16.5
		7	6	180	9.2	5.1
70	Y	21	6	846	88.7	10.5
		35	6	1759	289.2	16.4
		7	6	207	17.9	8.7
71	X	21	6	1044	55.4	5.3
		35	6	2239	177.5	7.9
		7	6	198	25.5	12.9
72	Y	21	6	779	66.2	8.5
		35	6	1792	263.3	14.7
All			1296	1096.4	835.3	76.2

[1]Values are presented as means. SD: Standard deviation; CV: coefficient of variation. BW: Body weight.
BD: Basal diet; PR: Basal diet + probiotic; PH: Basal diet + phytobiotic. F: female; M: male. X and Y correspond to two fast growing commercial poultry genetic lines.