eman ta zabal zazu

Universidad del País Vasco

Euskal Herriko Unibertsitatea

NEW INSIGHTS IN THE PATERNAL GENETIC LANDSCAPE OF SOUTHWESTERN EUROPE: DISSECTION OF HAPLOGROUP R1b-M269 AND FORENSIC APPLICATIONS

PATRICIA VILLAESCUSA URBANEJA

PhD Thesis

Directed by:

Professor Doctor María de los Ángeles Martínez de Pancorbo Gómez

BIOMICs Research Group

Department of Zoology and Animal Biology

University of the Basque Country UPV/EHU

Vitoria-Gasteiz, 2019

The present work has been carried out thanks to the grant of a predoctoral fellowship, within the program of Training of Research Staff by the University of the Basque Country UPV/EHU, which I have benefited from among the years 2015 and 2019. The funds for the accomplishment of the analysis have come from BIOMICs Research Group, consolidated by the Department of Education, Universities and Research of the Basque Government since 2012 (IT-424-07 and IT-833-13) and of the University of the Basque Country UPV/EHU (ELDUNANOTEK UFI11/32). In addition to that, SGIker (UPV/EHU, MICINN, GV/EJ, FSE) has provided technical and human support, especially from the DNA Bank of the University of the Basque Country UPV/EHU.

“El único modo de lograr lo imposible, es convenciéndose de que sí es posible”

Alicia,

Alicia en el país de las maravillas.

Acknowledgements

In the first place, I would like to extend my sincere gratitude to all the people who have helped, guided and encouraged me during the long and successful journey that was the path to obtain my doctoral thesis. Without their active support, this project would not have been possible.

I am remarkably indebted to my mentor Prof. Dr. Marian Martínez de Pancorbo, who believed in me and gave me the support to obtain my predoctoral fellowship. Her guidance and encouragement during these years have been essential in my training both as a researcher and as a person. In addition to that, I am extremely grateful to my lab colleagues, both present and past, who have shared experiences and accompanied me during the long lab hours. Miriam, Carol, Leire, Andrés, Endika, Maite, Tamara, Lourdes, David, Melania, Majo, Sergio, Amara, among others. *Sois los mejores compañeros de laboratorio que podría haber deseado, gracias por los innumerables buenos momentos, y por enseñarme y prestarme ayuda con todas las dudas, tanto profesionales como existenciales, que he tenido a lo largo de estos años.*

I would also like to thank Prof. Dr. Francesc Calafell, for his kind assistance and successful collaboration throughout the project.

I have to express my sincere appreciation to Prof. Dr. Lutz Roewer and to Dr. Maria Geppert for receiving me in Berlin. Thank you so much for your hospitality and guidance. Jessy, Jessman, Sascha, Marion, Steffi, Judith, Petra and the rest of the lab technicians, *vielen dank für ihren freundlichen empfang!*

Last but not least, my gratitude goes to my parents, whose support made this possible, as well as to all of my friends, who both directly or indirectly supported me during these years. Ailander, you have been there for me at all times, making the tough times less hard. No words can express my gratitude to you.

Thank you very much!

Patricia Villaescusa Urbaneja

Abstract

The human Y chromosome determines male sex and possess a haploid nature, escaping crossing over during meiosis. These traits make this chromosome male specific, which is transmitted from fathers to sons practically unchanged establishing paternal lineages. In the last years the study of genetic markers located in the human Y chromosome, that is, Y chromosome *short tandem repeats* (Y-STRs) and *single nucleotide polymorphisms* (Y-SNPs), has become relevant in the area of Forensic Genetics as a powerful tool to exclude male suspects from the involvement in a crime, identify the paternal lineage and biogeographical ancestry of male perpetrators, as well as to provide investigative clues to find an unknown male perpetrator or contributor or to a crime. In addition to that, the analysis of the Y chromosome can also be applied in other areas like population genetics and genetic genealogy for the study of paternity and/or kinship, detecting past and recent male mediated expansions, familiar searching, and evolutionary studies.

Although highly useful, the analysis of the Y chromosome is limited in Forensic Genetics due to its haploidy and inheritance, which makes it less effective for identification purposes in comparison with autosomal markers. Moreover, it cannot distinguish between individuals from the same paternal lineage. However, its utility cannot be disregarded since the vast majority of the crimes where DNA evidence is helpful are committed by male perpetrators.

The study of Y-SNPs has revealed that the distribution of paternal lineages, or haplogroups, around the word is restricted to concrete geographic areas at continental and regional level, allowing to reconstruct the evolutionary history of lineages. In this context, the current genetic makeup of Europe has been the object of multiple studies, as a controversy arose around the Paleolithic or Neolithic origin of the most common paternal lineage in West Europe, R1b-M269.

The main objective of the present doctoral thesis work focuses on the reconstruction of the most probable evolutionary scenario of the main European paternal lineage R1b-M269 in the Iberian Peninsula and Southwest Europe through the dissection in its subhaplogroups. The analysis of M269 and its sublineages in populations of Southwestern Europe allowed us to further characterize the paternal genetic landscape of that region, revealing a different distribution pattern from the one proposed before by other authors for R1b-S116, one of the main M269 sublineages, and resolving the paragroup S116* by the presence of the Iberian near-specific haplogroup R1b-DF27, which occupies a different geographic area than the other S116 subhaplogroups and is present in the Iberian Peninsula in frequencies over 30-50%.

On the other hand, the analysis and dissection of DF27 has revealed that this lineage is also present in Latin America due to historical events, and that some of its sublineages display moderate geographic differentiation in Iberia, making DF27 a potentially useful marker in Forensic Genetics for determining Iberian or southwest European paternal biogeographical ancestry. Furthermore, the estimated *times to the most recent common ancestor* (TMRCA) show that DF27 originated recently, around 4,000-4,200 years ago, at the transition between the Neolithic and the Bronze Age that remodeled the Y chromosome landscape of Europe.

In addition to that, in forensic and population genetics the necessity of new multiplex tools that allow the simultaneous genotyping of several genetic markers in a unique reaction has proven to be a critical demand. These multiplex tools can be composed of either Y-STRs and/or Y-SNPs, and are based on capillary electrophoresis, minisequencing, or massive parallel sequencing (MPS) technology, although the last one is not yet implemented in all Forensic Genetics laboratories. To respond to the increasing demand of Y-SNP multiplexes with higher haplogroup resolution, and of Y-STR multiplexes able to resolve the particular cases that the current Y-STR panels are not able to respond to, two novel multiplex panels were developed in the present work. On the one hand, the 15-plex Y-SNP minisequencing panel allows the fine subtyping of the haplogroup DF27, suitable for the inference of Iberian and southwest European paternal biogeographical ancestry. On the other hand, the *slowly mutating* (SM) Y-STR panel could be helpful in conjunction with current Y-STRs panels for confirming exclusions in kinship cases where minimal discrepancies in one or a few loci are reported using only the regular Y-STR panels, as well as for evolutionary studies.

All in all, the studies conducted within the present doctoral thesis work have provided new clues about the evolutionary history of the current paternal genetic landscape of Europe, as well as an extensive genetic reference dataset of forensic and population interest for future applications.

Resumen

El cromosoma Y humano determina el sexo masculino y posee una naturaleza haploide, sin estar sujeto en la mayor parte de su longitud a recombinación, debido a que escapa del entrecruzamiento cromosómico durante la meiosis. Estas características hacen que este cromosoma sea específico del sexo masculino y que se trasmita prácticamente sin cambios de padres a hijos, estableciendo lo que se conoce como linajes paternos. El estudio de los linajes del cromosoma Y, mediante polimorfismos de un solo nucleótido (*single nucleotide polimorphism*, SNP) en el cromosoma Y (Y-SNP), permite la reconstrucción de la historia evolutiva de los linajes paternos de la especie humana desde sus inicios en África hace al menos 300.000 años.

El conocimiento de la estructura y propiedades del cromosoma Y obtenido a través de la genética de poblaciones humanas mediante el análisis de sus marcadores genéticos es la base para el estudio de las migraciones humanas y sus aplicaciones en genética forense y otras áreas afines como la genealogía genética o la genética evolutiva. Estos marcadores pueden ser de dos tipos, los anteriormente mencionados Y-SNPs, y los microsatélites Y-STRs (*short tandem repeat*, STR). Su estudio permite conocer cuáles son los linajes paternos característicos o presentes en cada población humana y entender cómo se distribuyen en las mismas, permitiendo así diferenciar entre unas poblaciones y otras. Por ello, en los últimos años su estudio es relevante dentro del área de la Genética Forense debido a que el análisis de estos marcadores permite, entre otras cosas, excluir a un sospechoso masculino como contribuyente de restos biológicos presentes en la escena de un delito, identificar el linaje paterno y la ancestralidad biogeográfica de los sospechosos, y proporcionar pistas para la identificación del autor o contribuyente masculino de un delito. Además de sus aplicaciones forenses, los marcadores genéticos del cromosoma Y también pueden aplicarse al estudio de paternidad o parentesco biológico, la búsqueda de familiares desaparecidos y la detección de movimientos migratorios masculinos tanto pasados como recientes.

A pesar de su gran utilidad, el análisis del cromosoma Y presenta varias limitaciones que restringen su uso en rutina forense. Sus marcadores genéticos son menos efectivos para la identificación de individuos que los marcadores autosómicos debido a su modo de herencia, que impide distinguir entre los individuos de un mismo linaje paterno. Padres e hijos, además de familiares emparentados por línea paterna, poseen el mismo cromosoma Y. No obstante, debido al hecho de que la mayoría de los delitos con pruebas de ADN informativas son cometidos por individuos varones, la utilidad del análisis de marcadores genéticos del cromosoma Y en rutina forense es de gran ayuda. En particular, el análisis de estos marcadores resulta de gran valor en delitos

cometidos por varones, tal como agresiones sexuales, casos en los que el resto de marcadores genéticos (como los autosómicos) hayan fallado, o en casos en los que se quiera obtener información sobre la ancestralidad biogeográfica paterna de un sospechoso.

El estudio de los Y-SNPs ha revelado que las agrupaciones de linajes paternos, también llamadas haplogrupos, se distribuyen en áreas geográficas concretas a lo largo del mundo, tanto continental como regionalmente. De manera que ciertos haplogrupos son más comunes en determinadas zonas del planeta o en algunos grupos étnicos mientras que otros se encuentran más diseminados, dependiendo todo ello de la historia evolutiva de cada población o linaje. En el caso de Europa, los linajes paternos más comunes son R1b en el oeste del continente y R1a en el este, ambos pertenecientes al macrohaplogrupo R.

La actual composición genética de Europa ha sido objeto de múltiples estudios poblacionales desde hace varios años, y fruto de ello surgió una gran controversia alrededor del origen del haplogrupo más común en el oeste de Europa, R1b-M269. Las estimaciones de edad obtenidas a partir de los distintos estudios genéticos realizados por distintos autores situaron el origen de este linaje paterno tanto durante el periodo Paleolítico, como en tiempos más recientes, en el Neolítico. También surgió cierto debate alrededor de su lugar de origen, que según los autores que apoyan un origen de M269 en Paleolítico se situaría en la región Franco-Cantábrica y, según los autores que defienden su origen más reciente en el Neolítico, se situaría en el este de Europa. A fin de resolver las cuestiones sobre su origen, es necesario un mayor estudio de la estructura de M269 y su distribución a lo largo de Europa, cubriendo la mayor parte de territorio posible, ya que los estudios publicados hasta la fecha no han aportado suficiente información sobre algunas áreas concretas del continente, como la zona de la cornisa atlántica. También sería necesaria una revisión de los métodos y los parámetros utilizados para estimar la edad de dicho haplogrupo.

Teniendo en cuenta lo mencionado anteriormente, el objetivo principal de este trabajo de tesis doctoral se centra en la reconstrucción del escenario evolutivo más probable del principal linaje paterno europeo M269 en la Península Ibérica y el suroeste de Europa a través de la disección en sus subhaplogrupos. Esto permitirá caracterizar de manera detallada la distribución de los linajes paternos presentes en la Península Ibérica e inferir el papel de esta región en la historia evolutiva de Europa, lo cual explicaría la distribución actual de la mayoría de los haplogrupos del suroeste de europeo.

El análisis de M269 y de sus subhaplogrupos en poblaciones del suroeste de Europa nos ha permitido caracterizar el paisaje genético paterno en esa región, revelando que uno de los sublinajes principales de M269, R1b-S116, presenta un patrón de distribución distinto al

propuesto anteriormente por otros autores, observándose un gradiente de frecuencias decreciente con la distancia desde el norte de la Península Ibérica, la costa oeste francesa y las islas británicas. Por otro lado, el paragrupo de S116, S116*, fue prácticamente resuelto por la presencia del sublinaje R1b-DF27, que se distribuye en un área geográfica distinta de las de R1b-U152 y R1b-M529, los otros dos subhaplogrupos principales de S116.

En cuanto al origen de M269, los resultados obtenidos sugieren que se originó en el este de Europa, apareciendo sus sublinajes durante la ola de avance a medida que se extendió por el resto de Europa. De este modo, teniendo en cuenta los datos aportados por el análisis de Y-SNPs e Y-STRs, y los resultados obtenidos por otros autores en estudios paralelos, se podría descartar el origen de M269 en el área Franco-Cantábrica. El cálculo de la antigüedad de M269 ha reforzado la problemática existente alrededor de los métodos para la estimación de la edad de un haplogrupo a partir de Y-STRs, dependiente en gran medida de la tasa de mutación elegida para el cálculo. No obstante, los últimos estudios de secuenciación masiva de la región específica masculina del cromosoma Y parecen haber aportado una escala temporal fiable para la diversidad del cromosoma Y.

DF27 ha resultado ser un linaje paterno casi específico de la Península Ibérica, estando presente en frecuencias superiores al 30-50% y que disminuyen de manera drástica al 6-20% fuera de esa región. Además, también se ha detectado la presencia de este linaje en Latinoamérica debido a sucesos históricos, estando ausente en África y Asia. Por otro lado, la disección de DF27 ha permitido conocer que algunos de sus subhaplogrupos, como R1b-L176.2 y R1b-Z220, presentan cierta diferenciación geográfica dentro de la Península Ibérica, dividiéndola en este y centro-norte respectivamente. Todo ello hace de DF27 un marcador de potencial interés en Genética Forense para la determinación de la ancestralidad biogeográfica paterna Ibérica y/o del suroeste europeo.

Las altas frecuencias del paragrupo de DF27, DF27*, podrían indicar la presencia de nuevos sublinajes todavía no conocidos de este haplogrupo, aunque no haya sido observado un patrón de variación interna a través de los *median joining networks* construidos con los individuos que forman parte del paragrupo. No obstante, la caracterización de la secuencia completa del cromosoma Y a través de secuenciación paralela masiva (MPS) permitiría la resolución completa y fiable del paragrupo.

Por otra parte, las estimaciones del tiempo hasta el ancestro común más reciente (*time to the most recent common ancestor*, TMRCA) sitúan el origen de DF27 recientemente, hace 4.000-4.200 años durante el periodo de transición entre el Neolítico y la Edad de Bronce que remodeló el paisaje genético paterno de Europa. En cuanto al lugar de origen de DF27, aunque no ha sido

posible determinarlo con precisión, el noreste de la Península Ibérica es la región más probable de origen, teniendo en cuenta la diversidad interna de los Y-STRs y los TMRCAs estimados en distintas poblaciones.

Asimismo, en el área de la Genética Forense y de poblaciones ha surgido en los últimos años una creciente demanda de nuevas herramientas que permitan el genotipado simultáneo de múltiples marcadores genéticos en una única reacción. Las herramientas multiplex más utilizadas para el análisis de Y-STRs o Y-SNPs consisten en técnicas de análisis de longitud de fragmentos de STRs o minisecuenciación de SNPs mediante electroforesis capilar.

En el caso de los Y-SNPs, es necesario el desarrollo de paneles que permitan un alto nivel de resolución de distintos haplogrupos, mientras que en el caso de los Y-STRs, es de interés el desarrollo de nuevos paneles complementarios a los ya existentes. Para responder a la creciente demanda de herramientas multiplex, en el presente trabajo de tesis se han desarrollado dos paneles multiplex de Y-SNPs e Y-STRs para su uso en Genética Forense y de poblaciones.

Por un lado, se ha desarrollado el panel Y-SNPs 15-plex basado en la tecnología de minisecuenciación, que permite el subtipaje a alta resolución del haplogrupo DF27 y es una herramienta robusta y reproducible. Además de DF27 y sus subhaplogrupos, el panel también incluye otros sublinajes de M269 por encima de DF27, lo que hace su uso adecuado a la Genética Forense a la hora de inferir la ascendencia biogeográfica paterna Ibérica o del suroeste europeo de un vestigio, ya que varios de los Y-SNPs incluidos presentan una marcada o moderada diferenciación geográfica, en particular S116, U106, U152, M529, DF27, M167 y Z220.

Por otra parte, el panel 15-plex también se puede aplicar para el estudio de la introgresión de la población del suroeste de Europa en poblaciones que hayan sido destino de migraciones históricas españolas y portuguesas, como Latinoamérica, u otras áreas del mundo influenciadas históricamente por presencia hispana, entre ellas Flandes, Cerdeña, Sicilia o Filipinas. Aunque otros paneles de minisecuenciación publicados previamente por otros autores incluyen el haplogrupo R1b, no ofrecen un alto nivel de resolución de este haplogrupo. Por ello, el uso del panel 15-plex en combinación con otros paneles de minisecuenciación de Y-SNPs que incluyan una baja resolución del haplogrupo R1b puede ofrecer un mayor poder resolutivo para los haplogrupos europeos con un consumo mínimo de muestra de ADN, algo crítico en muestras de origen forense. El panel 15-plex también puede utilizarse en combinación con el análisis del ADN mitocondrial, cuya información sobre el linaje materno puede completar la información sobre la ancestralidad de los individuos.

Debido a la falta de muestras reales de algunos de los Y-SNPs incluidos en la herramienta 15-plex, se utilizó la técnica de mutagénesis dirigida para generar de manera *in vitro* las variantes derivadas y, así, comprobar la capacidad del panel para detectarlas. La mutagénesis dirigida es una técnica de biología molecular que permite la creación de mutaciones puntuales en el ADN, además de inserciones y deleciones. Por lo tanto, la mutagénesis dirigida se postula como una técnica muy adecuada para producir variantes genéticas *in vitro* para su uso en reacciones de minisecuenciación, y permitir así la inclusión de todas las variantes durante el proceso de optimización de paneles multiplex.

Por otro lado, el segundo panel desarrollado en el presente trabajo incluye seis Y-STRs de tasa de mutación baja, también llamados *slowly mutating* (SM) Y-STRs. Los paneles de Y-STRs de rutina forense actuales incluyen marcadores de tasa de mutación desde baja a alta, siendo los de alta tasa más adecuados para aplicaciones como la identificación de individuos y los de tasa media y/o baja para su aplicación en diagnóstico de parentesco, búsqueda de personas desaparecidas o estudios evolutivos.

A pesar de su gran utilidad, los paneles actuales no son capaces de resolver los casos de parentesco complejos, en los que las discrepancias mínimas son críticas y acaban siendo reportadas como exclusiones. Por ello, el panel SM Y-STRs es una herramienta de utilidad en Genética Forense en conjunción con los paneles de Y-STRs de rutina al incluir Y-STRs de baja tasa de mutación, adecuados para su uso en parentescos complejos. De este modo, el poder evaluar disparidades adicionales mediante el análisis de SM Y-STRs puede proporcionar evidencia adicional para la exclusión de parentesco biológico, ya que es más raro que se den eventos mutacionales en estos marcadores.

Además, los SM Y-STRs son también adecuados para su uso en estudios evolutivos, permitiendo optimizar y aumentar la resolución de la predicción de haplogrupos del cromosoma Y a partir de los Y-STRs. La adición de los seis SM Y-STRs a la hora de hacer la predicción de un haplogrupo incluyendo los Y-STRs de los paneles convencionales puede ayudar a obtener predicciones más fiables, al ser marcadores más estables debido a su tasa de mutación más baja. Los estudios de validación realizados con este panel han demostrado que es una herramienta reproducible, robusta y sensible para su uso con vestigios de origen forense, al obtenerse perfiles genéticos completos a partir de cantidades muy limitadas de ADN (200 pg) y en presencia de dos inhibidores típicos en rutina forense como son el ácido húmico y la hematina.

En conclusión, el presente trabajo de tesis doctoral ha proporcionado, por un lado, nuevas pistas sobre la historia evolutiva del acervo genético europeo actual, permitiendo caracterizar en mayor

detalle la distribución del haplogrupo M269 y varios de sus sublinajes y, por otro lado, dos nuevas herramientas de uso forense para el análisis multiplex de Y-SNPs e Y-STRs. Además, fruto del análisis de un gran número de marcadores en una extensa colección de muestras de distintas poblaciones mundiales se ha generado un extenso conjunto de datos de referencia de interés tanto forense como poblacional.

Table of contents

Abbreviations

ABC: approximate Bayesian computation

AMOVA: analysis of molecular variance

ASD: Average Square Distance

bp: base pair

CE: capillary electrophoresis

CNV: copy number variation

CODIS: combined DNA index system

DC: discrimination capacity

DNA: deoxyribonucleic acid

EMR: evolutionary mutation rate

ESS: European Standar Set

***et al.*:** and others (from latin: *et alii*)

FCA: factorial correspondence analysis

GD: genetic diversity

GMR: genealogical mutation rate

GWAS: genome-wide association study

HRM: High Resolution Melting

INDEL: insertion or deletion polymorphism

ISOGG: International Society of Genetic Genealogy

MDS: multidimensional scaling analysis

MPS: massive parallel sequencing

MRCA: most recent common ancestor

MSY: male-specific region of the Y chromosome

mtDNA: mitochondrial DNA

N: total number of samples

NGS: next generation sequencing

NRY: non-recombining region of the Y chromosome

P: significance value

PAR: pseudoautosomal región of the Y chromosome

PCA: principal component analysis

PCR: polimerase chain reaction

RFLP: restriction fragment length polymorphism

SBE: single base extension

SNP: single nucleotide polymorphism

STR: short tandem repeat

TMRCA: time to the most recent common ancestor

VNTR: variable number of tandem rpeats

YA: years ago

YCC: Y Chromosome Consortium

Y-SNP: Y chromosome SNP

Y-STR: Y chromosome STR

Table list

Figure list

1. Introduction

1.1 Human genetic variability

The sequence of the human genome is a well-organized library that encodes the genetic instructions for human molecular functions and contains rich information about human evolution. The *genome* is composed by deoxyribonucleic acid (DNA), which includes genes and non-coding DNA, as well as the genetic information stored in other organelles such as the mitochondria (mtDNA). More than six decades have passed since the definitive structure of the DNA was proposed by James Watson, Francis Crick and Maurice Wilkins [1,2] and, despite of that, our understanding of all the secrets hidden along the genome is far from complete.

In order to unravel these secrets in 1990 the Human Genome Project (HGP) was launched, the largest biomedical research project in history, with the involvement the International Human Genome Sequencing Consortium (IHGSC). In parallel, another project with the same aim was conducted outside government by the corporation Celera Genomics. This competition ended up with the concurrent release in 2001 of the draft sequences of the human genome [3,4], which provided a first overall view of the complete genome. Finally, the project was considered complete in 2003 [5], two years ahead of its original schedule, and in 2007 the fully sequenced genomes were released to the public [6,7]. As a follow-up of the HGP, in 2003 the project Encyclopedia of DNA Elements (ENCODE) was launched with the aim of defining all functional elements encoded in the human genome [8]. This project has provided interesting findings, like approximately 8.5% of the genome, which is composed of more than 3,000 billion base pairs, corresponds to DNA/protein binding regions and 2.9% corresponds to protein-coding gene exons, while among these last ones only 1.2% presents a specific function in protein coding [8–11].

Human *genetic variation* is defined as the genetic differences that are present in and among populations. Only a small fraction of the genome, 0.3%, differs between people and makes each individual unique with the exception of monozygotic twins [12]. Consequently, these variable regions of the human genome provide the capacity to use the information contained in DNA sequences to differentiate and to establish biological relationships among individuals through human identification and kinship testing.

Polymorphisms are known as variations in the form of different alleles, or variants, at a particular locus. There are three major forms of variation at DNA level: sequence polymorphisms, length polymorphisms and copy number variants (CNV) [13]. The study of these variants has medical, forensic and evolutionary applications, since they allow to differentiate between individuals, occur at different frequencies in different human populations, and some are related to diseases. The

efforts made by the scientific community for a better understanding of the human genome, the development of better statistical methods, and the advent of massive parallel sequencing (MPS) has ended up in the discovery of a great number of new polymorphisms in the last years.

DNA typing is a broad term that refers to a wide range of methods that allow studying genetic variations, which makes possible to obtain what is called a *genetic profile* and differentiate between individuals through the analysis of polymorphisms. Forensic Genetics is based on the analysis of genetic profiles, by the comparison of a profile obtained from a questioned sample with a reference or known sample, or DNA databases that contain profiles related to criminal cases [13,14].

1.2 Evolution and history of Forensic Genetics

1.2.1 Chronology of Forensic Genetics

Forensic Genetics can be defined as the application of genetics to the resolution of legal conflicts [15]. The first step in the development of Forensic Genetics was made by Karl Landsteiner in 1900, who described the ABO blood grouping system and observed that individuals could be classified in different groups according to their blood type [16]. This finding led to the realization that this variation was applicable in solving crimes and paternity testing cases [17].

During the first half of the twentieth century, the technique 'absorption-inhibition ABO typing' was developed, becoming a standard in forensic laboratories, and numerous blood group markers and soluble blood serum protein markers were discovered. Although their application to criminal casework was rather limited due to the quantity of biological material required to provide highly discriminating results, which is minimal or degraded in forensic cases, and the difficulty of analyzing body fluids other than blood [15,18].

The introduction of the first antigen on leucocytes, which was later known as HLA (Human Leukocyte Antigen), in 1958 by Jean Dausset [19] involved an important improvement in paternity testing, as the HLAs were more polymorphic than any other genetic markers used up to that date [15,20]. Nevertheless, the antigenic HLA determinations possessed the same technical limitations as the other serological genetic markers for their application in forensic cases.

The decade of 1980 can be considered as the starting point of a new era in Forensic Genetics. In 1980 the analysis of the first highly polymorphic locus was reported [21], called minisatellites or variable number of tandem repeats (VNTR). Just a few years later, in 1984, Alec Jeffreys realized the potential forensic application of the minisatellite DNA by discovering their high levels of

variability [22–24], and developed the technique for their analysis that generated the well-known multiband patterns known as DNA fingerprints (Figure 1). The first time that DNA testing was applied in a forensic setting was in 1986, in the case that culminated in the conviction of Colin Pitchfork for a double rape and homicide in Leicestershire [13,25].

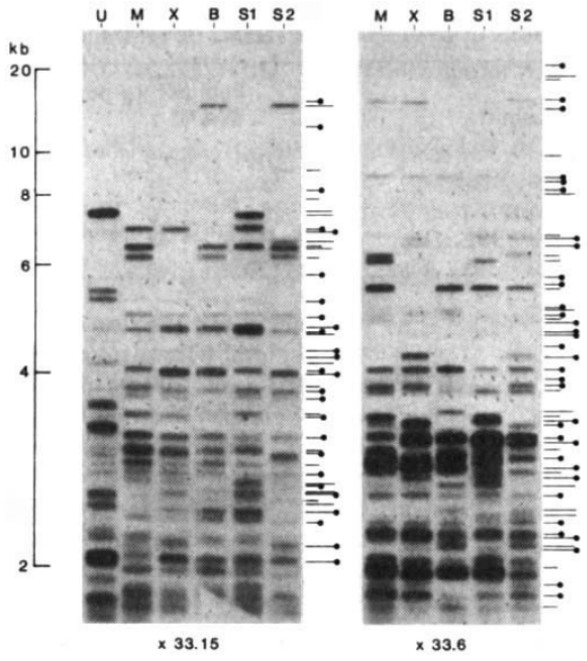


Figure 1. First application of *DNA fingerprint* in an immigration dispute. The DNA fingerprints correspond to a Ghanaian family involved in an immigration dispute. U: Unrelated individual; M: Mother; X: the boy in the dispute; B: Brother; S1: Sister 1; S2: Sister 2. Fragments present in the mother's DNA fingerprints are indicated by a short horizontal line; paternal fragments absent from M but present in at least one of the undisputed siblings (B, S1, S2) are marked with a long line. Maternal and paternal fragments transmitted to X are shown by a dot. Extracted from [24].

Although the analysis of minisatellite markers was very informative, the technique, called *DNA fingerprint*, was not very successful in forensics due to problems derived from the statistical evaluation of the evidence in cases of band matching, standardization, and quality requirement of the samples. In order to overcome the limited fragment resolution of the technique, forensic laboratories adhered to binning approaches, where fixed or floating bins were defined in relation to the observed DNA fragment size, and adjusted to the respective detection system [26–28]. For the reasons above mentioned, these multi-locus probes (MLP) were soon substituted in the forensic field by 'single locus probes' (SLPs).

Another critical point in the history of Forensic Genetics was the application of the polymerase chain reaction (PCR). The PCR was designed and named by Kary Mullis and colleagues from Cetus Corporation in 1983 [18,29,30], although the patent was not approved until years later in 1987 [31]. The successful application of PCR required considerable further development [32–34], as well as the

isolation of suitable heat-stable DNA polymerases available from thermophilic bacteria. This technique increased the sensitivity of DNA analysis to the point where DNA profiles could be obtained from only a few cells, enabling to analyze degraded DNA, a critical constraint at the time and even nowadays. It also made standardization easier by avoiding most of the statistical problems derived from the matching and binning of SLP bands [15].

The first application of PCR in a forensic case was for the analysis of single nucleotide polymorphisms (SNP) in the locus HLA-DQA1 (originally called HLA-DQα) using sequence-specific oligonucleotide probes (SSO) [35]. After the realization that it was feasible to analyze DNA markers through PCR, the first commercial kits became available. In forensic labs, the most popular kits were DQ Alpha AmpliType kit (Perkin-Elmer, Foster city, CA, USA), and the AmpliType Polymarker PCR Amplification kit (Perkin-Elmer, Foster city, CA, USA) [15,36].

Afterwards, the efforts of the forensic community were directed to the amplification of fragment length polymorphisms, among which microsatellites or *short tandem repeats* (STRs) are included. The minisatellite D1S80 was the first one that was used in forensic analysis, but these early PCR systems were soon replaced by the analysis of STRs, which are currently the most commonly used markers in Forensic Genetics [13,15]. Discovered in 1989 [37,38], they were applied for the first time in forensics in 1990, and their discrimination power was higher in comparison with the previously used genetic markers.

This was followed by a process of standardization of the techniques and the nomenclature of the markers, as well as an effort in quality control. As a result, the accreditation of forensic laboratories and professional certification of individuals became an important issue in Forensic Science [13,18]. Finally, during the decade of 1990 the first STR national database was released and the first fluorescent STR multiplex systems appeared [39–41].

Nowadays the most used methods in Forensic Genetics laboratories are the PCR systems based on STRs both for human identification and kinship testing, due to their higher power of discrimination, multiplex capability and rapid analysis speed [13]. The beginning of the 21st century has come with the evolution of molecular biology technologies that has enabled the discovery of new and more discriminative markers, as well as a trend in the last years to develop new molecular tools able to genotype more and more markers in a single PCR reaction. The more polymorphic markers that are studied, the more power of discrimination that can be achieved for the resolution of a case.

Apart from STRs, in the last 20 years other kind of markers have been studied for forensic analysis such as mtDNA, SNPs and insertion-deletion polymorphisms (INDELs). In the last decade, a huge progress has been in molecular biology since massively parallel sequencing (MPS) techniques, also known as next-generation sequencing (NGS), became available as of 2005. These platforms have also started to be used and evaluated for their use in Forensic Genetics [42,43]. Although its real application in routine casework is yet to be established, as its implementation in forensic laboratories has been debated due to the costs, the additional challenge of data processing, and sensitivity issues in comparison to traditional technologies. Figure 2 resumes the main highlights in the history of Forensic Genetics.

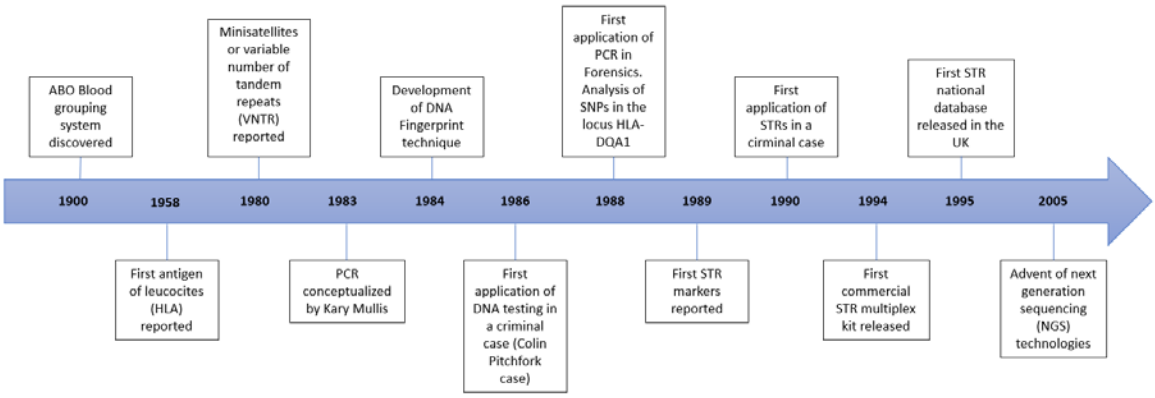


Figure 2. Main highlights in the history of Forensic Genetics.

1.2.2 Genetic markers commonly used in Forensic Genetics

Polymorphisms can be classified into three main groups:

- *Sequence variation polymorphisms:*

This type of polymorphism entails variation of one or more nucleotides in the DNA sequence. SNPs and the hypervariable regions of the mitochondrial DNA (HVI, HVII and HVIII) can be categorize into this group.

- *Length polymorphisms:*

Length polymorphisms are variants of the same locus differentiated by the number of nucleotides within the fragment of DNA. Minisatellites and microsatellites are two types of this variation.

- *Copy number variation (CNV) polymorphisms:*

These structural variants are segments of one kilobase (kb) or larger that are present at a variable copy number in comparison with a reference genome [44]. These polymorphisms can involve inversions, duplications, deletions and relocations of large segments of chromosomes [13].

In Forensic Genetics the most commonly analyzed polymorphisms are microsatellites, followed by SNPs.

1.2.2.1 Short Tandem Repeats (STRs)

Microsatellites or STRs are DNA regions with repeats units that are from 1 to 6 base pairs (bp) in length and are repeated typically 5-50 times [14,45]. STRs account for about 3% of the total human genome [14,46,47] and are scattered throughout the genome, occurring around every 10,000 nucleotides [47–49]. They have become the forensic markers of choice because of their high variability and easy amplification by PCR [14].

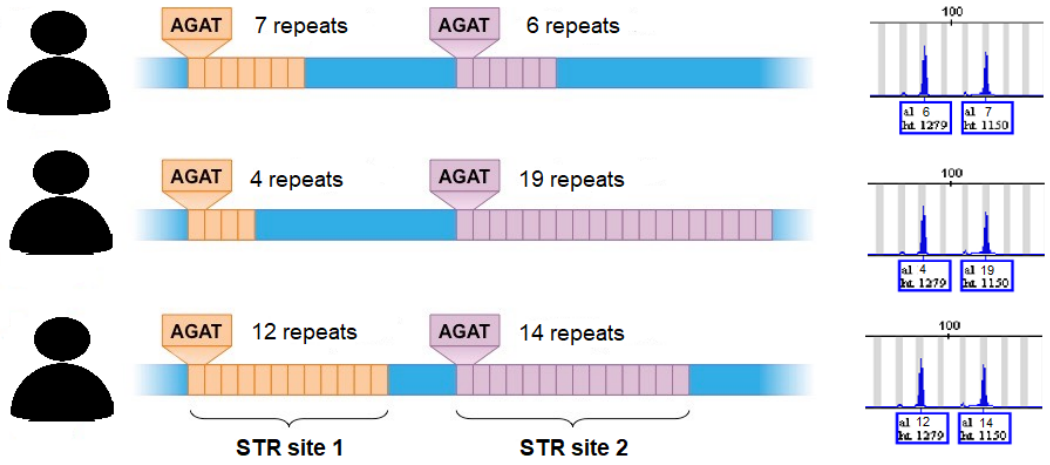


Figure 3. Example of DNA typing and profiling in different individuals.

The number of repeats in STR markers is highly variable among individuals, making these markers highly effective for human identification purposes. These markers may present a different number of alleles or repetitions at a specific locus (Figure 3). Eventually, all the genotypes of a set of markers comprise the genetic profile of an individual.

1.2.2.1.1 Types of STRs

STR repeat sequences are named by the length of the repeat unit, i.e. if there are three nucleotides in the repeats, they are called trinucleotides. Usually in the core repeat there can be mono-, di-, tri-, tetra-, penta-, and hexanucleotides [14]. The most popular STR systems for human identification

are tetranucleotides. On the one hand, di- and trinucleotides display greater 'stutter' percentages (typical STR amplification artifacts) than tetranucleotides (30% vs. 15%), and closely spaced heterozygotes are more difficult to resolve using size-based electrophoretic separations. On the other hand, penta- and hexanucleotides are less common in the human genome. STRs can also can be classified according to the rigor and degree of perfection of the repeat unit, into the categories 'simple', when it contains units of identical length and sequence; 'compound', when it comprises two or more adjacent simple repeats; or 'complex' , when it contains several non-consensus alleles that differ in size and sequence and are, therefore, difficult to genotype reproducibly [14,50] (Table 1).

Table 1. Types of variation in STR markers.

Repeat unit	Motif structure	Example repeat structure	Example STR
Simple			
Dinucleotide	$(AC)_n$	ACACACAC	YCAII
Trinucleotide	$(ATG)_n$	ATGATGATGATG	DYS481
Tetranucleotide	$(AATG)_n$	AATGAATGAATGAATG	TH01
Compound	$(TCTA)_m – (TCTG)_n$	TCTATCTATCTGTCTG	vWA
Complex	$(TCTA)_k – (AGTC)_m – (CCGA)_n$	TCTAAGTCAGTCCCGACCGA	D21S11

Finally, not all alleles for an STR locus contain simple repeat units, sometimes they can contain non-consensus alleles that fall in between alleles with full repeats units that are called *microvariants* [51].

1.2.2.1.2 Mutation rate

Mutation rate is defined as the number of mutations that can occur in a single generation. The mutation rates of human autosomal STRs are high, and oscillate between 10^{-2} and 10^{-6} per locus and per generation [52,53]. The major mechanism that causes microsatellite allele mutation is replication slippage or slipped-strand mispairing (SSM), although some authors also mention other mechanisms such as recombination or unequal crossing over during meiosis [53–58].

1.2.2.1.3 Nomenclature

In order to make possible interlaboratory reproducibility and comparisons in the forensic community a common nomenclature was developed. The nomenclature has been established by the DNA Commission of the International Society of Forensic Genetics (ISFG, https://www.isfg.org/), formerly known as the International Society of Forensic Haemogenetics, and has been updated when necessary [59,59–61]. Consequently, a repeat sequence is usually named

by the base composition of the core repeat unit and the number of repeat units in the 5' to 3' direction [14]. With simple repeats, the number of repeats units are counted. For compound repeats, alleles are designated by counting the total number of full repeats. With complex repeat systems, alleles can be identified according to their relative size compared to an allelic ladder containing sequenced alleles [62].

Nevertheless, currently most Forensic Genetics laboratories use commercially available kits that contain allelic ladders with alleles designated according to the ISFG rule and, therefore, do not have to worry about STR allele nomenclature [63].

1.2.2.1.4 STRs used in Forensic Genetics

For DNA typing to be useful in forensic casework across a wide number of jurisdictions and/or for international collaboration, it is necessary to analyze a common set of standardized markers. These common sets must be validated before their use in casework and DNA profiles are usually stored in national or international databases, which enable forensic laboratories to exchange and compare DNA profiles between individuals. DNA databases are indispensable tools available to law enforcement for fighting crimes, and their use allows to resolve forensic cases and human identification in natural or man-made disasters [64–66].

In the last years the forensic community has made significant efforts in the development of new multiplex analysis methods including the main STRs that compose the two main forensic databases, the combined DNA system (CODIS) managed by the FBI [67] and the European Standard Set (ESS) database [68]. Nowadays several commercial multiplexes are available which include the standardized STRs. As of 2015 [63], three commercial manufacturers (ThermoFisher, Promega and Qiagen) provide more than two dozen different STR kits that examine subsets of markers from a total of 29 autosomal STR loci, a sex-typing marker known as amelogenin, the Y-STR DYS391 and a Y chromosome deletion. Among then, the following kits are the most used: Identifiler®Plus, MiniFiler™, NGM SElect™, GlobalFiler™ (from Thermofisher Scientific, Wilmington, DE, USA), Power Plex® 16, Power Plex® ESI-17 (from Promega Corporation, Madison, WI, USA) and Investigator ESSplex SE QS kit (from Qiagen, Hilden, Germany).

1.2.2.1.5 Forensic DNA databases

A DNA database is a collection of computer files that contains entries of DNA profiles that can be searched to look for potential matches [14]. In Forensic Genetics, a DNA profile consist of a list of STR genotypes produced by the analysis of the defined core. These profiles usually come from forensic casework or a criminal offender who has been considered to legally qualify to enter de

database. The objective of forensic DNA databases is to aid law enforcement investigations by allowing effectively sharing genetic information of suspects in crime cases, and making associations between groups of unsolved cases. The larger these databases grow, the more effective they become, although some privacy and security concerns were raised in the last years.

Before the launch of the national DNA databases, not all the contained data was compatible between countries or laboratories, and the need for a compatible currency of data exchange motivated the selection of core STR loci. The first national database with forensic purposes was established in 1995 in the United Kingdom (NDNAD). Soon after that, other European countries started to create their own databases [64]. In 1989, the FBI in the United States proposed the creation of a national database for North America, which was termed the combined DNA index system (CODIS). Subsequently, the requirement to compare DNA results across different countries or jurisdictions motivated the establishment of a core set of STRs [69,70], which would first consist of 13 markers in the CODIS database, and of 7 STRs in the European database (known as the European Standard Set, ESS). As these databases grew in numbers, both core sets have been enlarged to avoid adventitious matches by adding 7 additional loci in 2015 (although it was not made effective until 2017), and 5 loci in 2009 respectively [14,68,71]. As of 2014, more than 100 countries have developed national databases for policing purposes [72].

In this context, in order to enable forensic scientists to keep up to date of the advancements in DNA typing, such as the discovery of new alleles or STR markers, the National Institute of Standards and Technology (NIST) launched in 1997 the Short Tandem Repeat DNA Internet Database, known as STRBase (http://www.cstl.nist.gov/biotech/strbase) [73].

1.2.2.1.6 Scientific working groups in Forensic Genetics

Scientific working groups (SWG) consist of scientific subject-matter experts who collaborate to both improve discipline practices and build consensus standards. In the Forensic Genetics community, several working groups and commissions have been created in the last years with the aim to promote accurate and reliable testing and to assist with quality assurance measures through the dissemination of scientific results and opinions, guidelines, education, orientation and communication amongst scientists. Table 2 shows the existing working groups worldwide as of November 2018.

Table 2. Major scientific association and working groups in the field of Forensic Genetics over the last years in the Unites Estates (EEUU), Europe (EUR), Latin America (LA), Asia and Oceania (OCEA).

Territory	Acronym	Name	Webpage
EEUU	AABB	American Association of Blood banks	http://www.aabb.org/
	ASCLD/LAB	American Society of Crime laboratory Directors/ Laboratory Accreditation Board	http://www.ascld.org/
	CAP	College of American Pathologists	http://www.cap.org/
	DAB	DNA Advisory Board	Not found
	NIST	National Institute of Standards and Technology	http://www.cstl.nist.gov/
	SWGDAM	Scientific Working Group on DNA Analysis Methods	http://www.swgdam.org/
EUR	DGAB	Deutschsprachige Arbeitsgruppe der ISFG	http://dgab-online.de/
	EDNAP	European DNA Profiling Group	https://www.isfg.org/EDNAP
	ENFSI	European Network of Forensic Science Institute	http://enfsi.eu/
	GeFI	Genetisti Forensi Italiani	http://www.gefi-isfg.org/
	GHEP-ISFG	Grupo de Habla Española y Portuguesa de la ISFG	https://ghep-isfg.org/
	IEWPDP	Interpol European Working Party on DNA Profiling	https://www.interpol.int/INTERPOL-expertise/Forensics/DNA
	ISFG	International Society of Forensic Genetics	https://www.isfg.org/
	STADNAP	Standardization of DNA Profiling Techniques in the EU	http://www.stadnap.uni-mainz.de/
LA	GITAD	Grupo Iberoamericano de Trabajo en Análisis de DNA	http://gitad.ugr.es/
	AICEF	Academia Iberoamericana de Criminalística y Estudios Forenses	http://www.aicef.net/
ASIA	AFSN	Asian Forensic Sciences Network	http://www.asianforensic.net/
	JSDPR	Japanese Society for DNA Polymorphism Research	http://dnapol.umin.jp/
OCEA	SMANZFL	Senior managers of Australian and New Zealand Forensic Laboratories	http://www.nifs.com.au/SMANZFL/SMANZFL.html

1.2.2.2 Single Nucleotide Polymorphisms (SNPs)

SNPs are base pair (bp) variations at specific locations in the genome and they represent the most abundant class of human polymorphisms [15,74] (Figure 4). The public dbSNP catalogue version 151 contains more than 113 million of validated SNPs

(https://www.ncbi.nlm.nih.gov/projects/SNP/snp_summary.cgi) in the human genome. A genetic variation at a single locus is not considered to be a SNP unless at least two alleles have a frequency over 1% in a large population of unrelated individuals [15]. The genomic distribution of SNPs is heterogenous, occurring more frequently in non-coding regions of the genome.

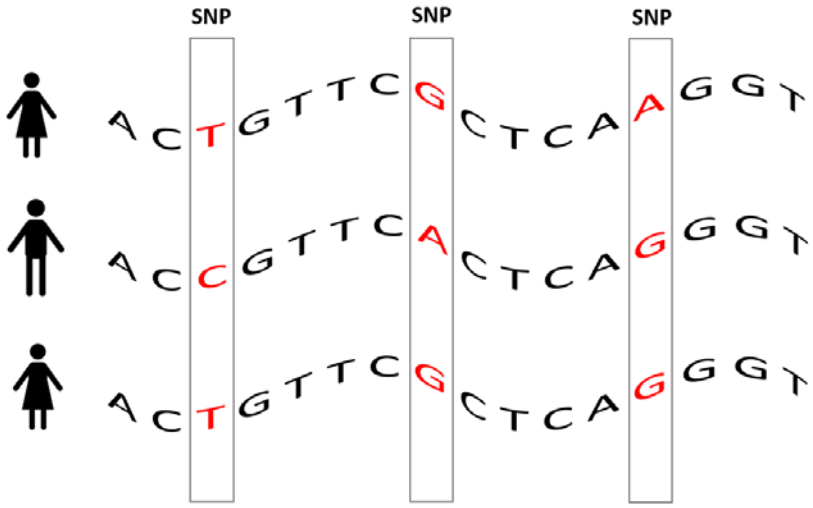


Figure 4. Diagram of a single nucleotide polymorphism (SNP).

The vast majority of SNPs are biallelic due to the low mutation rate at a particular bp position, and it is highly unlikely that two-point mutations happen at the same position over the time. For that reason, the analysis of SNPs has allowed to reconstruct the history of populations by studying the distribution of SNP alleles among present and past populations. In addition to that, SNPs can also be used to identify individuals, which makes them of high interest in Forensic Genetics, and some of them are also related to particular traits or diseases, which remarks their utility in medicine [15,74]. In the past decade genome-wide association studies (GWAS) have discovered several novel SNPs, with applications in different scientific fields [75]. Additionally, a large number of studies have contributed population data associated to some of these SNPs [76], and the forensic community has taken advantage of these resources to apply SNPs in forensic DNA analysis [74,77].

1.2.2.2.1 Advantages and disadvantages of SNPs versus STRs

STR are the preferred markers for forensic investigations due their high variability, the standardized commercial kits available, and the fact that they build up forensic DNA databases [14,15]. However, SNPs possess some qualities that makes them highly useful markers for forensic applications.

First, the amplification products from SNPs can be less than 100 bp in size, since the length of the product needs only to be the length of the PCR primers plus 1 bp. PCR primers are usually 15-18

long, which means that the PCR amplicon containing a SNP may be shorter than 40 bp. In contrast, the PCR products containing a STR locus are generally longer than 200 bp in order to allow their separation and identification by capillary electrophoresis. Thus, SNPs are able to recover more information from degraded DNA samples than STRs. Consequently, partial genetic profiles are obtained by the analysis of STRs when the DNA is degraded, whereas SNP typing of the same samples ends up in complete profiles[15].

Second, SNPs can be multiplexed to a higher level than STRs due to the fact that some detection methods, such as array hybridization, are not constrained by electrophoretic space [14]. Furthermore, the data interpretation derived from SNP typing results is simpler, as PCR artifacts known as 'stutters' are not formed in the amplification of SNPs, in contrast with the STR loci amplification. In addition to that, the sample processing and data analysis can also be more fully automated because there is no need of a size-based separation [14,15,74].

Third, SNPs are more stable than STRs. Mutations are a huge problem in kinship testing, and the mutation rates of the commonly used STRs are high, as detailed in previous sections, whereas the mutation rates for SNPs are estimated to be between 1.1-1.3 x 10^{-8} bp [78,79]. Consequently, the use of STRs in kinship testing can lead to genetic inconsistencies between child and a parent, since a few genetic inconsistencies are to be expected due to mutation events. By comparison, mutations will be extremely rare if SNPs are studied [15,80] and, for that reason, these markers pose as a useful tool for kinship testing in combination with STRs.

Fourth, unlike STRs, the analysis of SNPs allows a moderate power to predict ancestry, bio-geographical origin and certain physical traits, such as eye or hair color [14,77,81]. The forensic community has grown more and more interested in the SNPs related to these characteristics in the last years given their potential applications.

Although SNPs are full of advantages, they also possess significant limitations that challenge their use in forensic routine. The most critical disadvantage is their discrimination power. SNPs are less informative than STRs because most SNP loci are biallelic, meaning they have only two possible alleles, while the STR loci typically used in Forensic Genetics have 8-15 different alleles. The match probability for *n* SNP loci (*P*), assuming that all SNPs are in Hardy-Weinberg equilibrium, can be described as:

$$P = (p^2)^n + \big(2p(1-p)\big)^n + ((1-p)^2)^n$$

where *p* is the frequency of the least common allele and is constant for all loci. The highest match probability is obtained for *p* = 0.5. If *p* is between 0.2 and 0.5, 50 SNPs give a combined match

probability equivalent to that of 12 STRs [60]. Thus, in order to obtain a discrimination capacity similar to those of STRs a higher number of SNPs should be analyzed, around four times the number of STRs on average [60,82], though the number of SNPs needed may fluctuate in practice due to the variable allele frequencies of these markers in different populations.

SNPs possess better multiplexing capability than STRs, but the ability to simultaneously amplify enough SNPs in robust multiplexes to yield the discrimination power of the current STR panels from small amounts of DNA is still a challenge. The novel panels designed for MSP in the last years, which include large batteries of SNPs and allow analyzing hundreds of markers in multiple samples simultaneously [83–85], seem to have been able to overcome this limitation. However, degraded DNA samples still pose a challenge in these platforms [83,84].

On the other hand, SNPs are not really informative for samples that contain DNA mixtures, something that happens quite usually in trace samples from casework [74]. If both alleles of a SNP are detected in a mixture, there is no discrimination power and it is impossible to establish which individual contributed to the mixture. With STRs, since they are multi-allelic, it is possible to estimate the number of contributors in a sample from the number of detected alleles and, sometimes, to even determine the major and minor contributors based on the amplification strengths of the alleles [14,15].

Finally, forensic DNA databases are composed by STR profiles. To add SNP information to forensic databases would require analyzing the SNPs of all the included samples, which would be impossible in some cases due to the samples having been consumed or discarded, and the significant cost involved [86]. For that reason, STRs will probably continue to be the markers of choice in forensic casework while SNPs will be used as a supplementary tool, as they are useful for some specific applications [74,87]. Table 3 compares SNPs and STR markers.

Table 3. Comparison of SNP and STR markers. Adapted from [14].

Characteristics	SNPs	STRs
Genome abundance	Very high ≈ 1 in every kb	Less abundant ≈ 1 in every 15 kb
Variation type	Sequence variation	Length variation (repetition)
Number of alleles	Typically 2	Usually 5-20
Presence in nuclear DNA	Yes	Yes
Presence in mitochondrial DNA	Yes	No
Mutation rate	Low, nuclear 1.1-1.3 x 10^{-8}/ mitochondrial 2.7 x 10^{-5}	High, between 10^{-2}-10^{-6}
Stability across generations	High	Low
Amplicon size	Small, ≈ 60bp	Longer, ≈ 100-400 bp

Characteristics	SNPs	STRs
Artifacts	No	Stutters
Discrimination capacity	Low, 20-30% as informative as STRs	Very high
Existence of national databases	No	Yes
Mixture analysis	Limited	Very high
Application on identification	Requires a high number of markers	Requires a low number of markers
Application on degraded DNA	Yes	Limited, allelic loss or absence of profile
Application on complex kinships	Moderate	Low
Lineage information	Yes	No, very limited in Y chromosome STRs
Phenotypic information	Yes	No
Ancestry information	Yes	No
MPS analysis	Yes	Yes
Multiplexing capability	> 10 markers with multiple fluorescent dyes	<50 markers with conventional techniques, 100-1000 markers with MPS kits or microchips

1.2.2.2.2 SNP categories and applications

SNP markers were categorized at the 2007 congress of the ISFG [81]. SNPs of forensic interest possess specialized applications and, for that reason, four categories were defined: identity-testing SNPs (IISNPs), lineage informative SNPs (LISNPs), ancestry informative SNPs (AISNPs) and phenotype informative SNPs (PISNPs) [77,81].

1.2.2.2.2.1 IISNPs

Identity-testing SNPs provide information to differentiate between individuals and exclude those that cannot be the resource of an evidentiary sample, or cannot be a potential family member [77]. They can serve as supplemental markers when STR results are not definitive enough in cases like complex kinship tests [14], or when the DNA is highly degraded or in low template amounts and STRs do not provide complete genetic profiles or any profile at all [88]. For a SNP to be selected as a forensic identity marker it must possess high heterozygosity and low F_{ST}, as measured by Wright's F_{ST} [89].

Thus far, several panels of IISNPs have been developed [88]. The most used in forensic practice are the panels based in single base extension (SBE) or minisequencing technology [90]. Among them, the panels SNP*for*ID 52-plex [91] and Wang's 55-plex [90] are the ones with the highest number of SNPs

included. In the las 3 years three MPS forensic panels have been released which include IISNPs among other markers [83,84,92].

1.2.2.2.2.2 LISNPs

Lineage informative SNPs are sets of closely linked markers, with very little possibility of recombination between generations, so that each set behaves as a single locus [14,77]. Although each of the SNPs that make up a lineage marker is usually bi-allelic, the combination of the SNPs is equivalent to a single locus with many variants. Each of the combination of the variants is called *haplotype*. The most used LISNPs are the ones included in the Y chromosome (Y-SNPs) and the mitochondrial DNA (mtSNPs) [14,77,88]. All of these markers are transmitted uniparentally and, in the case mtSNPs and Y-SNPs, they have demonstrated informative geographical differentiation.

These markers can be useful in Forensic Genetics for defining both paternal and maternal lineages, tracing human migrations, and kinship analyses [77,93]. Y-SNPs and mtDNA can also aid in biogeographical ancestry prediction, although due to their inheritance caution should be taken when making ancestry predictions, since LISNPs provide lineage information about one of the parents and predictions on admixed individuals are often problematic and not reliable [94].

The major applications of these markers in casework have been for mass disaster victim identification, missing person cases, and to give clues about an unknown contributor to a crime scene when other markers, such as STRs, have given no results [74,95]. As expected, several multiplexes that contain LISNPs have been designed based on SBE technology [88] and some MPS kits also contain these types of markers as well.

1.2.2.2.2.3 AISNPs

Ancestry informative markers (AIMs) are distributed throughout the genome and occur at very different frequencies in the distinct populations of the world [96]. Thus, they are able to predict an individual's ancestral background. These markers can reveal the ancestral origin of a sample, but do not provide information about physical traits, as it is an indirect method of assessing phenotype [77]. The ideal characteristics of AISNPs are opposite to those of IISNPs, they should display low heterozygosity and high F_{ST} between populations [77,88].

In forensic casework, knowing the biogeographic origin of an individual can provide limited clues about the general appearance of a person and facilitate the crime investigation. STRs perform rather poorly for defining biogeographical ancestry and ethnicity due to the high degree of allele sharing between populations. For that reason, SNPs are better predictors of ethnicity, although

some caution should be taken since AIMS are not 100% accurate for predicting ancestral background [14].

Unlike LISNPs, AIMs are inherited jointly from both parents and owing to that, are able to predict ancestry in admixed individuals, who may not possess the expected phenotypic characteristics (like skin color). They can also be employed to efficiently estimate admixture proportions and to detect recent admixture in human populations [97].

One limitation of this type of SNPs is the need to define a number of candidate markers able to measure and assess differences in continental population structure at individual level [98]. Thus, a specific set of AISNPs should be used depending on the population. In this context, several multiplex tools have been released centered on global or population specific ancestry [97–100].

1.2.2.2.2.4 PISNPs

Phenotype informative SNPs, unlike AISNPs, can provide a direct and accurate genetic prediction of phenotypic traits and lead to the identification of the perpetrator of a crime due to the clues provided about their appearance [77,81].

The most obvious phenotype descriptors, or externally visible characteristics (EVCs), are pigmentation (skin, hair and eyes), hair morphology, body height, and facial features, which are all of them highly heritable [101–104]. The extraction of information on phenotype through the molecular analysis from biological crime scene samples is called Forensic DNA phenotyping (FDP) [15], which is important when no matching individuals can be found from an evidentiary DNA profile.

Most work on PISNPs has concentrated on pigmentation thus far, particularly skin, eye and hair color. Several SNPs related to those physical traits have been included in the forensic panels IrisPlex, HIrisplex, and 8-plex [104–106] which are able to predict eye color, eye/hair color and skin/eye color respectively. In the future, perhaps we will be able to reconstruct other highly heritable traits such as facial morphology or height, although sensitive considerations regarding privacy and ethics should be taken when performing these predictions.

1.2.2.2.3 SNP databases

SNP databases contain genomic information, allelic frequencies, and distribution of several SNPs in different human populations. Although most of these databases are not of forensic use, since they do not contain genotyping information related to casework, and the data is mostly derived from population studies, they can provide valuable information of SNPs related to forensic applications. Some of the most important databases providing SNP information are:

- *dbSNP* (https://www.ncbi.nlm.nih.gov/projects/SNP/index.html), which is a free public archive for genetic variation across different species developed by the National Center for Biotechnology Information (NCBI). This database assigns a SNP ID or "*rs"* number to each variation in order to identify it unambiguously.
- *Ensembl* (http://www.ensembl.org), is a genome database managed by the European Bioinformatics Institute and the Wellcome Trust Institute that provides a centralized resource of genomic information from various species.
- *SNPforID* browser (http://spsmart.cesga.es/snpforid.php), is an online repository of the data generated by the SNPforID [91,107].
- *UCSC Genome Browser* (http://genome.ucsc.edu/), is an on-line genome browser that offers access to genome sequence data from a variety of organisms.

1.3 Study of paternal lineages: The Y chromosome

The Y chromosome has a special primary role in humans, it determines male sex [108] and also has a critical role in spermatogenesis [109] and male fertility. Due to its distinct characteristics, namely male specificity, haploidy and absence of crossing over [110], the Y chromosome possess a particular interest in forensic, medical and population genetics [111].

The sex chromosomes began to diverge from a pair of homologous autosomes around 180 million years ago [112]. From then on, their evolutionary history has been strikingly different, as the X chromosome retained most of its original content and is highly conserved [113], while the Y chromosome has degenerated, losing lost most of its ancestral content [113].

1.3.1 The structure of the Y chromosome

The structure of the Y chromosome is highly complex, rich in segmental duplications and repeats that makes it almost impossible to assemble reliably using short-read sequencing technologies [114,115]. The presence of a Y chromosome usually leads to a male phenotype due to the expression of the gene SRY (sex-determining region Y) [116,117].

The Y chromosome is small, comprising about 60 megabases (Mb) and containing only a few genes [118] (Figure 5). At the tips of the short and long chromosomal arms, there are two segments of sequence homology, the pseudoautosomal regions 1 and 2 (PAR1 and PAR2), in which meiotic crossing over between the X and Y chromosomes occurs [119–124]. PAR1 is positioned in the short arm and is approximately 2.5 Mb in length [122,123], while PAR2 is in the long arm and is less than 1 Mb in size [121,123]. The remainder of the Y chromosome, which is about 95% its total length, is called

the male-specific region of the Y chromosome (MSY), which is located between the PAR regions and does not recombine [14,110,118]. Thus, it remains the same from father to son unless a mutation occurs. This region is also known as the non-recombining region of the Y chromosome (NRY), although due to the fact that abundant gene conversion and intrachromosomal recombination exists in this region [125], it is more correct to refer to it as MSY [118].

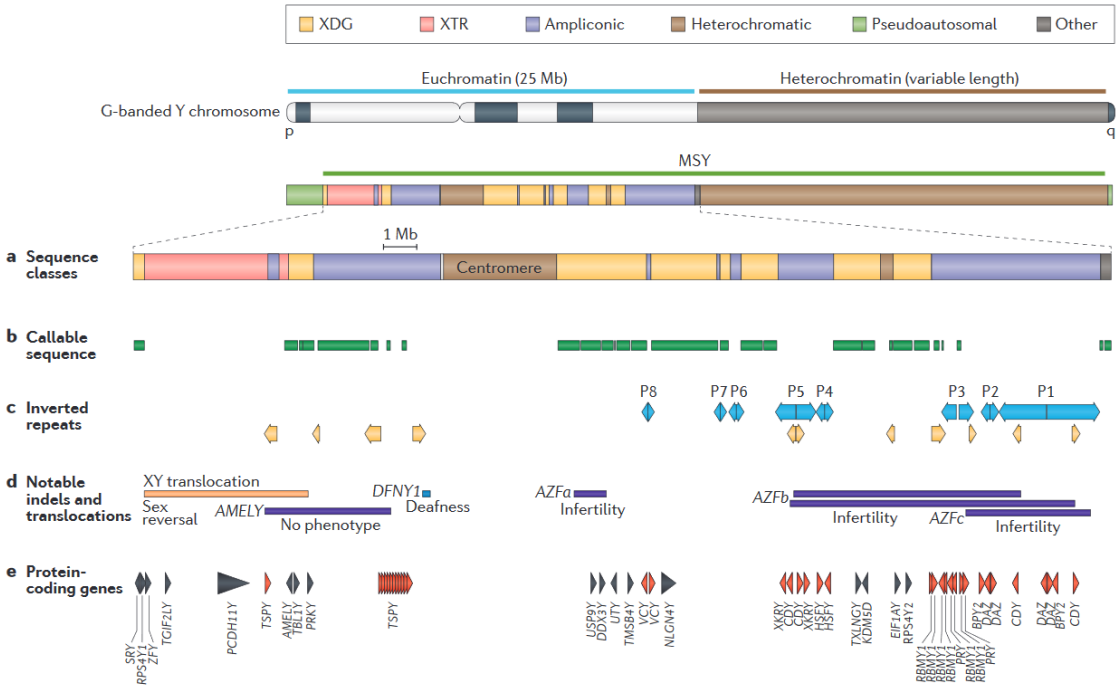


Figure 5. Detailed structure of the Y chromosome. a) Sequence classes. XDG: X-degenerate, XTR: X-transposed. b) Callable sequence. c) Inverted repeat sequences. d) Notable INDELs and translocations. e) Protein-coding genes. Extracted from [110].

The MSY region is composed of approximately one half of variable sized heterochromatin, while the remaining 23 Mb are composed of euchromatin [118]. The euchromatin is composed of three major sequence classes [118]:

- X-degenerate class (XDG) (Figure 5, a), which occurs in eight blocks of the Y chromosome with a length of 8.6 Mb. These sequences possess up to 96% nucleotide sequence identity to their X-linked homologues [126,127].
- X-transposed region (XTR) (Figure 5, a), which is a 3.4 Mb block of DNA that has been transferred from the X chromosome since the lineages of humans and chimpanzees diverged from each other [128]. These sequences are 99% identical to the sequences found in the X chromosome and do not participate in the crossing over during male meiosis.
- Ampliconic regions (Figure 5, c), which are intrachromosomal repeats of high sequence similarity and have a total length of 10.2 Mb [118]. There seven large blocks and the majority

of these sequences (60%) have intrachromosomal identities of 99.9% or greater, which makes it very difficult to tell one apart from another. Among the repeated sequences are large direct repeats and inverted repeats, as well as eight palindromes that collectively comprise 5.7 Mb.

The high interchromosomal and intrachromosomal similarity of the X-transposed region and the ampliconic regions makes analyzing some genomic areas and interpreting resequencing data very difficult [110], and only in 9.99 Mb of the total chromosome sequence are variants unambiguously callable [129] (Figure 5, b). Although the MSY region is not affected by crossing over during male meiosis, some level of recombination has been reported. Gene conversion occurs quite usually in ampliconic regions through non-allelic homologous recombination [110,125,130–134], and sometimes between non-pseudoautosomal sequences on the X and Y chromosomes that are very similar [132,135–137] (Figure 5, d).

During the last decade, the improved sequencing technologies have enriched our knowledge of the structure of the Y chromosome and the variation within the MSY. It is clear that this chromosome, previously neglected by geneticists and greatly ignored by GWAS [110,111], will provide critical information in forensic and medical genetics thanks to long-read sequencing technologies, which will probably enable us to access the entire sequence of the chromosome and reveal the information hidden in the repeated regions of the MSY [110].

1.3.2 Y chromosome markers

The Y chromosome, in opposite to autosomal markers, is a lineage marker and is passed down during generations without changing, constituting paternal lineages. In order to examine Y chromosome diversity two broad categories of DNA markers have been used: bi-allelic markers, such as Y chromosome SNPs (Y-SNPs) [138], and multi-allelic markers, like Y chromosome STRs (Y-STRs) [110,139,140]. The use of the Y chromosome in Forensic Genetics is limited due to its haploidy and inheritance, which makes it less effective for identification purposes in contrast to autosomal markers. It cannot differentiate between individuals from the same paternal lineage, meaning a father, his brother and his son will possess the same Y chromosome. However, its utility cannot be disregarded since a vast majority of crimes where DNA evidence is helpful involve male individuals as perpetrators [14].

1.3.2.1 Y-STRs

Currently, the most used Y chromosome markers in forensic casework are Y-STRs due to their higher mutation rate (10^{-3}) in comparison with Y-SNPs (10^{-9}) [14]. Y-STRs are used particularly in cases

where standard autosomal DNA profiling is not informative, crimes involving male perpetrators (especially sexual assault), characterization of paternal lineages of unknown male trace donors, and parentage/kinship analysis [95]. Y-STRs are described as defining *haplotypes*, combination of different loci alleles on a chromosome that are inherited or transmitted together and do not undergo recombination [15]. Thus, Y-STR haplotyping can exclude male suspects from involvement on a crime, identify the paternal lineage of a male perpetrator, highlight male contributions to a sample, and provide investigative leads to find unknown male perpetrators [95].

1.3.2.1.1 Types of Y-STRs

Some of the conventional Y-STR loci occur more than once due to the duplicated and palindromic structure of some regions of the Y chromosome and, therefore, produce more than one PCR product when amplified. Those are DYS385ab [141], DYS389 [142], DYF387S1a/b and DYS464 a/b/c/d [142–144]. Y-STR marker selection for forensic application is usually performed taking into account two main criteria, diversity measures in the populations [145–148] and/or mutation rate [149–151].

In forensic practice Y-STRs with low, medium, and high mutation rate use analyzed. Y-STRs with a higher mutation rate, or rapidly mutating Y-STRs (RM Y-STRs), possess a mutation rate of few mutations per 100 generations per each locus, and are more suitable for identification than those with low or medium mutation rate (few mutations per 1000 generations per locus) [95,148,150]. For that reason, they have captured the interest of the forensic community recently. Some RM Y-STRs (DYS570, DYS576, DYS449, DYS518, DYS627, and DYF387S1a/b) have already been included in some commercial kits, such as Yfiler™ Plus (Qiagen, Hilden, Germany) and PowerPlex® Y23 (Promega Corporation, Madison, WI, USA).

On the other hand, the conventional Y-STRs (with lower mutation rates) are better suited for parentage studies and for familiar searching than RM Y-STRs, as the occurrence of mutations with increased probabilities will trouble the estimation of paternity/kinship [95].

1.3.2.1.2 Minimal haplotype

A paternal lineage can be more accurately characterized by Y-STR haplotyping if more Y-STR markers are considered. Although in the last years the number of Y-STRs available for human and lineage identification has increased dramatically, in 1990 only a few loci were available, and to serve as a reference for forensic scientists in Y chromosome testing a core set of 9 Y-STRs was selected in 1997 called the “minimal haplotype” [147,152,153]. The minimal haplotype is defined by the markers DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393 and DYS385a/b. In 2003, two more markers were recommended to use by the SWGDAM: DYS438 and DYS439 [154].

1.3.2.1.3 Y-STR typing kits

As previously mentioned for STRs, forensic scientists rely heavily on available commercial multiplexes to perform DNA testing, and Y-STRs are no exception. Currently, the most widely used kits are Yfiler™ (Thermofisher Scientific, Wilmington, DE, USA), Yfiler™ Plus (Thermofisher Scientific, Wilmington, DE, USA), and Powerplex® Y23 (Promega Corporation, Madison, WI, USA). Other non-commercial multiplexes are also available which include RM Y-STRs, other markers of interest, or that serve to complete with more markers previous multiplexes [150,155–158] (Table 4).

Table 4. Most commonly used Y-STR markers in forensic DNA analysis. PPY23: PowerPlex® Y23; RM Y-STR: RM Y-STR panel [150]; 6-plex: 17 to 23 panel [155].

Y-STR marker	Minimal haplotype	Yfiler	Yfiler Plus	PPY23	RM Y-STR	6-plex
DYS19	X	X	X	X		
DYS385a/b	X	X	X	X		
DYS389I	X	X	X	X		
DYS389II	X	X	X	X		
DYS390	X	X	X	X		
DYS391	X	X	X	X		
DYS392	X	X	X	X		
DYS393	X	X	X	X		
DYS437		X	X	X		
DYS438	X	X	X	X		
DYS439	X	X	X	X		
DYS448		X	X	X		
DYS456		X	X	X		
DYS458		X	X	X		
DYS635		X	X	X		
Y-GATA-H4		X	X	X		
DYS481			X	X		X
DYS533			X	X		X
DYS549				X		X
DYS570			X	X	X	X
DYS576			X	X	X	X
DYS643				X		X
DYS449			X		X	
DYS460			X			
DYS518			X		X	
DYS627			X		X	
DYF387S1a/b			X		X	
DYS526a/b					X	
DYS547					X	
DYS612					X	
DYS626					X	
DYF399S1					X	
DYF403S1a/b					X	
DYF404S1					X	

1.3.2.2 Y-SNPs

Y-SNPs are distributed through the Y chromosome and they serve as markers of paternal lineage. They can be used to establish evolutionary lineages, broad biogeographical ancestry, and to trace

male-mediated demographic events [95,159,160]. The applications and limitations described in section *1.2.2.2* also apply in this case, in addition to limitations derived from the Y chromosome structure. The use of these markers is not extended in forensic casework, although they can provide useful clues when the DNA typing of other markers fail, or there is an unknown perpetrator. Y-SNPs are most used in population genetics and genetic genealogy, and they can also aid in paternity/kinship testing [159].

Since their mutation rate is lower in comparison with Y-STRs (10^{-9}), biogeographical ancestry signatures are kept much longer in Y-SNPs before it is diluted due to mutations. For that reason, Y-SNPs are more suitable for paternal biogeographical inference than Y-STRs. Furthermore, the Y chromosome escapes recombination and, once a mutation occurs, it is not removed from the gene pool unless no male offspring exists. The uniparentally inherited part of the genome, both Y chromosome and mtDNA, are more susceptible to genetic drift and, therefore, genetic differences can occur between geographic regions. On the other hand, some elements of human culture, like patrilocal residence, polygyny, and male-mediated migrations have also made the Y chromosome more suitable for ancestry analysis [95].

1.3.2.2.1 Y chromosome haplogroups and global distribution of paternal lineages

Y chromosome *haplogroups* are related sets of Y chromosomes that are collectively defined by specific Y-SNPs [110]. Due to the evolution and migration of human groups across the globe, haplogroups display a strong geographical differentiation at continental level, and some even at regional level [140,161]. Table 5 provides a broad picture of the distribution of paternal lineages worldwide, which can also be visualized in Figure 6.

Table 5. List of Y-SNPs defining the main Y chromosome haplogroups and their geographical distribution. NA: Not available. Adapted from [95].

Haplogroup	Defining Y-SNP	Rs number	Geographic distribution
A00-L1086	L1086, L1159, L1284	NA, NA, NA	Central Africa
A0-V148	V148, V166, L896, L991	rs181335666, rs187287389, NA, NA	Central Africa, West Africa
A1-M31	M31, P82, V4	rs369315948, NA, rs187409543	West Africa, North Africa
A2-V50	V50, L602	rs189205028, rs576471146	Southern Africa, Central Africa
A3-M32	M32	rs558241924	East Africa, Southern Africa
B-M60	M60, M181, V244	rs2032623, rs2032599, rs112298449	Central Africa, Southern Africa, East Africa
D-M174	M174, CTS94, JST021355	rs2032602, rs199881488, rs2267802	East Asia

Haplogroup	Defining Y-SNP	Rs number	Geographic distribution
E-M96	M96, M40, P29	rs9306841, rs9786608, rs60115999	Africa, West Asia, Southern Europe
E-V13	V13, V36	rs368031074, rs371443469	Southern Europe
E-M293	M293	rs9341316	Southern Africa
E-M75	M75	NA	Central and South Africa
E-M35	M35, L336	rs375228668, rs112779735	Africa, Ashkenazi and Sephardic Jews
C-M130	M130, M216	rs35284970, rs2032666	Central Asia, Northern Asia, North America, East Asia, Southeast Asia, Wallacea, Near Oceania, Remote Oceania, Australia
C-M8	M8, M105	rs3899, rs2032612	Japan
C-V20	V20	rs182352067	Southern Europe
C-M356	M356	NA	South Asia, Central Asia
C-B65	B65	rs374541802	Indonesia, Philippines
C-M38	M38	rs369611932	Wallacea, Near Oceania, Remote Oceania
C-M208	M208	rs2032659	Near Oceania, Remote Oceania
C-P33	P33	NA	Remote Oceania
C-PH41	PH41, PH338	NA, NA	Australia
C-M217	M217, P44, Z1453	rs2032668, NA, NA	Central Asia, Northeast Asia, South Americans (Ecuadorian Native Americans)
C-P39	P39	NA	Northern America
F-M89	M89, M213, P14	rs2032652, rs2032665, rs9786420	South Asia
G-M201	M201, P257	rs2032636, rs2740980	West Asia, Europe, Central Asia
G-P15	P15, U5, L31	rs370167410	Mediterranean Europe, Northern Europe
H-L901	L901, M3035	rs567848586, rs74378870	South Asia
I-M170	M170, M258, U179	rs2032597, rs9341301, rs2319818	Europe, West Asia
I-M253	M253, L80	rs9341296, rs35960273	Northern europe
I-M438	M438, P215	rs17307294	South-eastern Europe
J-M304	M304, P209	rs13447352, rs17315835	West Asia, North Africa, Horn of Africa, Southern Europe, Central Asia, South Asia
J-M267	M267, M497	rs9341313, rs371666197	Middle East, Caucasus, North-East and North Africa
J-M172	M172, L228	rs2032604, rs371968167	Middle east, Caucasus, Mediterranean area
L-M20	M20	rs3911	South Asia, West Asia
M-P397	P397, P399, PR2099	NA, NA, rs369017623	Wallacea, Near Oceania, Remote Oceania, Australia
M-P34	P34	NA	Near Oceania
M-M10072	M10072, FGC38729, Z33118	rs566812523, NA, rs368850080	Australia
N-M231	M231	rs9341278	Northern Asia, Northern Europe
O-M175	M175, P186	rs2032678	East Asia, Southeast Asia, Remote Oceania
Q-M242	M242	rs8179021	Northern Asia, Central Asia, Americas
Q-M3	M3	rs3894	Americas
Q-Z780	Z780	NA	Americas
R-M207	M207	rs2032658	Europe, West Asia, Central Asia, South Asia, North Africa, Central Africa
R-M420	M420	rs17250535	Eastern Europe
R-M458	M458	rs375323198	Eastern Europe, Caucasus region
R-Z284	Z284	rs767265794	Northwest Europe
R-Z93	Z93	rs566323605	South Asia, Central Asia

Haplogroup	Defining Y-SNP	Rs number	Geographic distribution
R-M343	M343, M415	rs9786184, NA	Western Europe
R-V88	V88	rs180946844	Africa, Southern Europe
R-M269	M269, M520	rs9786153, NA	Western Europe
R-Z2103	Z2103, Z2105	rs567703217, rs544980517	Eastern Europe, West Asia
R-M412	M412	rs9786140	Western Europe
R-L11	L11, S127	rs9786076	Western Europe
R-S116	S116, P312	rs34276300	South-western Europe
R-U106	U106, S21	rs16981293	North Central Europe
R-M529	M529, S145	rs11799226	British Isles
R-U152	U152, S28	rs1236440	Alps, France, Western Poland
R-M479	M479	rs372157627	South Asia, Central Asia
S-M254	M254	rs9341297	Near Oceania
T-M184	M184	rs20320	West Asia, Horn of Africa, North Africa, Southern Europe, South Asia

The relationship between Y-SNPs and geographical regions has been established thanks to population studies, which have generated valuable population data for many geographical regions. Some areas, however, have been less studied and the biogeographical information provided by Y-SNPs is limited, especially in the case of recently discovered markers whose distribution is not yet known [95].

In some cases, Y-SNPs with strong geographical differentiation can display strong or moderate correlation with an associated Y-STR based haplotype. In these instances, a Y-SNP haplogroup could be inferred from associated Y-STR haplotypes. This is true for the broadly studied European haplogroups R1a (M420) and R1b (M343) (Figure 6, Table 5). However, many other haplogroups are not that studied, and not many Y-STRs with strong geographic signatures are known [95]. In this context, many online haplogroup prediction tools are available, such as Whit Athey's haplogroup predictor (http://www.hprg.com/hapest5/) and Nevgen haplogroup predictor (http://www.nevgen.org/), whose accuracy depend mainly on the number of Y-STR haplotypes from which allele frequencies can be calculated [162,163]. Haplogroup prediction can be an easy and quick tool when Y-SNP information is not available, but the percentage of error present on the estimation, especially at subhaplogroup level, and in those lineages less studied, is a huge limitation that cannot be disregarded. For that reason, haplogroup prediction from Y-STR haplotypes is not an accurate enough method and Y-SNP typing is necessary to define the specific paternal lineage of a sample unambiguously [163].

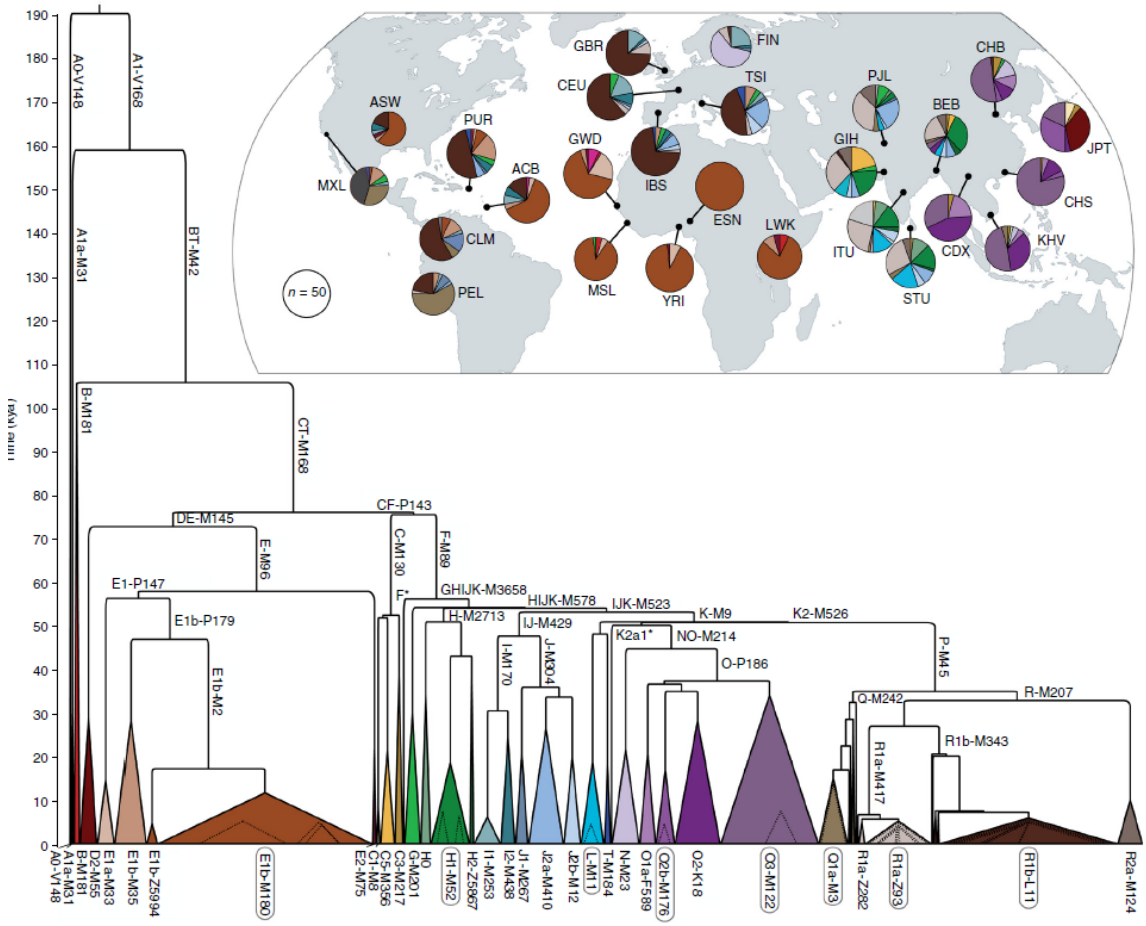


Figure 6. Y chromosome phylogeny and global haplogroup distribution. The branch lengths are proportional to the estimated times between the successive splits, occurring the most ancient division around 190,000 years ago. The colored triangles represent the major clades, and the width of each base is proportional to one less than the corresponding sample size. Dotted triangles represent the ages and sample sizes of the expanding lineages. Inset, world map indicating, for each of the 26 populations, the geographic source, sample size, and haplogroup distribution. Samples correspond to 1,244 male individuals from five global superpopulations sequenced on the Phase 3 of the 1000 Genomes Project [164]. Figure extracted from [165].

1.3.2.2.2 Phylogeny

The appearance of modern humans goes back to a single common origin in Africa 300,000-200,000 years ago. During this long history, there have been enough generation steps to allow the appearance of mutations that have end up creating continental differences at various Y-SNPs. These Y-SNPs can be used to establish a robust phylogeny using the principle of maximum parsimony [110], as each one of them first originated at some branch of the genealogical or phylogenetic tree that unites all humans. Therefore, Y-SNPs display a hierarchical structure that is organized in main haplogroups or lineages (defined by one or more concrete Y-SNPs) that can be dissected in subhaplogroups or sublineages, which are located below in the phylogenetic tree and possess the Y-SNP that defines the sub-branch as well as the one that defines the main branch.

In the last decade, several studies using MPS technology have provided a large number of undiscovered Y-SNPs and have allowed to reconstruct and calibrate a reliable reference phylogeny for the human Y chromosome [165–169] (Figure 6). The first efforts to develop a nomenclature for haplogroups were performed by the Y Chromosome Consortium (YCC) in 2002 [170], given the strong confusion that arose the appearance of several nonsystematic nomenclatures for the haplogroups. In 2014, a minimal reference phylogeny for the human Y chromosome was presented, which represent a reduced version of the tree including only the principal branches together with the broad geographic distribution [171]. This phylogeny is available online through Phylotree-Y (http://www.phylotree.org/Y/tree/index.htm). The International Society of Genetic Genealogy (ISOGG) offers online a more detailed phylogeny that is updated regularly (https://isogg.org/tree/index.html).

Despite the great efforts made by the community to develop a consensus Y-SNP nomenclature there is still confusion regarding this issue, since several Y-SNPs possess more than one denomination and some authors use one of the denominations while others use another one. For that reason, in order to avoid ambiguous haplogroup assignation, is best to refer to Y-SNPs using the SNP ID number provided by the database dbSNP (if it is assigned), and/or the most common denomination provided by both phylogenies mentioned above.

1.3.3 Y-SNP typing technologies

A great variety of genotyping techniques are available for the analysis of SNPs [74], and the increasing interest on these markers during the last 20 years has resulted in the development of more and more novel typing platforms. These methods are based in four common technologies: hybridization, primer extension, ligation, or invasive cleavage [15,74]. The techniques that will be described below are the most commonly used in Forensic Genetics and are also applicable for the typing of SNPs located in other regions of the genome.

1.3.3.1 SNaPshot™ minisequencing

The minisequencing or single base extension (SBE) method is one of the most relevant and commonly applied technology for forensic DNA analysis due to its sensitivity, high multiplexing capability, and the advantage of not requiring additional equipment apart from the one already available in any forensics laboratory [88,90,172]. Furthermore, there is a high number of available panels of different types of SNPs of forensic interest [88].

Minisequencing is a genotyping method that is based on primer extension techniques [74], where a detection primer is designed to anneal to the target DNA immediately upstream of the SNP of

interest and is then extended by DNA polymerase using fluorescently labelled single nucleotides (ddNTPs) [74,88,172] (Figure 7). Thus, only the single base that is complementary to the SNP of interest is added to the primer. The resulting dye-labeled products are most commonly detected by capillary electrophoresis (CE) [172], although other techniques are also available like fluorescence detection, matrix assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry (MS), and microarrays, on which the multiplexing capability will depend [74]. This technique was developed in 1990 [173,174] and first validated for forensic casework few years later for the detection of mitochondrial sequence polymorphisms [175].

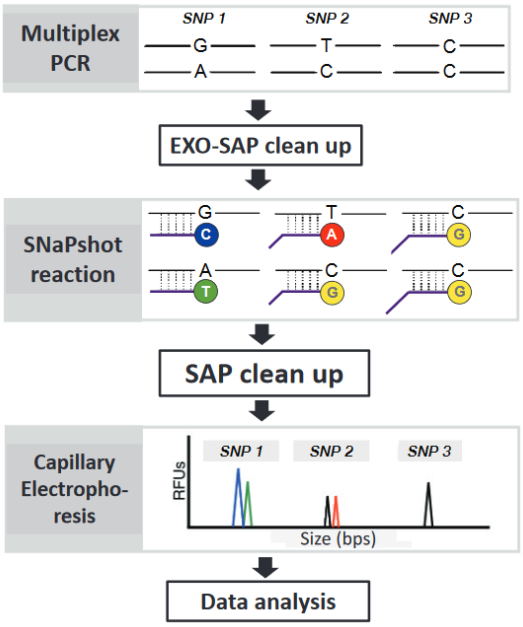


Figure 7. Outline of the SNaPshot™ process steps and depiction of the SBE reaction. Dye-linked terminating ddNTPS are shown as circles with their relevant colors. Extracted from [172].

The most common commercial minisequencing assay is the SNaPshot™ Multiplex kit by ThermoFisher (Thermofisher Scientific, Wilmington, DE, USA). Although there are not commercially available multiplex panels for the analysis of Y-SNPs, several studies have developed reliable multiplex tools that enable to genotype global haplogroups [176–179] or more regional lineages [178,180–182] in forensic and ancient DNA samples [183]. These Y-SNPs panels are highly useful, but some limitations exist regarding the haplogroup resolution they are able to achieve, as well as the number of multiplex reactions needed for their application, as some panels require more than one multiplex reaction [177,180,181]. Consequently, there are not minisequencing panels available that dissect each major branch of phylogenetic tree, as some haplogroups are more studied than others. Thus, there is a requirement to develop more multiplex tools that allow to dissect more specific haplogroups, achieving high resolution in a minimal number of reactions.

1.3.3.2 High Resolution Melting (HRM)

High Resolution Melting (HRM) is the quantitative analysis of the melt curve of DNA fragments following a real-time PCR (RT-PCR) amplification. The use of this technique in Forensic Genetics is not as widespread as SNaPshot™ genotyping, but recent studies have shown the potential of HRM for SNP genotyping in forensic samples [184–186]. HRM is a simple, cost effective, sensitive, closed tube SNP genotyping technique with high throughput potential [185].

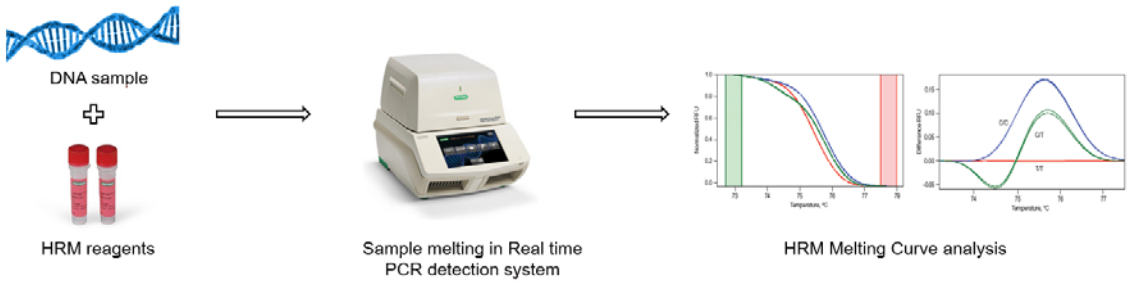


Figure 8. HRM workflow. Adapted from Bio-Rad (http://www.bio-rad.com).

This method needs a RT-PCR to be performed first, were the target SNP is amplified in a short fragment using a saturating concentration of double-stranded DNA (dsDNA) binding dye, like EvaGreen or LCGreen [187,188]. When the PCR products transitions from double stranded to single stranded conformation, a rapid loss of the fluorescent signal happens, where the fluorescence is proportional to the amount of remaining double stranded DNA. The temperature at which half of the DNA is in single stranded conformation is termed the "melting temperature" (T_m) [189–191]. The T_m and the morphology of the PCR product melting profile are then used to identify the presence of sequence variation (Figure 8). The melting profile, as well as the T_m, is dependent on the features of the sequences like GC content, GC distribution, length, and sequence of the PCR amplicon [189,192]. HRM is more sensitive for smaller fragments, as increasing the size of the PCR products may reduce T_m differences between alternative genotypes and increase the difficulty of data interpretation. For that reason, it is recommended to use fragments between 40-100 bp for SNP genotyping, or at least no longer than 300 bp [188,193–197].

The major advantages of HRM are the speed of the analysis, as the acquisition of the melting profiles occurs within the instrument immediately after the PCR, the elimination of post-PCR sample manipulation, which reduces the time of labor and the risk of contamination, and the low cost in comparison with other techniques [185,189,198]. Moreover, it can be applied to degraded DNA due to the short amplicons needed. The main limitation of HRM in Forensic Genetics is its reduced multiplexing capability, only multiplexes with up to 6 markers have been described [199,200]. Although

multiplexing is possible, it is more limited than in other genotyping technologies due to the fact that product detection occurs in a limited range of temperature (60-95 °C) [185,186,199].

1.3.3.3 TaqMan™ assays

The TaqMan™ assay is based in the 5' nuclease activity of the Taq polymerase, which displaces and cleaves the oligonucleotide probes hybridized to the DNA generating a fluorescent signal [201,202]. This method has been one of the most popular probe-based assays in forensics, especially for human DNA quantification and sex identification through RT-PCR [203].

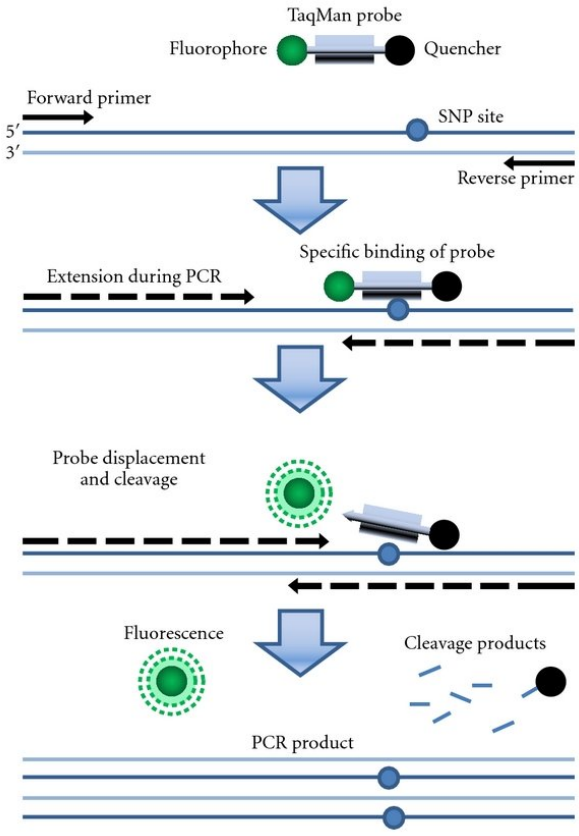


Figure 9. Representation of TaqMan™ assay genotyping. Extracted from [204].

Two primers with reporter fluorescent dyes attached to the 5' end and a quencher attached to the 3' end are required, one complementary to the ancestral allele of the variant and the other to the derived allele [202]. When the probes are not hybridized to the target DNA the quencher interacts with the fluorophore, quenching the fluorescence. During the PCR annealing step, the TaqMan™ probes hybridize to the target DNA, and in the extension step the Taq polymerase cleaves the 5' fluorescent dye by its 5'nuclease activity, which leads to an increase in the fluorescence of the reporter dye. The genotype of a sample is then determined by the measuring of the signal intensity of the two different dyes [74] (Figure 9).

Even if this method is specific and very sensitive for Y-SNP analysis, its multiplexing capability is very limited, as it also happens with HRM. It is only possible to analyze simultaneously up to 4 different SNPs. Considering that, currently the application of this technique in Forensic Genetics is mostly limited for DNA quantification and assessment, used in common quantification kits such as Quantifiler (Promega Corporation, Madison, WI, USA).

1.3.3.4 High density SNP arrays

High density SNP arrays allow hundreds of thousands or even millions of SNPs to be genotyped in parallel. Their main limitation is the high cost, the high rate of null results, and the requirement of high amounts of DNA, which is often not available from casework samples that come from minimal biological stains.

In this method, oligonucleotides are attached to a solid support to create a microarray and are then hybridized with fluorescent labelled PCR products that contain the target SNP sequence. Fluorescence intensity is then translated into nucleic acid abundance. SNP arrays have been typically used in GWAS to associate genetic variants with diseases or particular phenotypic traits [74]. In addition to that, some companies that offer genealogical services, like 23andme (https://www.23andme.com/en-int/) or FTDNA (https://www.familytreedna.com/), use high-density SNP arrays to provide genetic testing for biogeographical ancestry, including a high number of autosomal, mitochondrial, and Y chromosome SNPs.

Overall, although the price of the arrays has lowered in the last years [205], this method offers the opportunity of genotyping a large number of SNPs at a scale that far exceeds current forensic casework requirement [74].

1.3.3.5 Massive parallel sequencing (MPS)

Over the last years, the SNP genotyping platforms based on massive parallel sequencing (MPS) have popularized not only in medical genetics, but also on Forensic Genetics for analyzing SNPs related to phenotype and ancestry, and STRs. The first limitation of these platforms was the cost, the DNA amount needed for the analysis, and the complex data processing. However, the last efforts made by the forensic community have allowed to generate and validate platforms compatible with casework samples [83,84].

These methods, which are considered second-generation sequencing methods [43,206], offer the advantages over capillary electrophoresis systems of more deep sequence information and an increased discrimination power of a forensic profile through the possibility to analyze

simultaneously hundreds of markers in one multiplex PCR, not only SNPs but also STRs and complete genome sequence. Likewise, the complexity of data processing and interpretation, and the influence of external factors (i.e. costs, instrument performance, staff training, and quality levels of reagents) may increase as well [84].

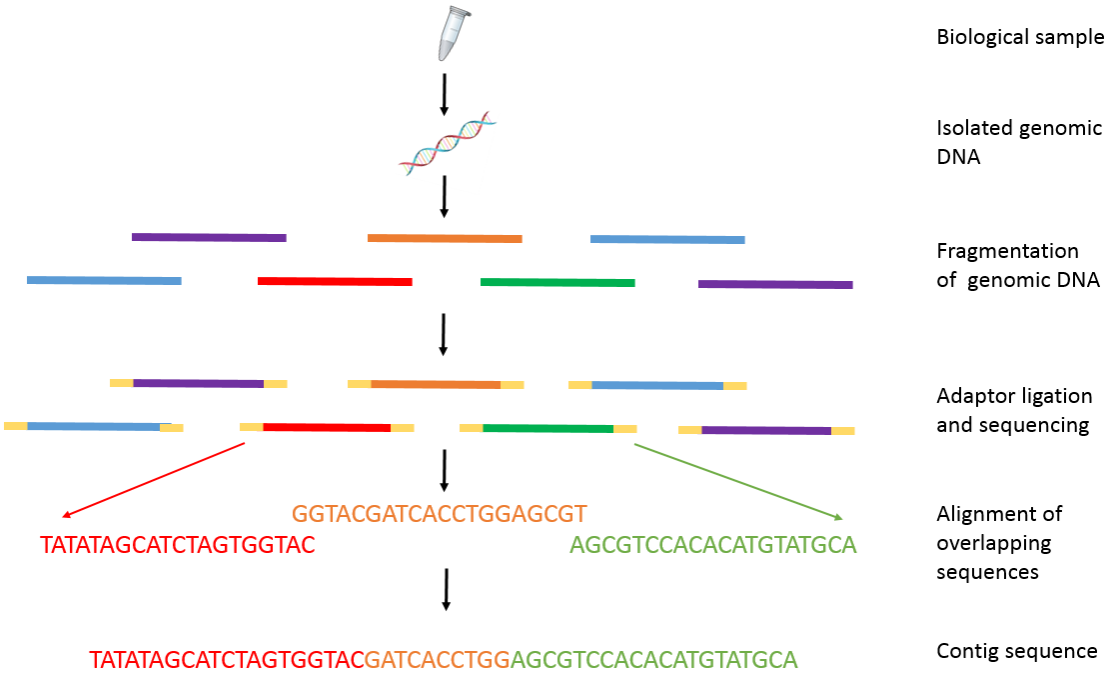


Figure 10. Basic principle of massive parallel sequencing technologies. Adapted from [207].

The most widely used MPS techniques in forensics are PCR based capture methods combined with sequencing-by-synthesis using reversible dye terminators, represented by the Illumina® MiSeq®/NexSeq platforms, and ion semiconductor sequencing, represented by the Ion Torrent™ platforms Personal Genome Machine ™ (PGM)/S5 [42,43] (Figure 10). From 2014 on, the first MPS commercial kits specifically designed for forensic applications were released by the companies Thermo Fisher Scientific (HID-Ion AmpliSeq™ Identity, HID-Ion AmpliSeq™ Ancestry, and mtDNA Whole Genome Panel), Verogen (ForenSeq™ DNA Signature Prep Kit), Qiagen (Qiagen SNP ID-kit, Globalfiler™ system) and Promega (PowerSeq™ systems), which include autosomal SNPs, Y-SNPs and/or STRs (autosomal, Y-STRs, and/or X-STRs). Furthermore, interlaboratory validation studies [84], as well as commercial kit validation studies are already available for both platforms [83,208–210]. A survey performed in 2017 among European forensic laboratories reported that the most used kits were the ones supplied by the companies Verogen (formerly Illumina) and Thermo Fisher Scientific in an equal proportion [43,211]. Those kits are Forenseq™ DNA signature kit [43,208,211–216], and the HID-Ion Ampliseq™ panels [217–221].

By now it seems clear that before MPS becomes a routine Forensic Genetic diagnostic tool there is still a lot of work left to do, like developing guidelines and accommodating the national DNA databases to accept markers identified by MPS [222]. What is more, at the moment not all forensic laboratories can afford to implement MPS technologies to their routine analysis.

1.3.4 Y chromosome genetic databases

As mentioned in previous sections, DNA databases are essential for forensic casework and population genetics. Given the particular nature of the Y chromosome, specific databases have been devoted to store information of Y chromosome markers.

1.3.4.1 Y-STR databases

A number of online Y-STR databases exists which contain genetic profiles of anonymous individuals used to estimate the frequency of specific Y-STR haplotypes (Table 6). Some of them are of strict forensic use, while others contain Y-STR haplotypes associated with specific individuals and family names gathered by genetic genealogy companies that try to make genealogical connections.

Table 6. Summary of available online Y-STR databases (as of November 2018). Adapted from [14].

Database	Use	Number of samples	Number of Y-STR markers tested	Website
Y-STR Haplotype reference database (YHRD)	Forensic	265,324	7-29	http://www.yhrd.org
US Y-STR Database (US Y-STR)	Forensic	32,972	11-23	http://www.usystrdatabase.org/
Yfiler haplotype database	Forensic	11,393	17	http://www6.appliedbiosystems.com/yfilerdatabase/
Family tree DNA (FTDNA)	Genealogy	137,512	12-67	https://www.familytreedna.com/projects.aspx

Among the detailed databases in Table 6, the largest and most used in forensics is the YHRD, created by Lutz Roewer and Sascha Willuweit in 2000 [223]. This database contains results from more than 250,000 samples with at least the minimal haplotype loci results from different populations

and countries around the world. Searches in the database may be conducted by population group or geographic location, and in addition to that, it also offers tools for statistical calculations such as mixture analysis, kinship analysis, analysis of molecular variance (AMOVA) and multidimensional scaling (MDS).

Genetic genealogy databases are not typically used for Y-STR forensic haplotype frequency estimates, as they include limited information, but they can be helpful for associating a concrete Y-STR haplotype with a particular family surname [159], in case it was necessary for casework investigation [14].

1.3.4.2 Y-SNP databases

Most Y-SNP information can be found in general online SNP databases (detailed in section 1.2.2.2.3), but there are only a few of them that are exclusively dedicated to these markers and their phylogeny. Apart from their position in the Y chromosome phylogeny, other information like alternative names, chromosome position, distribution and age are also offered. Some of them even include Y-SNP typing results. Given the confusion that has arisen with Y-SNP nomenclature [170] it is highly recommended to check the phylogenies managed by Phylotree Y and/or the ISOGG (Table 7).

Table 7. Summary of available Y-SNP databases (as of November 2018).

Database	Phylogeny information	Y-SNP results	Webpage
Y-STR Haplotype reference database (YHRD)	Yes	Yes	http://www.yhrd.org
Phylotree Y	Yes	No	http://www.phylotree.org/Y/tree/index.htm
International Society of genetic Genealogy (ISOGG)	Yes	No	https://isogg.org/tree/index.html
Ybrowse from the ISOGG	Yes	No	http://ybrowse.org/gb2/gbrowse/chrY/?
Family Tree DNA (FTDNA)	No	Yes	https://www.familytreedna.com/projects.aspx
Yfull	Yes	No	https://www.yfull.com/tree/

None of the detailed databases, except the YHRD, is used in forensic routine. Nevertheless, these databases are still highly popular in population genetics and genetic genealogy thanks to amateur genealogists and citizen scientists, a huge and active community that has provided knowledge in the genealogical field and even contributed with novel information in the Y chromosome phylogeny [224].

1.3.5 Applications of the analysis of the Y chromosome

In the previous sections of the present work the utility of the analysis of the Y chromosome and its markers has been addressed. Although small, this chromosome offers a wide range of applications in different fields like forensic, population and evolutionary genetics, genealogy and demography.

1.3.5.1 Forensic Genetics

1.3.5.1.1 Paternity and kinship testing

Father and sons, as well as all the males of the same paternal lineage share the same Y chromosome, which its transmitted almost unchanged to the next generation. This feature allows to reconstruct familiar relationships through the analysis of Y-STRs and Y-SNPs, and is also relevant in disaster victim and missing person identification [14,95,159].

In paternity testing, analyzing the Y chromosome is relevant in deficiency cases, when autosomal profiling is impossible or difficult, where the putative father of a male child is unavailable due to different circumstances, like being deceased. Y-STR analysis allow to exclude putative fathers, while inclusion can be difficult due to all individuals of the same lineage sharing the same Y chromosome. For this purpose, Y-STRs with low or medium mutation rated will be used, as finding the same haplotype indicates biological paternity [95].

1.3.5.1.2 Biogeographical ancestry

The analysis of Y-SNPs can allow to determine the paternal lineage of an individual. Since the distribution of haplogroups is not random, the detection of a particular paternal lineage can give clues about the paternal biogeographical ancestry of an individual [14,95,225].

In forensic practice this application should be taken with caution, since Y-SNPs only give information about the paternal side, and lineage prediction on admixed individuals is not considered to be reliable [94]. For that reason, it is best to combine Y-SNP analysis with mtDNA and AIM genotyping in order to provide an effective biogeographical ancestry prediction.

1.3.5.1.3 Mixture analysis

The Y chromosome may aid in some mixture cases, particularly where autosomal tests are limited by the evidence, like the presence of high levels of female DNA in the presence of low amounts of male DNA or mixtures with a high number of male individuals, as it happens in ‘gang rapes’. This

situation usually happens in the following cases: sexual assault evidence from vasectomized or azoospermic males, and blood-blood or saliva-blood mixtures where the absence of sperm does not allow the differential extraction of male DNA [14,226].

1.3.5.2 Population Genetics

1.3.5.2.1 Population stratification

The relationship between Y chromosome markers, both Y-STRs and Y-SNPs, paternal lineages and surnames can allow to detect population stratification, both past and more recent. Population stratification can go unnoticed due to not being able to detect it through the genome analysis of living individuals, or owing to the scarcity of available ancient DNA profiles from past individuals [159,227]. The genealogical data associated to Y-SNPs makes it possible to indirectly study the population differentiation of distinct time periods [228].

1.3.5.2.2 Male mediated expansion

Y chromosome markers are as popular for detecting migration patterns as their maternal counterpart, mtDNA. This is due to its strong geographic differentiation linked to patrilocal marriages [229], the wide-range of mutation rates of its markers, and the fact that most societies are patriarchal [110,159]. Customs surrounding marriage practices influence the migration behavior of the different sexes, and can affect the diversity of the Y chromosome. In the last years, several Y chromosome resequencing studies [129,165,168,230,231] have coincided in detecting population bursts of expansion within specific paternal lineages in the past thousand years [110].

1.3.5.2.3 Time to the most recent common ancestor (TMRCA)

Investigating how an ancestral population diverges to give rise to distinct subpopulations remains a fundamental pursuit in population genetics. Through the analysis of Y-STRs and Y-SNPs the age of a concrete paternal linage can be estimated, that is, the time since the haplogroup-defining mutation occurred and, thus, provide insights into the origin and history of particular populations, haplogroups or past human migrations [140,167,232,233].

1.3.5.3 Evolutionary Genetics

By comparing the differences between different Y chromosomes within concrete patrilineal lineages mechanisms driving sex chromosome evolution can be studied [159,234]. The analysis of Y chromosome variation in deep-rooting pedigrees has been useful for the detection and

characterization of the evolution of new regions of the Y chromosome structure [235] and the selection of relevant Y-SNPs in the Y chromosome reference phylogeny [171,236].

1.3.5.4 Genetic Genealogy

Classical genealogy is based on the research of archival evidence, like original genealogical records and documents in civil or parish records to reveal familiar or genealogical connections [110,159]. Genetic genealogy combines the use of this type of evidence with genetic tests, which allows to establish genetic relatedness in absence of these documents.

In this context, genealogical data can help to verify the observed genetic population structure and correctly interpret it, estimate more reliably the time scale of the gene flow events detected, and determine the temporal genetic differentiation within a concrete population [227]. Furthermore, a relationship exists between Y chromosome haplotypes, both based on Y-STRs and Y-SNPs, and patrilineal surnames [237,238]. The study of this relationship in different countries has revealed the effects of past social structures in the current diversity of the Y chromosome [238–241].

1.3.5.5 Demography

The study of the Y chromosome enables to estimate and compare rates of extra-pair paternity (EPP) within and between human populations, both past and present. The EPP rate can be directly estimated from mismatches in Y chromosomal genotypes between pairs of individuals that share a common paternal ancestor based on genealogical evidence [242]. The study of EPP can help solve demographic historical questions and ascertain bias both in evolutionary demographic studies and the analysis of biological traits [159,243,244].

1.4 Evolution and history of the genetic makeup of Europe

Europe is a geographical area without well-defined boundaries located in the western part of the Eurasian subcontinent. Directly connected with Asia through the Middle east, and with North Africa through the Mediterranean Sea, this continent was populated during prehistoric times due to several waves of human migrations, and those demographic events have shaped the current gene pool of the European populations. Thanks to the insights provided by ancient DNA and modern population studies, it has been possible to reconstruct the most probable evolutionary scenario that determined the ancestry of the modern Europeans [245,246].

1.4.1 The paternal genetic landscape of Europe

The genetic makeup of Europe has been defined by complex processes such as human migrations and population settlements influenced by environmental change, historical conquests of the territory and cultural progress [245,247]. The analysis of Y-SNPs has enabled to establish the haplogroup composition of the current European population.

The study of several genetic markers from autosomal DNA and the Y chromosome revealed a southeast-northwest frequency cline [247–252]. Nowadays the most common paternal lineage in Europe is the macrohaploup R, defined by the Y-SNP M207 (Figure 11) (Table 5). In Europe haplogroup R is divided in two main subhaplogroups: R1a, which is defined by M420 and is more common in Eastern Europe in frequencies between 3-60% [247,252–254], and R1b, which is defined by M343 and is more common in Western Europe in frequencies up to 90% [247,252,254]. The major Western European paternal lineage is the R1b subhaplogroup M269 [255–258], which displays frequencies between 12-90% in this geographic area [257–259] (Figure 12).

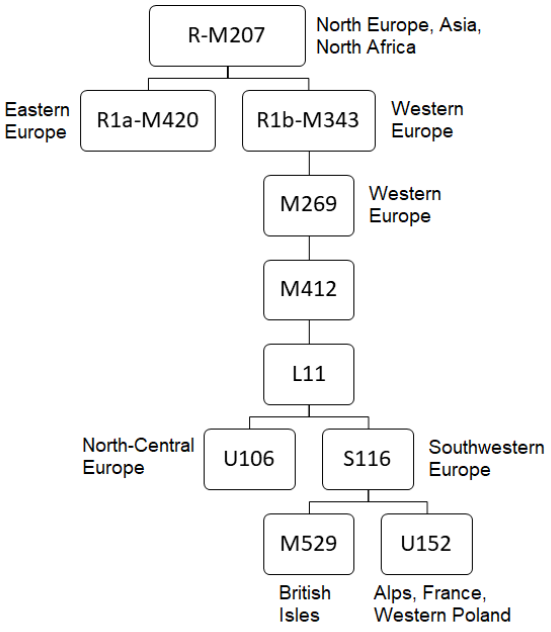


Figure 11. Simplified phylogeny of haplogroup R. The haplogroup assignment follows the minimal reference phylogeny for the human Y chromosome [171], supplemented with the more detailed tree maintained by the International Society of Forensic Genetics (ISOGG).

M269 shows a northwest-southeast frequency cline, from high frequencies in the west to lower frequencies in Eastern Europe [257–260]. As it happens with its mother haplogroup, M269 is also divided in many sublineages, such as M412 and L11, which are restricted to Western Europe [258,259]. The L11 derived sublineages S116 (also known as P312) and U106 are distributed in North-Central

Europe and Southwestern Europe respectively [258–260]. Finally, S116 is also divided in two main subhaplogroups that are geographically localized: M529 in the British Isles, and U152 in the Alps, France and Western Poland [258–260].

Apart from the mentioned R derived lineages, other minor haplogroups can also be observed in the European continent, such as E, J, I and G [261,262], defined by the Y-SNPs M96, M304, M170, and M201 respectively (Table 5). E haplogroup has been observed to display frequencies up to 25% [247,261], while J and I reach frequencies up to 27% [247,261] and 45% [247,262] depending on the geographic location. G is present only in scarce frequencies, between 5-15% [247].

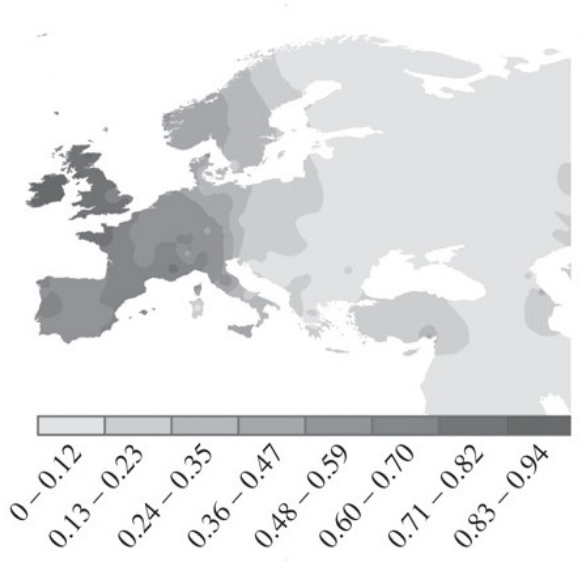


Figure 12. Frequency distribution of the haplogroup R1b-M269 in Europe. Extracted from [259].

1.4.2 Genetic history of Europe

Uniparental markers such as the Y chromosome and mtDNA are easy to track through generations and provide a unique male and female perspective of the human evolutionary history [263]. Their analysis, along with autosomal data, is crucial as they provide insights into demographic and social factors like sex-biased introgression or mobility, which allow to reconstruct the structure and movements of past populations [264]. For that reason, they are valuable markers in paleogenetics.

The current genetic diversity of Europe has been shaped by three critical major demographic episodes [245,265,266]:

- The arrival of anatomically modern humans 45,000 years ago (ya) to Europe during the Paleolithic.

- The last glacial maximum (LGM) between 27,000-16,000 ya, where human populations retreated into refugia in the Iberian, Italian and Balkan peninsulas, and the posterior northward recolonization 14,000 ya.
- The arrival of agriculture, which originated in the Near East, in southeast Europe during the Neolithic transition, and its spread throughout the rest of the continent between 9,000-5,000 years ago (Figure 13 A).

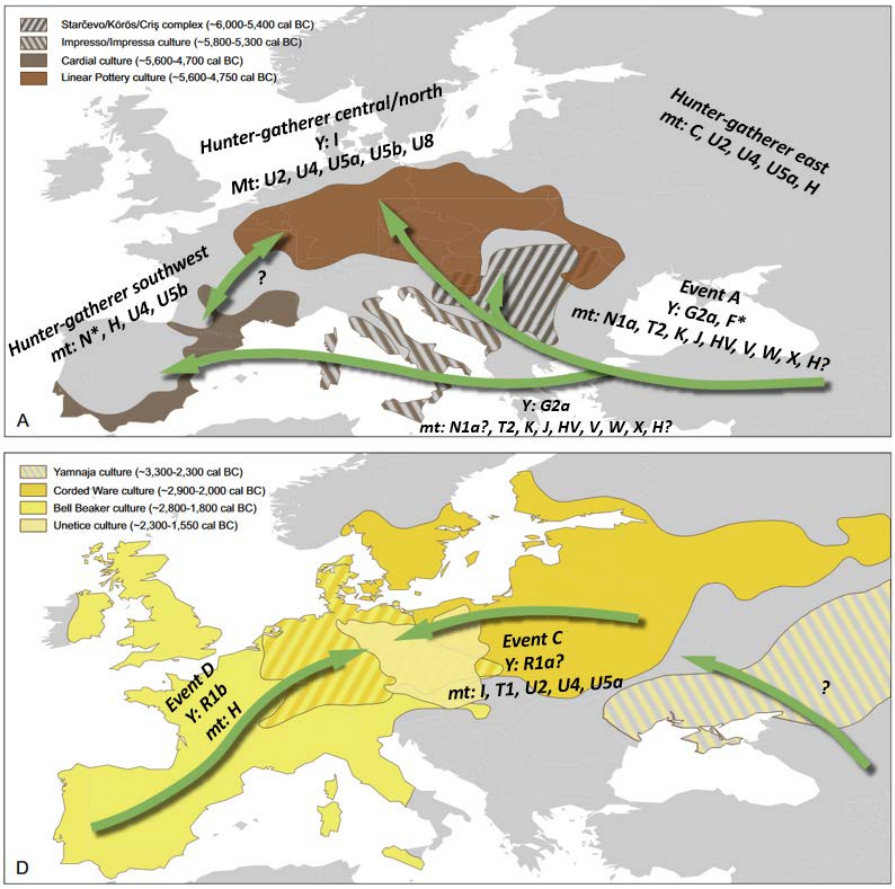


Figure 13. Summary of population dynamic events during the Neolithic Period in Europe. Different shadings and patterns denote the geographic distribution of cultures during this period: A) Early Neolithic. D) Late Neolithic/Early Bronze Age. Event A: The impact of incoming farmers during the Early Neolithic. Event C: Period of renewed genetic influx during the Late Neolithic with variable regional repercussions. Striped areas indicate archaeological culture for which ancient DNA data is not available so far. Green arrows display potential geographic expansion routes and their associated paternal or maternal lineages. Extracted from [266].

The first two episodes, that is, the arrival of anatomically modern humans to Europe (45,000 ya) and the LGM where human populations retreated to refugia (27,000-16,000 ya), occurred during the Paleolithic (up to 11,500 ya) and Mesolithic (11,500-5,000 ya) periods in Europe. The humans that inhabited the continent during the Paleolithic and Mesolithic periods were Hunter-gatherers (HG) [245,266]. Ancient DNA studies have revealed that the predominant Y chromosome haplogroup in HGs from North and Central Europe (8,000 ya) was I-M170 [267–269], which has been proposed to

originate in the Paleolithic and expand after the LGM [262]. Likewise, the most common mtDNA haplogroup in the Upper Paleolithic (50,000-11,500 ya) and Mesolithic (11,500-5,000 ya) HGs from Central, East, North and Southern Europe was U and its derived subhaplogroups [267,270–274] (Figure 13 A).

The genomic data (mtDNA and autosomal DNA) recovered from European Mesolithic individuals confirms a genetic discontinuity between Mesolithic and early farming individuals [265,267,273,275,276], and suggest that the population structure of pre-agricultural Europe was more complex than previously thought [266]. Furthermore, ancient DNA analysis has suggested that modern-day Europeans derive from at least three highly differentiated populations: west european HGs, who contributed ancestry to all Europeans but not to near easterners; ancient north eurasians, which are related to Upper Paleolithic siberians [276]; and early european farmers, of Near Eastern origin but who also possessed West European HG related ancestry [267].

The Neolithic transition emerged in the Near East around 12,000 years ago and is described as one of the most fundamental cultural changes in human history [277,278], which involved the transition from foraging to agriculture and animal domestication with a more sedentary way of life [278,279]. It reached Southeastern Europe around 7,000 years ago, expanding to the rest of the continent at later times. The available data from Central, South and Eastern European early farmers (7,000-5,000 ya) suggest that the predominant Y chromosome haplogroup was G2a-P15 (Table 5), which is rare in modern Europeans [280–284], followed by the I sublineage I2-PF3835 [269,283,285,286]. Other haplogroups also observed in early farming sites are F-M89, and E1b-M35 [280,281,283,284]. Conversely, the most common mtDNA haplogroups from Central and Southwest European farmers in the early Neolithic period (7,500-5,000 ya) were N1a, T2, K, J, HV, V, W, and X, while in the late Neolithic higher frequencies of U were observed [266,267,272,280,281,284,287–292] (Figure 13 A).

The most prevailing current European paternal lineages, R1a and R1b, have not been reported in the fossil record until the late Neolithic (4,000-3,000 ya), in remains associated to the cultures Bell Beaker in Western Europe [288] and Corded Ware in Eastern Europe [264], which coexisted for more than 300 years and overlapped in Central Europe until they were replaced by the Unetice culture in the Bronze Age (Figure 13 D). Likewise, the Bell Beakers show higher affinity with modern South Europeans, while the Corded Ware individuals show greater similarities with modern Eastern Europeans [266]. Both cultures have been associated with migration processes during the late Neolithic [289]. The succeeding historical period, the Bronze Age, started around 4,300 years ago and is associated with changes in burial practices, the spread of horse riding, and developments in weaponry [293,294].

1.4.3 The controversy of the origin of R1b-M269

The origin of the major haplogroup in modern West European males has been the subject of heated controversy due to the differing estimated times to the most recent common ancestor (TMRCA) obtained by different authors, which placed its origin either in the Paleolithic or the Neolithic. On the one hand, some authors obtained older TMRCAs for M269, placing its origin in the Paleolithic, after the LGM, and assuming a postglacial expansion from the Franco-Cantabrian refuge, which would explain the current pattern of frequencies [247,252,294,295]. These authors based their theories not only in Y chromosome data, but also in mtDNA haplogroup information [296], particularly H lineage, which shares a similar pattern of frequencies as R haplogroup. Another study supported the postglacial expansion of M269, but suggests that this haplogroup could have had a parallel expansion from a refuge located in Anatolia (Eastern Europe) towards Southeast Europe, based on the Y-STR variance within M269 [297].

On the other hand, Balaresque and colleagues [257], based on the higher diversity of Y-STR haplotypes observed in Eastern European M269 males in contrast with the frequency cline of this haplogroup, estimated the origin of the lineage around 6,000 years ago in Eastern Europe, during the Neolithic. These authors applied a germinal mutation rate instead of an evolutionary one for calculating the TMRCA [257]. However, this proposal was strongly challenged by Busby and colleagues [259], who recalculated the diversity of the Y-STR haplotypes within M269 in a larger and geographically broader sample, and observed a homogeneous Y-STR variation in the whole group of European samples. A posterior study that analyzed some M269 subhaplogroups in a larger sample of Europe obtained coalescence times compatible with Balaresque's proposal, suggesting a spread of M269 also in the Neolithic [258]. Moreover, they linked the spread of M269 with the *Linearbandkeramik* culture, which spread throughout Northern Europe around 7,500 years ago, from Hungary to France. Finally, another theory suggested that M269 arrived to the Iberian Peninsula during the Neolithic and linked the appearance and subsequent dissemination of the sublineages S116 and M529 with the bell Beaker expansion northwards [298].

The controversy cannot be more interesting, as it has generated a cordial and productive discussion to unravel the evolutionary history of M269. The analysis of ancient DNA from both Paleolithic and Neolithic remains, the better characterization of this haplogroups in more areas of Europe, as well as the availability of more whole sequences of the Y chromosome that will allow to make more robust date estimations, may definitely reveal the complete history of the West European major lineage.

1.4.4 Molecular dating of paternal lineages

The most recent common ancestor (MRCA) is the most recent individual from which all the individuals are directly descended. The MRCA can sometimes be determined by constructing a pedigree but, in general, it is impossible to identify the exact MRCA in a large set of samples. What can be done is to estimate the time at which the MRCA lived, that is, the time to the most recent common ancestor (TMRCA), by using genetic data. A rooted phylogeny can provide a relative chronology for genetic changes. Mutations that appear closer to the tips of a phylogenetic tree must have occurred after those that are closer to the root within the same clade. In order to provide reliable estimates, it is necessary to construct a chronological record, that is, to provide a context to the genetic data by placing the timing of a change in a wider context, maybe related to an archaeological culture or a paleoclimatological event [299].

The chronological record can be constructed by establishing molecular clocks, which are processes of variation that change predictable with time. This process can be mutation, recombination or genetic drift. Apart from that, it is also necessary to calibrate the molecular clock, that is, to know the rate of change. And the way to do that is by calculating the mutation rate [300]. The non-recombining portions of the genome, that is, the Y chromosome and mtDNA, contain haplotypes that can be related by a single most parsimonious phylogeny and allow to date all the nodes of a phylogeny [299].

1.4.4.1 TMRCA calculating methods

Several methods exist to calculate the TMRCA of a set of chromosomes that share a common mutation at a unique marker. These methods can be classified into those that involve a population model and those that do not need that model [301]. This last class uses summary statistics of intra-allelic diversity in order to date an allele, which increase linearly with time. For non-recombining haplotypes, mutation drives diversification and, consequently, it represents the molecular clock [299]. Those summary statistics are Rho (ρ) [302] and the average square distance (ASD) [303,304]. Rho needs to construct a phylogeny while ASD does not need to. The main limitation of these methods is that it is difficult to get a true estimate of their 95% confidence limits, although they can provide unbiased point estimates. The source of error is usually the uncertainty in the parameters and/or in the demographic history of a population [299].

On the other hand, model-based TMRCA estimation methods are based on the "coalescent theory" [305,306] and are more computationally intense. They are also called 'Bayesian approach' methods. Their main advantage is that they use all the information present in the genetic data,

suggesting the most likely phylogenetic tree and its age, and a wide range of parameters can also be estimated. However, the accuracy of the estimates is prone to bias resulting from choosing the appropriate demographic model [299,300]. Three popular softwares that have implemented these methods are BATWING [307], Genetree [308] and BEAST [309], among others.

1.4.4.2 Mutation rates

Two types of markers can be used to calculate the TMRCA of Y chromosome haplogroups: the rapidly mutating Y-STRs, and the slowly mutating Y-SNPs [300]. The calibration of the Y chromosome mutation rates can be performed using three approaches [299,300]:

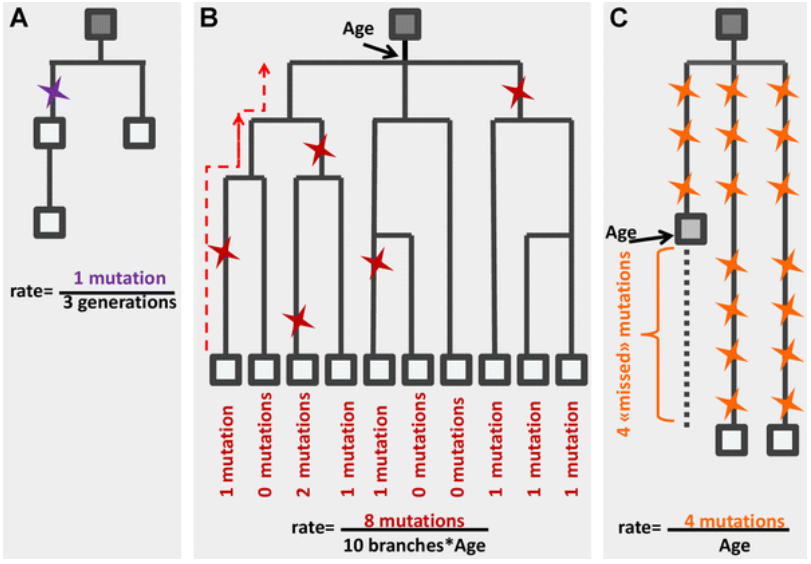


Figure 14. Three approaches to estimate the mutation rate on the Y-chromosome. A: Genealogical approach. Mutations separating members of the pedigree are counted and divided by the number of generations. B: Calibration approach. The average number of mutations from the MRCA to the modern samples divided by the TMRCA, which is assumed to coincide with a population event of known date. C: Ancient DNA approach. The older the ancient sample is, the less time it has had to accumulate mutations. Thus, the number of "missed" mutations is proportional to the (radiocarbon) age of the sample. Extracted from [300].

- I) Genealogical approach, based on pedigrees, counting mutations along the genealogical line and dividing by the number of generations or years. This results in a genealogical mutation rate (GMR) (Figure 14 A).
- II) Calibration approach, where the average number of mutations from the MRCA to the modern samples is divided by the TMRCA, which it is assumed to coincide with a historical event of known time (Figure 14 B). It is also known as the evolutionary mutation rate (EMR)
- III) Ancient DNA approach, based on phylogenetic lineages whose evolution has stopped long time ago. The older an ancient sample is, the less time it has had for mutations to accumulate than present-day samples. The number of 'missed' mutations is proportional to

the age of the ancient sample. This approach uses real samples from human evolutionary history, whose age is often known through radiocarbon dating (Figure 14 C).

In the last years researchers have estimated the TMRCA of Y chromosome haplogroups using both Y-STRs and Y-SNPs, selecting both the EMR or the GMR. Selecting one mutation rate or the other to estimate haplogroup age from Y-STRs has been the subject of great controversy [167,310–312]. The last studies conclude that GMR often provides better estimates for haplogroups that are younger than 7,000 years, while the EMR can estimate correctly (or overestimate) haplogroups older than 15,000 years. However, if available, Y-SNP based ages should be employed as they provide more reliable estimations [300].

2. Hypothesis and objectives

2.1 Hypothesis

The study of the human Y chromosome due to its haploid character and its male-specificity is a powerful tool in forensic analysis and genetic genealogy. In Forensic Genetics, Y chromosome testing is widely used particularly in cases where standard autosomal DNA profiling is not informative, as well as to exclude male suspects for involvement in a crime, identify the lineage and the paternal biogeographic ancestry of male perpetrators, and provide investigative leads for finding unknown male perpetrators [95,110]. Likewise, in genetic genealogy the Y chromosome is vastly used for kinship analysis, evolutionary studies, familiar searching, and surname and demography studies among others [159,160]. The markers of choice for these types of analysis are the well-known Y chromosome *short tandem repeats* (Y-STRs) and *single nucleotide polymorphisms* (Y-SNPs).

Up to now, the study of Y-SNPs has enabled to know that particular paternal lineages are restricted to specific geographic areas at continental and regional levels, and the analysis of these lineages is highly useful to reconstruct the evolutionary history of the human species [161]. These Y-SNPs are also of great interest in Forensic Genetics, as they allow to link a concrete biogeographical ancestry with a vestige.

The current genetic makeup of Europe is the result of many population migrations and settlements influenced by climate, cultural progress, and historical conquest of territory among other causes [245]. The most common paternal lineage in Europe is R1b-M269, which is shared by 40-90% of the males in Central and Western Europe and follows a cline of increasing frequencies from East to West that peaks in the British Isles and Northern Iberia [257,258]. Its origin has been the subject of heated controversy due to the discrepancies existing in the estimated times to the most recent common ancestor (TMRCA), which place the advent of this haplogroup in the Franco-Cantabrian refuge during the Paleolithic [247,294,295], or more recently during the Neolithic in Eastern Europe [257,258].

This controversy has made clear the requirement of a more precise understanding of the structure and distribution of haplogroup M269, of forensic and population interest, with more comprehensive sampling schemes including more information regarding some areas of Southwest Europe like the Atlantic coast and Iberia.

The **first hypothesis** of the present thesis work is based on the fact that a better understanding of the structure of M269 through its dissection in its subhaplogroups, would allow to obtain more

reliable age estimations and adjust or even rewrite the theories of the European peopling, clearing the controversy regarding its origin.

In addition to that, in forensic and population genetics the use of multiplex tools for genotyping markers of forensic interest like STRs and SNPs has been well established. The use of these panels allows the simultaneous genotyping of several markers in a unique reaction (or a reduced number of them), and are a time saving and cost-effective solution of easy implementation in any laboratory without the requiring of additional equipment. Currently, there are several commercial Y-STR panels available, which include STR markers that exhibit high to low-range mutation rates and are applicable for identification and kinship testing. However, due to the nature of the markers included in these panels there may be limitations for their application in particular cases, like exclusion cases with minimal discrepancies or evolutionary studies. Similarly, several Y-SNPs panels are available in the literature that include markers related to the most common Y chromosome haplogroups or some concrete sub-branches. These multiplex tools are based on the minisequencing technique, with capillary electrophoresis as its detection system. Although in the last years they have gained popularity in Forensic Genetics due to their application for inferring biogeographical paternal ancestry, there exists limitations in their power of population discrimination, the number of multiplex reactions necessary, and DNA sample consumption.

In view of the above mentioned, the **second hypothesis** of the present work is centered around the demand to develop more efficient multiplex STR and SNP panels of forensic application. It is probable that new panels used in conjunction with the current ones will allow, on the one hand, to resolve those particular cases that current Y-STR panels are not able to respond to and, in the other hand, to obtain higher haplogroup resolution in Y-SNP panels that will enable to improve male lineage discrimination in concrete branches.

2.2 Objectives

Main objective

The main objective of this doctoral thesis work is to reconstruct the most probable evolutionary scenario of the main European paternal lineage R1b-M269 in the Iberian Peninsula and Southwest Europe through the dissection in its subhaplogroups by the analysis of Y chromosome SNPs (Y-SNPs) and innovative statistics. This will allow us to carefully characterize the paternal ancestry landscape of the Iberian Peninsula and infer the role of this area in the European evolutionary history, which will explain how the vast majority of the Southwest European gene pool is distributed the current way. Furthermore, the genetic data generated in this study will be of great interest in Forensic Genetics for the detection of Iberian and/or Southwest European paternal biogeographical ancestry, as it will provide novel and more detailed information of the distribution of lineages below R1b-M269, improving the resolution at European regional level.

Specific objectives:

1. To analyze R1b-M269 paternal lineage and its current sublineages in populations of Atlantic Europe and the Iberian Peninsula, which will allow to define in detail the distribution of M269 in Southwest Europe and to obtain new clues about its evolutionary history.
2. To characterize the structure and spatial distribution of the Iberian near-specific paternal lineage R1b-DF27 in Southwest European populations through the dissection in its sublineages, with the aim to estimate its time of origin, as well as to model its expansion in the phylogenetic context and the related demographic events.
3. To design and optimize a new minisequencing method that allows the simultaneous analysis of 15 Y-SNPs for the fine subtyping of the Iberian paternal lineage R1b-DF27, with applicability in both forensic and population analysis.
4. To design, optimize and validate a novel panel of six Slowly Mutating Y-STRs, which can be used in conjunction with the existing multiplex commercial kits for forensic casework, particularly in complex kinship cases and in optimizing the prediction of paternal ancestry based on current Y-STR panels with medium-high mutation rates.

3. Materials and methods

3.1 Human DNA samples

3.1.1 Population samples

In the present doctoral thesis work several populations have been studied (Table 8). All the samples were obtained from volunteer male donors following the ethical principles of the 2000 Helsinki Declaration of the World Medical Association. For each study, the corresponding favorable ethical approvals were obtained.

Table 8. Summary of the population samples analyzed in the present doctoral thesis. N= number of individuals; BNADN= Banco Nacional de ADN Carlos III – Spanish national DNA bank (BNADN Ref. 12/0031); The samples from UPV/EHU Biobancos del Instituto de Salud Carlos III (Sección Colecciones) ref. C.0000214 were collected by BIOMICs Research Group once favorable ethical reports were obtained (Faculty of Pharmacy UPV/EHU, September 2008 CEISH/119/2012).

Study	***Population***	***Sample Size***	***Provided by***
Study Number 1 N= 1560	Alicante	N= 116	Miguel Hernández University – UMH Pathology and Surgery Dept.
	Andalucía	N= 100	BNADN
	Asturias	N= 63	BNADN
	Barcelona	N= 100	BNADN
	Basque Country	N= 341	UPV/EHU Biobancos del Instituto de Salud Carlos III (Sección Colecciones) ref. C.0000214
	Cantabria	N= 96	University of Cantabria – UC
	Galicia	N= 70	BNADN
	Madrid	N= 99	BNADN
	Portugal (Porto)	N= 110	National Institute of Legal Medicine and Forensic Sciences of Porto
	Brittany (Brest)	N= 145	University of Bretagne Occidentale – UBO
	Ireland	N= 146	Trinity Biomedical Sciences Institute – Academic Unit of Neurology
	Denmark	N= 174	University of Copenhagen – Forensic Medicine Dept.
Study Number 2 N= 591	Aragón	N= 92	University of Zaragoza – Forensic Medicine Dept.
	Asturias	N= 63	BNADN
	Basque Country	N= 340	UPV/EHU Biobancos del Instituto de Salud Carlos III (Sección Colecciones) ref. C.0000214
	Cantabria	N= 96	University of Cantabria – UC
Study Number 3 N= 2990	Alicante	N= 142	Miguel Hernández University – UMH Pathology and Surgery Dept.
	Andalucía	N= 100	BNADN
	Aragón	N= 92	University of Zaragoza – Forensic Medicine Dept.
	Asturias	N= 63	BNADN
	Barcelona	N= 571	BNADN; University Pompeu Fabra – Institute of Evolutionary Biology
	Basque Country	N= 340	UPV/EHU Biobancos del Instituto de Salud Carlos III (Sección Colecciones) ref. C.0000214
	Cantabria	N= 96	University of Cantabria – UC
	Castellón	N= 49	University Pompeu Fabra – Institute of Evolutionary Biology

Study	*Population*	*Sample Size*	*Provided by*
Study Number 3 N= 2990	Galicia	N= 70	BNADN
	Girona	N= 131	University Pompeu Fabra – Institute of Evolutionary Biology
	Lleida	N= 104	University Pompeu Fabra – Institute of Evolutionary Biology
	Madrid	N= 99	BNADN
	Mallorca	N= 48	University Pompeu Fabra – Institute of Evolutionary Biology
	Pyrenees	N= 46	University Pompeu Fabra – Institute of Evolutionary Biology
	Tarragona	N= 120	University Pompeu Fabra – Institute of Evolutionary Biology
	Valencia	N= 79	University Pompeu Fabra – Institute of Evolutionary Biology
	Alsace	N= 80	University of Santiago de Compostela – Forensic Science Institute
	Auvergne	N= 89	University of Santiago de Compostela – Forensic Science Institute
	Brittany (Brest)	N= 145	University of Bretagne Occidentale – UBO
	Île-de-France	N= 91	University of Santiago de Compostela – Forensic Science Institute
	Midi-Pyrénées	N= 67	University of Santiago de Compostela – Forensic Sciences Institute
	Nord-Pas-de-Calais	N= 68	University of Santiago de Compostela – Forensic Sciences Institute
	Provence – Alpes Côte d'Azur	N= 45	University of Santiago de Compostela – Forensic Sciences Institute
	Ireland	N= 146	Trinity Biomedical Sciences Institute – Academic Unit of Neurology
	Portugal (Porto)	N= 109	National Institute of Legal Medicine and Forensic Sciences of Porto
Study Number 4 N= 24	Basque Country	N= 21	UPV/EHU Biobancos del Instituto de Salud Carlos III (Sección Colecciones) ref. C.0000214
	Barcelona	N=2	BNADN
	Brittany (Brest)	N= 1	University of Bretagne Occidentale – UBO
Study Number 5 N= 628	Europeans from Spain	N= 319	UPV/EHU Biobancos del Instituto de Salud Carlos III (Sección Colecciones) ref. C.0000214, BNADN
	Asians from Thailand	N= 102	Colorado College – Molecular Biology Dept.
	Africans from Malawi	N= 31	University of Santiago de Compostela – USC
	Native Americans from Guatemala	N= 50	UPV/EHU Biobancos del Instituto de Salud Carlos III (Sección Colecciones) ref. C.0000214
	Hispanics from Colombia	N= 60	University of Antioquia
	Hispanics from Nicaragua	N= 66	University of Zaragoza – Forensic Medicine Dept.

3.1.2 Control samples

Human control DNA 2800M (Promega® Corporation, Madison, WI, USA) was used to set up PCR amplification conditions (*Studies Number 4 and 5*) and to conduct sensitivity and stability studies (*Study Number 5*).

3.2 DNA extraction

Human DNA from buccal swabs of the populations of Brittany and Alicante (*Study Number 1*) was isolated by organic extraction. The population of Aragon (*Study Number 2*) was extracted from blood stains in FTA papers with Chelex-100® chelating resin suspension (Sigma-Aldrich Corporation, St. Louis, MO, USA). The remaining population DNA samples had already been previously extracted.

The applied protocols for each of the above-mentioned methods are described below:

1. *Organic extraction*
    - In a laminar flow cabinet, cut the swab tip or half of the swab tip and put it into a centrifuge tube.
    - Add 500 µl of lysis buffer (50 µl NaCl 1 M, 50 µl EDTA 100 mM, 50 µl Tris 100 mM, 100 µl SDS 10%, 20 µl DTT 1 M, 205 µl Milli-Q water).
    - Add 25 µl of Proteinase K (20 mg/ml) and mix gently.
    - Incubate for 1 hour at 64 °C with shaking (750 rpm).
    - Mix for 5-10 seconds using a vortex mixer and then centrifuge.
    - In order to recover the amount of lysis product retained in the swab we used the double-tube method. The double-tube was constructed as follows:
        - Prepare 0.5 ml centrifuge tubes with a hole in their base and put them opened inside a 1.5 ml centrifuge tube.
    - Transfer with great caution the swabs to the double-tube, close the lid and centrifuge for 5 minutes at 13,000 rpm.
    - Discard the 0.5 µl centrifuge tube and transfer the recovered lysis product to the initial centrifuge tube.
    - In a fume cabinet, add 500 µl of phenol: chloroform: isoamyl alcohol (25:24:1) to each tube and shake by hand thoroughly for 5-10 minutes.
    - Centrifuge for 10 minutes at 13,000 g.
    - Remove carefully the upper aqueous phase and transfer it to a new tube.

- Add one volume of phenol: chloroform: isoamyl alcohol (25:24:1), similar to the one recovered previously, and shake by hand thoroughly for 5-10 minutes.
- Centrifuge for 10 minutes at 13,000 g.
- Remove carefully the upper aqueous phase and transfer it to a new tube.
- Add the following reagents in the listed order to the recovered aqueous phase and turn over the tubes 30-50 times:
  - 1/10 volumes of AcNa 2 M
  - 1 µl of glycogen (20 mg/ml)
  - 2 volumes of absolute ethanol at -20 °C
- Incubate the samples at -20 °C for 60-120 minutes to precipitate the DNA from the sample.
- Centrifuge for 20 minutes at 13,000 g.
- Carefully remove the supernatant without disturbing the DNA pellet.
- Add 1 ml of ethanol 70% and turn over the tube several times.
- Centrifuge for 5 minutes at 13,000 rpm.
- Discard the supernatant without disturbing the DNA pellet.
- Remove the remaining ethanol in the DNA concentrator (about 15-20 minutes) at 45 °C.
- Resuspend the DNA pellet in 30 µl of sterile Milli-Q water.
- Mix in a vortex for 10-20 seconds and centrifuge
- Incubate for 10 minutes at 65 °C and 800 rpm in a thermomixer.

2. *Chelex-100® chelating resin suspension*
   - Incubate the buccal swabs or FTA paper in 1 ml of sterile Milli-Q water at room temperature for 30 minutes, with shaking every 5 minutes.
   - Discard the swab.
   - Centrifuge for 1 minute at 10,000-15,000 g.
   - Discard the supernatant without disturbing the pellet, except for around 50 µl.
   - Resuspend the pellet in the remaining volume using the vortex gently.
   - Add 150 µl of Chelex-100® resin at 5%.
   - Incubate at 56 °C for 15-30 minutes.
   - Mix for 5-10 seconds using a vortex.
   - Incubate for 8 minutes at 100 °C.
   - Mix for 5-10 seconds using a vortex.
   - Centrifuge for 2-3 minutes at 10,000-15,000 g.

- Transfer the supernatant containing the DNA to another tube, with care of not taking any of the resin.

3.3 DNA quantification

DNA was quantified by using two methods: 1) Scientific NanoDrop™ 1000 Spectrophotometer (Thermofisher Scientific, Wilmington, DE, USA), and 2) Quanti-iT PicogGreen™ dsDNA Assay Kit (Thermofisher Scientific, Wilmington, DE, USA). After that, the DNA was diluted in Milli-Q water to a 1-3 ng/µl concentration.

3.4 Y chromosome phylogeny

The nomenclature of the Y-SNPs analyzed in the present doctoral thesis work follows the minimal reference phylogeny for the human Y chromosome [171], supplemented when necessary with the more detailed tree maintained by the International Society of Genetic Genealogy (ISOGG; https://isogg.org/tree/index.html; v.12.53, February 2017). The analyzed Y-SNPs correspond to the diagnostic positions that determine the corresponding haplogroups (Figure 15).

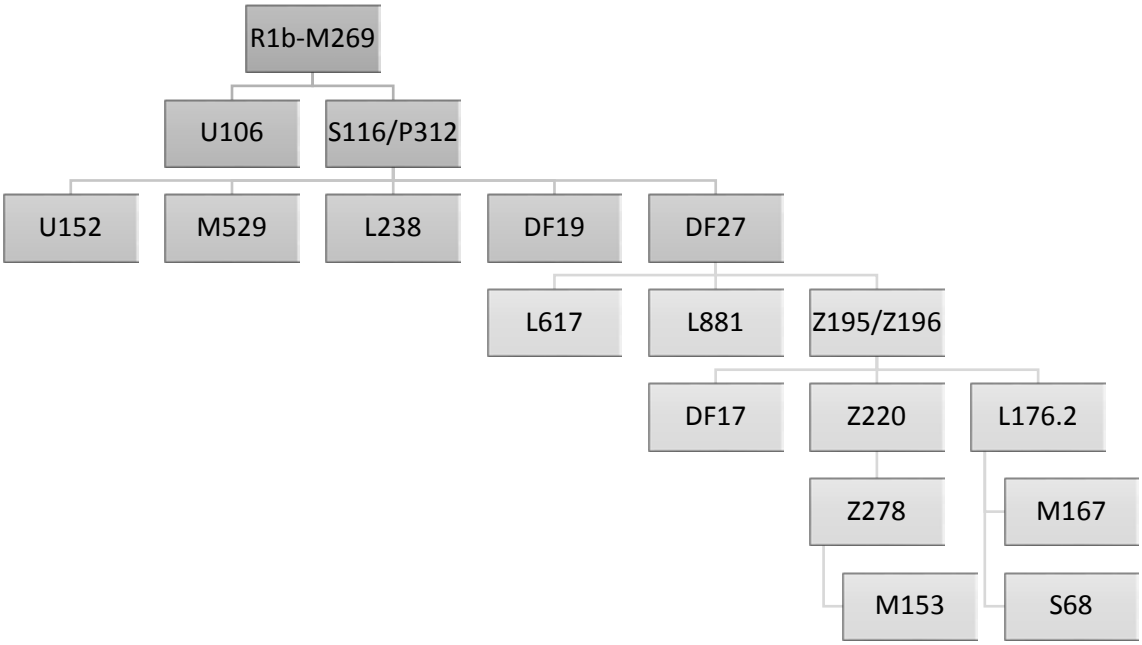


Figure 15. Simplified phylogenetic tree of the R1b-M269 haplogroup.

3.5 DNA amplification

The general workflow followed for the analysis of Y-SNPs and Y-STRs is as the one described in Figure 16. Part of the samples included in *Study Number 3* were analyzed in the Institut de Biologia Evolutiva from the University Pompeu Fabra (CSIC-UPF) by an Open Array panel as described in Solé-Morata and colleagues [240].

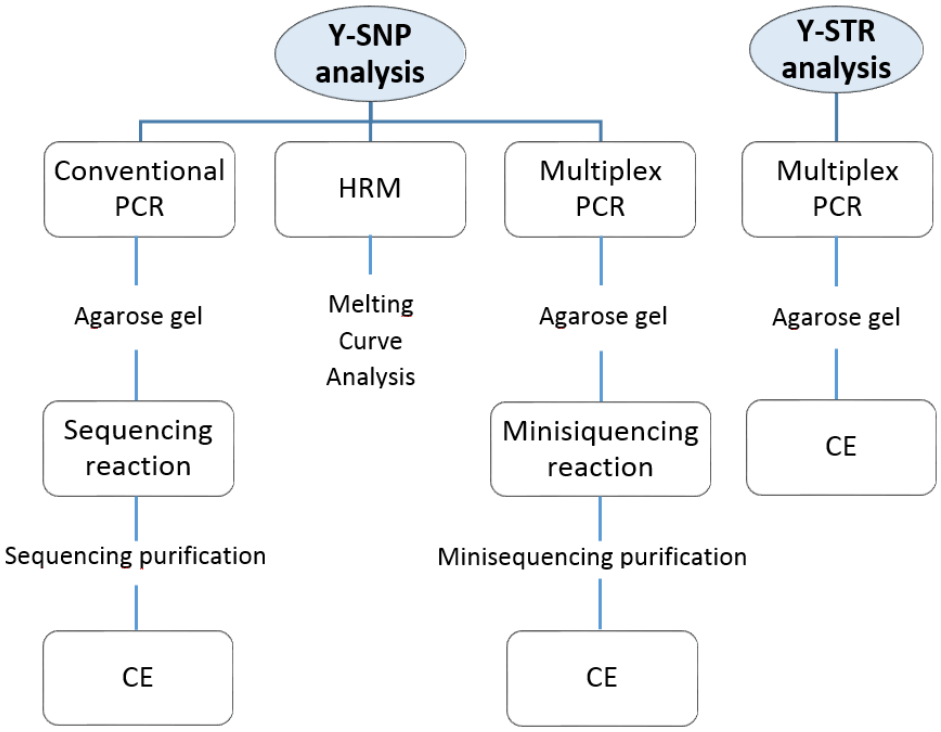


Figure 16. A schematic representation of the general workflow followed for the analysis of Y-SNP and Y-STRs. CE: capillary electroforesis.

3.5.1 Primer design and optimization

PCR amplification primers were designed with the software PerlPrimer v.1.1.21 [313] or manually. The specificity of the primers and their non-homology with the X-chromosome and other genome regions were confirmed with Primer-BLAST (https://www.ncbi.nlm.nih.gov/tools/primer-blast/). Potential unfavorable interactions between primers were checked with the web-based version of AutoDimer [314].

The optimal annealing temperature for each primer was assessed by testing the performance of the primers in temperatures between 55-62 °C.

3.5.2 PCR amplification

Conventional PCR amplification was performed for the analysis of the Y-SNPs M153, DF17 and L617, as well as for confirming by Sanger sequencing the corresponding genotypes of each HRM cluster. The reactions were carried out with 9.05 µl Milli-Q water, 0.6 µl of dNTPs (Bioline Reagents, London, UK), 0.45 µl of $MgCl_2$ 50 mM (Bioline Reagents, London, UK), 1.5 µl of buffer (Bioline Reagents, London, UK), 0.45 µl of each primer (10 µM), 0.3 µl of bovine serum albumin (Roche, Basilea, Switzerland), 0.2 µl of BIOTAQ™ DNA polymerase 5 U/µl (Bioline Reagents, London, UK) and 2 ng of DNA in a final volume of 15 µl.

The amplifications were carried out on the C1000™ Thermal Cycler (Bio-Rad, Hercules, CA, USA) under the conditions detailed in *Study Number 1*, which consisted in: pre-incubation at 98 °C 5 min; 35 cycles at 98 °C 30 sec, 30 sec at the corresponding annealing temperature (see *Studies Number 1* and *2*); 30 sec at 72 °C, and a final incubation at 72 °C for 10 min.

For the analysis of conventional Y-STRs (see *Study Number 1*) the kit AmpFLSTR Yfiler Amplification kit (Thermofisher Scientific, Wilmington, DE, USA) was employed following the recommendations of the manufacturer.

The site-directed mutagenesis conducted in *Study Number 4* was carried out with 8.9 µl of Milli-Q water, 0.6 µl dNTPs 10 mM (Bioline Reagents, London, UK), 0.6 µl $MgCl_2$ 50 mM (Bioline Reagents, London, UK), 1.5 µl buffer 10x (Bioline Reagents, London, UK), 0.3 µl bovine serum albumin (10x) (Roche, Basilea, Switzerland), 0.45 µl of each primer (Forward an Reverse) at 10 µM, 0.2 µl of BIOTAQ™ DNA polymerase 5 U/µl (Bioline Reagents, London, UK) and 2 ng of DNA.

3.5.3 High Resolution Melting (HRM)

All Y-SNPs were analyzed by HRM unless otherwise specified. The reaction contained 2.5 µl of SsoFast EvaGreen Supermix (Bio-Rad, Hercules, CA, USA), 0.5 µl of each primer at 1 µM and 1 ng of DNA in a final volume of 5 µl, and was carried out in a C1000™ Thermal Cycler (Bio-Rad, Hercules, CA, USA) equipped with a CFX96™ Real-Time PCR Detection System (Bio-Rad, Hercules, CA, USA). The conditions for amplification and melting were the following: 2 min at 98 °C; 35 cycles at 98 °C for 30 sec, 30 sec at the corresponding annealing temperature (see *Studies Number 1* and *2*); 30 sec at 95 °C, 2 min at 60°C, and finally the melting cycle from 65 °C to 95 °C with an increase of 0.2 °C/sec, for detecting the different allelic variants.

Data interpretation was carried out using the software Precision Melt Analysis v.1.2 (Bio-Rad, Hercules, CA, USA). Only high-quality amplification and melting curves with a cluster assignment

over 95% of confidence were considered. Positive and negative controls for each haplogroup, confirmed by Sanger sequencing, were used for performing the assignment of the corresponding allelic variants of every cluster.

3.5.4 TaqMan™ assay

In *Study Number 1* the Y-SNP M269 was analyzed using a TaqMan™ predesigned assay (Thermofisher Scientific, Wilmington, DE, USA), following the manufacturer's guidelines. Allelic discrimination analysis was performed with a 7000 Real-Time PCR System (Applied Biosystems, Foster City, CA, USA).

3.6 Agarose gel electrophoresis

Amplification success of DNA samples and the performance of the PCR reaction was assessed by agarose gel electrophoresis. PCR products were migrated in 1.5% agarose gels in TBE 1x solution with GelRed (Biotium, Fremont, CA, USA) at 100 V for 30 minutes and visualized with UV light in an UVIdoc gel documentation system (Uvitec, Cambridge, UK).

3.7 DNA sequencing

3.7.1 PCR product purification

Nucleotide and primer excess from the PCR amplification were cleaned enzymatically by adding 0.28 U of exonuclease I (exo) (Takara Bio, Kusatsu, Japan) and 0.72 U of shrimp alkaline phosphatase (SAP) (Takara Bio, Kusatsu, Japan) to 2 µl of PCR product and incubated for 45 min at 37 °C followed by 15 min at 80 °C.

3.7.2 Sequencing reaction

The sequencing reactions were carried out using BigDye Terminator v3.1 Cycle Sequencing Kit (Thermofisher Scientific, Wilmington, DE, USA) under the following conditions: 5.5 µl of Milli-Q water, 0.7 µl of BDT, 0.25 µl of the corresponding primer at 10 µM and 3.5 µl of PCR product in a final volume of 10 µl. The sequencing reaction consisted on 1 min at 96 °C, 25 cycles at 96 °C for 10 sec, 5 sec at 55 °C and 1.15 min at 60 °C in a C1000™ Thermal Cycler (Bio-Rad, Hercules, CA, USA).

3.7.3 Sequencing product purification

In order to eliminate the unincorporated BigDye terminators and salts the purification was performed using BigDye Xterminator™ Purification Kit (Thermofisher Scientific, Wilmington, DE, USA) in the following conditions: 22.5 µl of SAM solution, 5 µl of Xterminator solution and 10 µl of sequencing product. The samples were mixed for 45 min at 2,000 rpm in darkness and then centrifuged for 4 min at 2,000 rpm.

3.7.4 Capillary electrophoresis and data analysis

Sequencing products were analyzed by mixing 5 µl of Hi-Di™ Formamide (Thermofisher Scientific, Wilmington, DE, USA) and 5 µl of purified product. Capillary electrophoresis was conducted on an ABI PRISM 3130 genetic Analyzer (Thermofisher Scientific, Wilmington, DE, USA) with POP-7® polymer and a capillary of 36 cm. The results were analyzed using the software Sequencing Analysis v.5.2 (Thermofisher Scientific, Wilmington, DE, USA) and ChromasPro v.1.5 (Technelysium Pty Ltd, Brisbane, Australia).

3.8 DNA pyrosequencing

The Y-SNP L617 was analyzed by pyrosequencing in *Study Number 2*. PCR amplification was carried out using Qiagen HotStart Master Mix Kit (Qiagen, Hilden, Germany), 7.5 µM of biotinylated primer, 15 µM of nonbiotinylated primer and 1 ng of DNA. Pyrosequencing was carried out using the SQA kit (Biotage, Uppsala, Sweden) on a PSQ 96MA Pyrosequencer (Biotage, Uppsala, Sweden). Data interpretation was performed using the PyroMark Q96 MD Software (Qiagen, Hilden, Germany).

The applied protocol for the above-mentioned method is described below:

- Turn-on the Pyromark Q96MD.
- Turn-on the thermoblock at 80 °C.
- Fill the cuvettes of the vacuum platform with 100 ml of the buffers in the following order:
  - Cuvette 1: EtOH 70%
  - Cuvette 2: Denaturing solution (NaOH 2 M)
  - Cuvette 3: Pyromark Wash Buffer (Qiagen, Hilden, Germany) 1/10 dilution
  - Cuvette 4: Milli-Q water
- For the Binding Buffer mix (BBM) for each sample add 40 µl of Binding Buffer, 26 µl of Milli-Q water and 2 µl of sepharose.

- Add 12 µl of amplified product (diluted to 1/2) to the BBM and mix it for 20 minutes at 700 rpm in an orbital shaker.
- For the Annealing Buffer (ABM) for each sample mix, 11.64 µl of Annealing Buffer and 0.36 µl of primer at 10 mM were mixed and added to a blank plate.
- The amplicon cleaning was performed in the vacuum workstation (Qiagen, Hilden, Germany) in the following way:
    - With the vacuum switch on, lower the vacuum tool into the wells of the PCR plate for 15 sec to capture the beads with the amplification product.
    - With vacuum on, flush the tool with 70% EtOH (cuvette 1) for 5 sec.
    - With vacuum on, flush the tool with Denaturation Solution (cuvette 2) for 5 sec.
    - With vacuum on, flush the tool with Wash Buffer (cuvette 3) for 10 sec.
    - Align the vacuum tool with the blank plate that contains the ABM and switch the vacuum off.
    - Lower the vacuum tool into the wells and gently shake from side to side to release the beads.
    - Put the plate with the ABM and the clean amplicon in the thermoblock at 80 °C for 2 min.
    - Remove the plate and let it cool down at room temperature for 4-5 min.
    - With vacuum on, clean the tool with Milli-Q water (cuvette 4) and flush the filter probes.
- Prepare the reagents of the pirosequencing reaction by adding the corresponding quantities of enzyme, substratum and nucleotides in the Holder Box, with care of no leaving any air bubbles.
- Prior to the startup of the pyrosequencer a dispensation test was made.
- If the dispensation test is correct, introduce the pyrosequencing plate and start the machine.

3.9 Development of multiplex systems for the analysis of Y-SNPs and Y-STRs

3.9.1 Marker selection

3.9.1.1 Study Number 4

The selected Y-SNPs correspond to the diagnostic positions that determine the main subhaplogroups of the lineage R1b-DF27 and some sublineages of R1b-M269 above DF27. The

markers were chosen from the updated versions of the Y chromosome phylogenetic tree, as detailed in section 3.4.

3.9.1.2 Study Number 5

Y-STR markers were selected from the extensive study of Ballantyne and colleagues [151]. The main criteria for the marker selection were a low mutation rate (~ 10^{-4} mutations/generation) [151], and a gene diversity generally > 0.4 according to the data reported in the literature [126,146,157,158,315–317].

3.9.2 Primer design and optimization

Amplification primer design was performed as described in section 3.5.1. For both multiplex panels, final optimal concentrations for each primer in the reaction mix were adjusted in line with the different electropherogram intensities.

The miniprimers detailed in *Study Number 4* were designed manually, and in order to assure the separation of the extension primers during capillary electrophoresis their lengths were adjusted by adding tails of neutral sequence on the 5' end [318]. Amplification fragments differed in 5 bp to allow a clear electrophoretic separation.

The site-directed mutagenesis primers for DF19 and L881 in *Study Number 4* were designed manually by inserting the necessary nucleotide to produce the derived variant in the primer sequence [319]. The changed nucleotide was placed as close as possible to the 5' extreme of the primer in order to prevent primer-DNA hybridation problems due to the mismatch produced by the changed nucleotide.

The forward primers detailed in *Study Number 5* were modified by the addition of a fluorescent dye at their 5' end: 5-FAM (Abs. = 495 nm; Em. = 520 nm), YAKIMA YELLOW (Abs. = 530 nm; Em. = 549 nm), ATTO 550 (Abs. = 554 nm; Em. = 576 nm) and ATTO 565 (Abs. = 563 nm; Em. = 596 nm). The selected markers were distributed in the multiplex by the expected amplicon length, using a four-dye chemistry. The design was designed as an open system, where other markers of interest could be easily added along the four-dye layout in order to complement the multiplex if needed.

3.9.3 Multiplex PCR

PCR reaction of the Y-SNP 15-plex panel (*Study Number 4*) consisted of 5 µl of Qiagen Multiplex PCR Kit (Qiagen, Hilden, Germany), 1 µl of 10x primer mix, 3 µl of sterile Milli-Q water, and 1 ng of DNA in a final volume of 10 µl. Amplification was carried out in a C1000™ Thermal Cycler (Bio-Rad, Hercules, CA, USA) under the following conditions: 95 °C for 15 min; 3 cycles at 95 °C for 30 sec,

63 °C for 45 sec, and 72 °C for 30 sec; 15 cycles at 95 °C for 30 sec, 63 °C for 45 sec (with decrements of 0.2 °C per cycle) and 72 °C for 30 sec, 20 cycles at 95 °C for 30 sec, 60 °C for 45 sec, and 72 °C for 30 sec; and a final extension of 7 min at 72 °C.

The multiplex amplification for the 6-plex Y-STR panel (*Study Number 5*) consisted of 5 µl of Qiagen Multiplex PCR kit (Qiagen, Hilden, Germany), 0.5 µl of primer mix (final concentration of 0.2 µM), 1 ng of DNA, and Milli-Q water for a final reaction volume of 10 µl. The PCR was performed in a GeneAmp 9800 Thermal Cycler (Thermofisher Scientific, Wilmingtong, DE, USA) under the following cycling conditions: an initial denaturation at 95 °C for 15 min, followed by 30 cycles of 94 °C for 30 sec, 65 °C for 90 sec, 72 °C for 90 sec, and a final extension at 72 °C for 10 min.

Amplification success was assessed as described in section 3.6. From this point on the workflow follows as described in Figure 16 for each panel.

3.9.4 Minisequencing reaction

The multiplex minisequencing PCR contained 2 µl of SNaPshot™ Multiplex Kit reaction mix (Thermofisher Scientific, Wilmington, DE, USA), 0.7 µl of 10x minisequencing primer mix, 3.3 µl of sterile Milli-Q water, and 1 µl of purified multiplex PCR product (see section 3.7.1), in a final volume of 7 µl. The PCR was carried out in a C1000™ Thermal Cycler (Bio-Rad, Hercules, CA, USA) under the following conditions: 25 cycles at 96 °C for 10 sec; 50 °C for 5 sec; and 60 °C for 30 sec.

3.9.5 Minisequencing purification

For the elimination of the remaining dideoxinucleotides, the minisequencing products were purified adding 0.75 U of SAP (Takara Bio, Kusatsu, Japan) to 2 µl of minisequencing product and incubated for 60 min at 37 °C, followed by 15 min at 80 °C.

3.9.6 Capillary electrophoresis

The Y-SNP minisequencing products (*Study Number 4*) were analyzed by mixing 9.75 µl of Hi-Di™ Formamide (Thermofisher Scientific, Wilmington, DE, USA), 0.25 µl of GeneScan™ 120LIZ (Thermofisher Scientific, Wilmington, DE, USA) and 5 µl of purified product, and then denatured at 96 °C for 6 min.

The Y-STR amplification products (*Study Number 5*) were analyzed mixing 1 µl of PCR product with 9 µl of Hi-Di™ Formamide (Thermofisher Scientific, Wilmington, DE, USA) and 0.5 µl of GeneScan™ 500LIZ (Thermofisher Scientific, Wilmington, DE, USA).

For both panels, capillary electrophoresis was carried out as detailed in section 3.7.4. Data interpretation was performed with GeneMapper ID v.4.0 software (Thermofisher Scientific, Wilmington, DE, USA).

3.9.7 Reproducibility

The reproducibility of both multiplex panels (*Studies Number 4* and *5*) was validated by the analysis of negative controls and DNA samples several times, by different researches, on different days, and in different thermal cyclers. Afterwards, the mobility of the peaks of the different allelic variants in the electropherograms, both Y-SNPs and Y-STRs, was compared.

3.9.8 Sensitivity and stability assays

In order to evaluate the robustness of the novel Y-STR panel (Study Number 5) sensibility and stability studies were conducted. Sensitivity indicates the ability to obtain reliable results from a range of DNA quantities that allows to determine the upper and lower limits of detection of the assay [320]. To evaluate the minimum quantity of DNA required to obtain reliable results, that is, complete profiles, human control DNA 2800M (Promega® Corporation, Madison, WI, USA) was analyzed in triplicate in the following amounts of DNA: 10, 1.6, 1, 0.4, 0.2, 0.1, 0.05 and 0.025 ng.

Stability states the ability to obtain reliable results from DNA recovered from biological samples deposited on various substrates and subjected to various environmental and chemical insults [320]. To evaluate the stability of the panel presented in *Study Number 5*, 1 ng of human control DNA 2800M (Promega® Corporation, Madison, WI, USA) was analyzed in duplicate in the presence of two common inhibitors in forensic casework: humic acid and haematin (Sigma-Aldrich Corporation, St. Louis, MO, USA). The study was performed using the following concentrations of inhibitors: 3,000, 2,000, 1,000, 500, 300, 250, 200, 100, 50 and 25 ng/µl.

3.10 Statistical analyses

3.10.1 Population genetic parameters

The absolute and relative Y-SNPs frequencies were estimated by direct counting (*Studies Number 1, 2 and 3*). In *Study Number 3*, individuals with partial genetic information were present. In order to estimate the probabilities of each individual with missing genotypes to belong to each possible subhaplogroup, information corresponding to full genotypes over R1b, S116 or Z195 was used applying the formulas below.

Let *a* be the absolute frequency of haplogroup M269 (xS116) in a sample of *n* Y chromosomes; similarly, let *b*: S116 (xDF27), *c*: DF27 (xZ195), *d*: Z195 (xL176.2, xZ220), *e*: L176.2 (xM167), *f*: M167, *g*: Z220 (xZ278), *h*: Z278 (xM153), and *i*: M153. Let *s*=*a*+*b*+*c*+...+*i*. We have three types of samples with partial information: R1b-M269 without further subtyping (let its frequency be *j*), S116 (xU152, xM529, xZ195), but not typed for DF27 (call it *k*), and Z195 (xZ220), not typed for L176.2 (*l*). *j* individuals may belong to any of the *a*,...,*i* subhaplogroups with probability *a/s*,...,*i/s*; *k* can be DF27 (xZ195) with probability *c/(b+c)*, and Z195 (xZ220, xM167) can be either Z195 (xL176.2, xZ220) with probability *d/(d+e)* or L176.2 (xM167) with probability *e/(d+e)*. Combining these probabilities and turning them into estimated relative frequencies (which we denote with a circumflex over each letter), we have

$$\hat{c} = \frac{c\left(1 + \frac{j}{s} + \frac{k}{b+c}\right)}{n}$$

$$\hat{d} = \frac{d\left(1 + \frac{j}{s} + \frac{l}{d+e}\right)}{n}$$

$$\hat{e} = \frac{e\left(1 + \frac{j}{s} + \frac{l}{d+e}\right)}{n}$$

$$\hat{f} = \frac{f\left(1 + \frac{j}{s}\right)}{n}$$

$$\hat{g} = \frac{g\left(1 + \frac{j}{s}\right)}{n}$$

$$\hat{h} = \frac{h\left(1 + \frac{j}{s}\right)}{n}$$

$$\hat{\imath} = \frac{i\left(1 + \frac{j}{s}\right)}{n}$$

The frequencies corresponding to Y chromosome haplogroups were represented in contour maps using the software SURFER v.12 (Golden Software, Golden, CO, USA) by the krigging method.

Allele and haplotype frequencies for Y-STRs were calculated by mere counting using the software Arlequin v.3.5 [321].

3.10.2 Forensic parameters

Genetic diversity (GD), defined as the probability that two randomly chosen haplotypes are different in the sample, was calculated using the software Arlequin v.3.5 [321]. The formula applied is described below.

$$\widehat{H} = \frac{n}{n-1}(1 - \sum_{i=1}^{k} p_i^2)$$

$$V(\widehat{H}) = \frac{2}{n-1}\left\{2(-2)\left[\sum_{i=1}^{k} p_i^3 - (\sum_{i=1}^{k} p_i^2)^2\right] + \sum_{i=1}^{k} p_i^2 - (\sum_{i=1}^{k} p_i^2)^2\right\}$$

where *n* is the number of gene copies in the sample, *k* is the number of haplotypes, and p_i is the sample frequency of the *ith* haplotype [322].

The discrimination capacity (DC), that is, the number of different haplotypes observed in a given population, was calculated by dividing the number of different haplotypes by the total number of individuals in the population.

3.10.3 Population differentiation

The pairwise genetic distances F_{ST} and R_{ST}, and the corresponding significance *p* values between the analyzed populations were calculated with Arlequin v.3.5 [321](10,000 permutations). This method is based on estimating the proportion of genetic variation found within and between populations, since it allows quantitatively comparing the difference between different populations. The significant *p* values were adjusted with the sequential Bonferroni correction (α) [323] in order to account for potential Type I errors due to the multiple comparisons performed. Values below α represent genetic heterogeneity, and higher values represent the absence of genetic heterogeneity. The formula applied is described below.

$$\alpha = \frac{0.05}{\left[\frac{(n-1)}{2}\right] * n}$$

where *n* is the number of populations.

To obtain a representation of the pairwise genetic distances, 2D- and 3D-nonmetric multidimensional scaling (NMDS) analysis were performed using the software PAST v.3.04 [324], and the x-y-z coordinates were represented using the *rgl* package (http://cran.r-project.org/package=rgl) for R software [325].

The genetic structure of the different populations was studied by analysis of molecular variance (AMOVA) with Arlequin software v.3.5 [321] and factorial correspondence analysis (FCA) with Genetix v.4.05.2 [326]. In order to quantify patterns of population structure principal component analysis (PCA) was carried out with the software IBM SPSS Statistics v.19 (IBM Corporation, Armonk, NY, USA). Additionally, spatial genetic patterns were studied trough spatial PCA (sPCA) in *Study Number 1*, implemented using the algorithm provided in the R software package *adegenet* (http://adegenet.r-forge.r-project.org/) [327,328]. This method calculates components based on the genetic variance between populations and their spatial autocorrelation. The most informative components are those with the absolute highest eigenvalues. For instance, the most positive are associated with positive spatial autocorrelation (global structure), and the most negative are associated with negative spatial autocorrelation (local structure). A global structure implies that each sampling location is genetically closer to its neighbors than randomly chosen locations, as occurs with spatial groups, clines or intermediate states. In contrast, a stronger genetic differentiation among neighbors than among random pairs of populations characterizes a local structure.

3.10.4 Phylogenetic relationships

Phylogenetic analysis allows to understand the evolutionary history and the relationships between lineages, as well as between different populations. These relationships can be ascertained thanks to phylogenetic reconstruction methods, which evaluate the heritable traits analyzed. Phylogenetic relationships were calculated through the Median Joining algorithm (MJ network) using the software Network v.5.0 [329] in *Studies Number 1* and *2*. This software builds all the shortest and simplest phylogenetic trees possible for the analyzed individuals.

3.10.5 TMRCA estimation

The time to the most recent common ancestor (TMRCA) of different haplogroups was estimated with the algorithms Rho (ρ), implemented within Network software v.5.0 [330], a weighted version of Rho computed with an *ad hoc* R script (http://github.com/fcalafell/weighted_rho), and the average square distance (ASD) by using the Kilin-Klyosov TMRCA calculator [331]. For the calculations the following mutation rates were considered:

- The evolutionary mutation rate (EMR) of 6.9 x 10^{-4} locus/25 years established by Zhivotovsky and colleagues [310] (*Study Number 1*).
- A mean germ line mutation rate (GMR) of 1.37 x 10^{-3} per locus per generation for the Y-STRs considered in the calculation obtained from the YHRD (https://yhrd.org) [223], and a

generation time of 30 years, which translates into a rate of 728 years/mutation (*Studies Number 2* and *3*)

- A global Y chromosome mutation rate of 0.888 x 10^{-9} per year [165,332] taking into account the 10.36 Mb of the Y chromosome amenable to short read sequencing and SNP detection [165], which translates into a rate of 108.7 years/mutation (*Study Number 3*).

The Rho statistic (ρ) [302,333] is defined as the average number of mutations *l* along *m* unique haplotypes sampled from *n* individuals. Each line stems from a defined ancestral node given a resolved gene tree. Under the hypothesis that the phylogenetic tree under consideration is correct:

$$\rho = (n_1 l_1 + n_2 l_2 + \cdots + n_m l_m)/n$$

with variance:

$$\sigma^2 = (n_1^2 l_1 + n_2^2 l_2 + \cdots + n_m^2 l_m)/n^2$$

The weighted version of ρ, ρ_w, leverages on the relatively precise knowledge of the mutation rate of each Y-STR. Thus, it considers that mutations at slow mutating STRs take longer to accumulate than mutations at faster mutating STRs. It is defined as:

$$\rho_w = \frac{1}{N}\sum_{i=1}^{k} n_i \left(\sum_{j=1}^{S} \left(|X_{ji} - X_{jm}| \right) \cdot \frac{\bar{\mu}}{\mu_j} \right)$$

where *N* is the number of chromosomes, *k* is the number of different haplotypes, n_i is the absolute frequency of the *ith* haplotype, *S* is the number of different STRs, X_{ji} is the allelic state of the *ith* haplotype at the *jth* STR, X_{jm} is the median allele at the *jth* STR, $\bar{\mu}$ is the average mutation rate and μ_j is the mutation rate of the *jth* STR. The standard deviation of ρ_w is given by:

$$sd(\rho_w) = \frac{1}{N}\sqrt{\sum_{i=1}^{k} n_i^2 \left(\sum_{j=1}^{S} \left(|X_{ji} - X_{jm}| \right) \cdot \frac{\bar{\mu}}{\mu_j} \right)}$$

and age, as in ref. [333], is estimated as:

$$T = \rho_w \cdot \bar{\mu}$$

where $\bar{\mu}$ is now expressed in years per mutation.

The average square distance (ASD) [334,335] is defined as:

$$ASD = V_A + V_B + (\mu_A - \mu_B)^2$$

where V_A, V_B, μ_A and μ_B are the variances and means, respectively, of allele size in populations A and B. In order to remove the dependence of ASD on population size and to decrease its variance the following distance, based on the Stepwise Mutation Model, was introduced [334,335]:

$$ASD\ (\delta\mu)^2 = (\mu_A - \mu_B)^2$$

3.10.6 Demographic model evaluation

In order to test alternative demographic models and to estimate their parameters approximate Bayesian computing (ABC) was performed in *Study Number 3*. ABC allows to deal with mathematical models where likelihood calculations have failed, and since it is a likelihood-free inference, the algorithms sample from the posterior distribution of the parameters by finding values that yield simulated data sufficiently resembling the observed data [336].

One million simulations were run with the software fastsimcoal2 [337,338], either with a constant population size (drawn from a lognormal distribution between 100 and 100,000), or with an exponential growth that started *Tstart* generations ago. In the growth model, the effective population sizes before (*Na*) and at the end (*Nc*) of the growth were drawn in the same way of the constant model and conditioned to *Na* < *Nc*. *Na* refers to a time *Tstart* drawn from a uniform distribution between 50 and 350 generations. Y-STR mutation rates were taken as fixed given the high precision with which they are known, but the value of the geometric parameter for the Generalized Stepwise Mutation model was sampled from a uniform distribution with limits (0; 0.8). To summarize the data, the mean and the standard deviation over loci of four statistics were calculated: the number of different haplotypes (*K*), the haplotype diversity (*H*), the allelic range, and the Garza- Williamson's index. Posterior probabilities of the models were calculated by means of the simple rejection algorithm [339] as well as of the weighted multinomial logistic regression [340], evaluating different thresholds for both methods to check the stability of the results as in Vai and colleagues [341]. For parameter estimation, Euclidian distance was calculated between the simulated and observed summary statistics, and retained the 5% of the total simulations corresponding to the shortest distances. Posterior probability for each parameter was estimated using a weighted local regression [342], after a logtan transformation [343].

4. Results

4.1 Study Number 1

'New clues to the evolutionary history of the main European paternal lineage M269: dissection of the Y-SNP S116 in Atlantic Europe and Iberia'

The present study corresponds to the attainment of the objective 1: *To analyze R1b-M269 paternal lineage and its current sublineages in populations of Atlantic Europe and the Iberian Peninsula, which will allow to define in detail the distribution of M269 in Southwest Europe and to obtain new clues about its evolutionary history.*

As mentioned in the following study, the present genetic makeup of Europe is the result of several human migrations and settlements influenced primarily by climate, cultural development and historical conquests of territory. The origin and expansion of the paternal lineage R1b-M269, the most common haplogroup in West Europe, has been the subject of great controversy. Some authors placed its origin in the Paleolithic, during the post-glacial expansion from the Franco-Cantabrian refuge or from Eastern Europe. Other authors, on the other hand, shifted the origin of M269 in Eastern Europe from the Paleolithic to the Neolithic due to the younger coalescent times obtained by applying germinal mutation rates rather than evolutionary mutation rates. In order to address this controversy and to complete the distribution of M269 along Southern Europe and the Atlantic coast this study offers the deepest dissection of S116, one of the main derived subhaplogroups of M269.

In this study, a total of 1,560 individuals from the Iberian Peninsula (Galicia, Asturias, Cantabria, Basque Country, Barcelona, Alicante, Andalucía, Madrid, Portugal) and Atlantic Europe (Brittany, Ireland, Denmark) were genotyped for the Y-SNPs M269, L11, U106, S116, U162, M529, DF27, DF19, and L238 by TaqMan™ assays or by High Resolution Melting (HRM). Pairwise F_{ST} values between the populations studied were calculated and Median Joining Networks, spatial principal component analysis (sPCA), and maps of haplogroup distribution were also constructed.

Our results reveal that the M269 derived branch S116 displays frequency peaks and a spatial distribution that differs from the one previously proposed, which peaks in the Upper Danube basin and Paris. In contrast, our study shows that S116 displays the highest frequencies along the Atlantic coastline and the British Isles. Apart from that, the dissection of S116 in its sublineages revealed an outstanding frequency of DF27 in the Iberian Peninsula, which is located in a different geographic area than the rest of the S116 sublineages and displays a restricted distribution pattern. Finally, the frequency distribution of M269 subhaplogroups, the absence of sublineages over S116 in the Franco-Cantabrian area, the homogeneity of Y-STR diversity within M269 in

Europe, and the origin of new sublineages such as L11 on the migration wave of M269 support an origin of M269 in Eastern Europe, with the appearance of its sublineages, like S116, during the advance of the lineage through West Europe.

In sum, the importance of continuing the dissection of M269 lineage in different European populations is reflected in the present research, since the discovery and study of new sublineages have provided new clues in the distribution and origin of M269. Until the release of this article, despite the number of previous studies conducted on this haplogroup, there was a still a need to collect more population data of Southwest European populations to properly characterize the distribution of the sublineages of M269.

This study has resulted in an international publication in the journal *European Journal of Human Genetics* under the heading '*New clues to the evolutionary history of the main European paternal lineage M269: dissection of the Y-SNP S116 in Atlantic Europe and Iberia*' in March 2016. Q1, IP: 4.287. The publication is shown below.

European Journal of Human Genetics (2016) 24, 437–441

www.nature.com/ejhg

Article

New clues to the evolutionary history of the main European paternal lineage M269: dissection of the Y-SNP S116 in Atlantic Europe and Iberia

Laura Valverde[1,a], M. José Illescas[1,a], Patricia Villaescusa[1], Amparo M. Gotor[1], Ainara García[1], Sergio Cardoso[1], Jaime Algorta[2,3], Susana Catarino[2], Karen Rouault[4], Claude Férec[4], Orla Hardiman[5], Maite Zarrabeitia[6], Susana Jiménez[7], M. Fátima Pinheiro[8], Begoña M. Jarreta[9], Jill Olofsson[10], Niels Morling[10], Marian M. de Pancorbo[1,*]

[1]BIOMICs Research Group, Lascaray Research Center, University of the Basque Country UPV/EHU, Vitoria-Gasteiz, Spain.

[2]Progenika Biopharma SA (a Grifols company), Bizkaia Technology Park, Derio, Spain.

[3]Department of Molecular Biology, Faculty of Science and Technology, University of the Basque Country UPV/EHU, Bilbao, Spain.

[4]Inserm UMR1078, Génétique, Génomique fonctionnelle et Biotechnologies, Brest, France.

[5]National Neuroscience Centre, Beaumont Hospital, Dublin, Ireland.

[6]Forensic and Legal Medicine Area, Department of Physiology and Pharmacology, University of Cantabria, Cantabria, Spain.

[7]Forensic Medicine Division, Department of Pathology and Surgery, University Miguel Hernandez, Elche, Alicante, Spain.

[8]Forensic Genetics Department, National Institute of Legal Medicine and Forensic Sciences, Porto, Portugal.

[9]Laboratory of Genetics and Genetic Identification, Department of Pharmacology, University of Zaragoza, Zaragoza, Spain.

[10]Section of Forensic Genetics, Department of Forensic Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark.

[*]Correspondence author

[11]These authors contributed equally to this work.

Received 17 January 2015; revised 19 April 2015; accepted 29 April 2015; published online 17 June 2015.

Abstract

The dissection of S116 in more than 1500 individuals from Atlantic Europe and the Iberian Peninsula has provided important clues about the controversial evolutionary history of M269. First, the results do not point to an origin of M269 in the Franco-Cantabrian refuge, owing to the lack of sublineage diversity within M269, which supports the new theories proposing its origin in Eastern Europe. Second, S116 shows frequency peaks and spatial distribution that differ from those previously proposed, indicating an origin farther west, and it also shows a high frequency in

the Atlantic coastline. Third, an outstanding frequency of the DF27 sublineage has been found in Iberia, with a restricted distribution pattern inside this peninsula and a frequency maximum in the area of the Franco-Cantabrian refuge. This entire panorama indicates an old arrival of M269 into Western Europe, because it has generated at least two episodes of expansion in the Franco-Cantabrian area. This study demonstrates the importance of continuing the dissection of the M269 lineage in different European populations because the discovery and study of new sublineages can adjust or even completely revise the theories about European peopling, as has been the case for the place of origin of M269.

Introduction

The current genetic makeup of Europe is the result of many population migrations and settlements influenced principally by climate, cultural progress and the historical conquests of territory.[1,2] The genetic evidence provided by the analysis of the Y chromosome (Ychr), which is a valuable tool for the study of the evolution of the paternal lineages because of its uniparental mode of inheritance, has revealed that a large majority of the individuals currently in Central and Western Europe (40–90%) belong to a single lineage, R-M269.[2,3] The lineage M269 has its maximum frequency in the Franco-Cantabrian area, and it shows a cline of decreasing frequency with distance. This has led to numerous theories about the role of the Franco-Cantabrian region in European genetic history.

To date, the most widely accepted theories have argued that this pattern of frequencies may be the result of origin in, and subsequent postglacial expansion from, the Franco-Cantabrian refuge.[2–4] Another theory, based on the variance of Y-STR haplotypes within M269, also supports its postglacial expansion but argues that M269 could have had a parallel expansion from a refuge in Eastern Europe (Anatolia).[5] The new theory proposed by Balaresque *et al*,[6] based on the higher diversity of Y-STR haplotypes in Eastern European M269 individuals than in Western European ones, concludes that there is a single origin for haplogroup M269 in Eastern Europe. In addition, the Balaresque *et al*[6] theory shifts the origin from the glacial period to the Neolithic, because they apply germinal mutation rates rather than evolutionary, generating younger coalescence times.

The arrival of M269 from Eastern Europe proposed by Balaresque *et al*[6] has been strongly refuted by Busby *et al*.[7] Busby et al recalculated the diversity of Y-STRs haplotypes within M269 in a larger and geographically broader sample, indicating not higher diversity in Eastern Europe but a homogeneous background of microsatellite variation in the whole European sample.[7]

The dissection of haplogroup M269 has shown a wide range of European areas possessing geographically located subhaplogroup expansions,[8,9] which provides useful information for reconstructing the phylogeographic history of this lineage. However, the study of these sublineages, far from helping to find a consensus about the origin, growth and history of this great lineage, has increased the controversy. Myres *et al*[9] analysed M269 and the sublineages M412, L11, U106, S116, U152 and M529. The obtained coalescence times and frequency distribution patterns led them to conclude that the current distribution of M269 sublineages is owing to allele surfing at the periphery of the westwards expansion of M269. Therefore, Myres *et al*[9] proposed the origin of M269 in Eastern Europe, similar to Balaresque *et al*,[6] but earlier during the Mesolithic period.

Finally, a different theory, not supported to date, argues that M269 entered the Iberian Peninsula in the late Neolithic and that its subhaplogroups S116 and M529 would appear during the expansion of the Bell Beakers northwards.[10]

This multi-sided debate affects not only European paternal lineages but also maternal lineages. In principle, the task of inferring the evolutionary histories of paternal lineages is actually more complicated than that of the maternal lineages, because the increased size and complexity of the Ychr makes the development of comprehensive and complete time-scaled phylogenetic trees more arduous than for mitochondrial DNA (mtDNA). In addition, mtDNA has more information in aDNA.

However, despite this, there is currently a major controversy about the origin and expansion of maternal haplogroup H, which shares a similar pattern of frequencies with paternal haplogroup R, and for which a similar and contemporaneous history has been suggested[3] (Supplementary Box 1).

The controversy cannot be more interesting. Efforts to unravel the evolutionary history of the most frequent haplogroups in Europe have generated a cordial and productive discussion about new calculation methods and new approaches for the study of these haplogroups and sub-haplogroups.

Our study goes deeply into the study of the M269 sublineages of the European Atlantic coast and the Iberian Peninsula. This territory has a high frequency of the still-unresolved paragroup S116* (×U152, ×M529) (data from[7,9]). Therefore, this study offers the deepest analysis of haplogroup S116 made to date in Europe. These new data, as well as their comparison when possible with previous Ychr and mtDNA data, resolve important questions and offer novel clues about the

evolutionary history of M269, in addition to finding new sublineages with important and restricted geographic locations.

Materials and methods

A total of 1560 healthy, unrelated males from the Iberian Peninsula (Galicia, Asturias, Cantabria, Basque Country, Barcelona, Alicante, Andalucía, Madrid, Portugal) and Atlantic Europe (Brittany (Brest), Ireland, Denmark) were studied (Supplementary Table S1). The Y-SNPs M269, L11, U106, S116, U152, M529, DF27, DF19 and L238 were analysed by TaqMan assays (Applied Biosystems, Carlsbad, CA, USA) or by High Resolution Melting Technology (for further details see Supplementary Box 2 and Supplementary Table S2). Individuals from Basque Country were also genotyped for a set of 17 Y-STR loci using the AmpFlSTR YfilerTM kit (Applied Biosystems).

Maps of haplogroup frequency distribution were constructed using the Surfer Golden software v 10.0.500 (Golden Software, Golden, CO, USA) by the kriging method. The spatial genetic patterns were studied through spatial principal component analyses (sPCAs) using the R software package adegenet (R Foundation for Statistical Computing, Vienna, Austria; http://adegenet. r-forge.r-project.org/). Genetic distances (Fst) between populations based on haplogroup frequencies were calculated using the Arlequin v 3.1 (University of Bern, Bern, Switzerland) software and plotted in Multidimensional Scaling graphs using the PAST software (University of Oslo, Oslo, Norway). The phylogenetic relationships of Y-STR haplotypes were estimated by median joining networks using NETWORK v 4.5.1.6 (Fluxus Technology Ltd., Kiel, Germany). Higher phylogenetic weight was allocated to the loci with lower mutation rate,[11,12] lower variance (VL, Kayser *et al*[13]) and higher linearity (D, Busby *et al*[7]; calculated with the actual range published in YHRD, Willuweit *et al*[14]; Supplementary Box 4). Coalescent times were estimated using the Network software and the evolutionary mutation rate 6.9 x 10^{-4}/locus/25 years, established by Zhivotovsky *et al*[15] and confirmed by Shi *et al*[16] for the set of YSTRs analysed here. Further details about statistical treatment can be found in the Supplementary Box 2.

Data generated in this study can be accessed in Supplementary Tables S1. The Basque Y-STR–Y-SNP haplotype data have been uploaded to the public database YHRD under accession numbers YA003672-77, YA003718 and YA004063.[14]

Results and discussion

The Y-SNPs M269, L11, U106, S116, U152, M529, L238, DF19 and DF27 were analysed in 1560 individuals from 12 different populations from the Atlantic Coast and the Iberian Peninsula (Supplementary Tables S1).

Surprisingly, the inclusion of new populations from the Atlantic Coast and Iberia in this study has identified a frequency distribution of haplogroup S116 that differs from the previously proposed distribution. Myres *et al*[9] proposed a frequency peak in the Upper Danube Basin and Paris, with declining frequency towards Italy, Iberia, southern France and British Isles. By contrast, these new data show maximum frequencies in northern Iberia, the western coast of France and the British Isles, raising questions about the possible expansion of this lineage during the early Neolithic LBK culture (Linearbandkeramik or Linear Pottery culture), as proposed by Myres *et al*.[9]

Supplementary Figure S1 shows distribution maps that compile all of the frequency data for M269 sublineages published to date (more than 16 000 male individuals;[7,9,17] present study) but at a lower level of resolution than that achieved in the current study. From the maps, it can be appreciated that M269 sublineages show distinct areas of distribution in Europe: U106 is distributed in the countries of Central- Northern Europe, and S116 occurs in Western and South-western Europe. With regard to the sublineages of S116, U152 is more common in northern Italy and the Alpine region, whereas M529 is more common in the British Isles and Brittany. However, there is a large percentage of S116 individuals unassigned to any of these sublineages, described here as paragroup S116* (×U152, ×M529). The frequency of this paragroup reaches approximately 50% in the Iberian Peninsula and exceeds 80% in the Basque region. It has also been observed in the area of Brittany and the British Isles, but the frequencies there do not exceed 20%.

The dissection analysis of S116 has provided very informative results for further completing the history of M269. The paragroup S116* (× U152, × M529) has been largely resolved owing to the discovery of the highly frequent sublineage DF27 in the Iberian Peninsula. DF27 has a frequency of 40–48% in Iberia but reaches frequencies over 60% in the Franco-Cantabrian region, particularly in the Basque population. However, outside the Iberian Peninsula, the frequency is below 20% (Supplementary Figure S2 and Supplementary Table S1). Thus, the sublineage S116-DF27 is located in a different geographic area than that occupied by the other S116 sublineages M529 and U152 (Supplementary Figure S1).

The DF19 and L238 sublineages show very low frequencies in Western Europe. The DF19 sublineage was not detected in any individuals, and L238 was detected only in one individual from Brest (Brittany) (Supplementary Table S1).

The new population data highlight the high frequencies of M529 found in Brest (>50%) (Supplementary Figure S1), outside the British Isles, which may raise doubts about whether it originated in the European continent or in the British Isles.

The sublineage U152 shows a striking distribution in the Iberian Peninsula (Supplementary Figure S1), where frequency peaks appear in the coastal corners in the SW (southern Portugal, 13%), NW (Galicia, Asturias, 8%) and NE (Barcelona, Alicante, 6%), and the minimum lies in the Basque region (2%). In Europe, haplogroup U152 has its maximum in the Alpine region, and thus perhaps its frequency pattern could be explained by a migration from the Alpine region of origin to the Iberian Peninsula, along the coast, avoiding areas historically known to have remained more isolated, as is the case with Basque Country.

After analysing all five sublineages so far known for S116 (U152, M529, L238, DF19 and DF27), some individuals remained who did not belong to any of these five sublineages, and they were assigned as belonging to the new, more concise S116* paragroup (× U152, × M529, × L238, × DF19, × DF27) (hereafter called S116*). The maximum frequency of S116* has been found in Irish (17%) and Basque (12%) populations. In both populations, the vast majority of individuals belonging to the S116 haplogroup belong to their respective M529 or DF27 sublineage, and those who do not belong to either of these sublineages belong almost entirely to paragroup S116* (Supplementary Table S1). Only the discovery of new Y-SNPs will determine whether these individuals can be assigned to new sublineages, which may be identical or different between Ireland and Basque country, providing more clues about the genetic relationship and evolution between the two populations.

One of the main reasons leading to the proposal of the hypothesis of origin and/or expansion of M269 from the Franco-Cantabrian refuge is its maximum frequency and pattern of decreasing frequency with increasing distance from this area. The Basque population is located in the heart of the refuge area, and our results indicate that almost all of their M269 lineages belong to sublineage S116 (Basque Country; M269-82%; S116-80%, Supplementary Table S1). If M269 had originated in this area, it would seem logical to find higher variability of M269 sublineages, such as M269xL11, L11 or U106*. Thus, the dissection of M269 in the refuge area raises questions about its origin in this region. Unfortunately, the homogeneity in the variability of Y-STRs within M269 makes it impossible to pinpoint a more likely origin,[7] but the frequency distribution of M269 sublineages in the European continent suggests an origin in the East with a subsequent migration westwards, with the appearance of its sublineages during the advance of the migration wave.[9]

However, the Basque region has maximum frequencies of S116 and its sublineages S116* and DF27, the latter showing a decreasing gradient with distance. Meanwhile, M529 and U152 frequencies are extremely low. This may indicate that this region is a source for S116 and its sublineage DF27. Myres *et al*[9] proposed the Upper Danube basin and Paris area as the geographic

sources of S116. The patterns of frequencies obtained here also suggest that S116 emerged on the crest of the wave of migration but somewhere closer to the Franco-Cantabrian region. Thus, a possible evolutionary scenario of these lineages may be chronologically as shown in Figure 1: (1) origin of M269 in Eastern Europe; (2) origin of L11 on the wave of the westward advance of M269;[9] and (3) colonization of the entire continent by L11, as evidenced by the high frequency of L11* in different parts of the Atlantic coast, from the Baltic to the southern coast of Portugal (data from[7,9]) (L11 origin has been hypothesized in the map in Northern Europe); (4) origin of U106 from L11 individuals who inhabited the southern coast of the North Sea; (5) origin of S116 from L11 individuals inhabiting the Eastern Cantabrian coast, that is, the area of the Franco-Cantabrian refuge; and (6) origin of the DF27 sublineage from S116 individuals inhabiting the refuge area, while other S116 individuals spread to the rest of Iberia and Europe along the Atlantic and Mediterranean coasts, originating M529 and U152, respectively. Subsequently, (7) the U152, M529 and DF27 subtypes spread and came to occupy their present territories, with U152 and M529 re-entering the Iberian Peninsula (Figure 1). U152 and M529 may have re-entered the Iberian Peninsula during one of the numerous subsequent migrations to Iberia, during either Neolithic or historical times, that is, with the arrival of Phoenicians, Carthaginians, Romans, Goths or Vikings.[18]

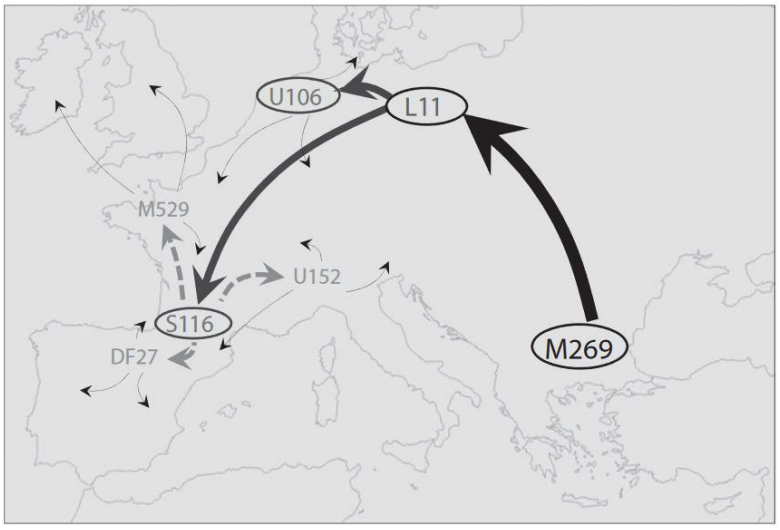


Figure 1 Evolutionary proposal for sublineages of M269 in Europe. Arrows start at the most likely places of origin and indicate the direction of expansion. The older the movement, the thicker the arrow. The thinner arrows indicate the current distribution of the younger sublineages here studied.

To delve into the phylogenetic structure of S116 and DF27 haplogroups, a median joining network was performed with 15 Y-STR haplotypes of only M269 Basque native individuals (Supplementary Figure S3 and Supplementary Table S3). Thus, the study of the potentially ancient lineages that have inhabited the Franco–Cantabrian region until today is intended (Supplementary Box 3).

The phylogeny was constructed following carefully selected settings (Supplementary Box 4). The network showed a bipartite structure with two main groups corresponding to the individuals belonging to the S116* and DF27 haplogroups (Supplementary Figure S3). In addition, haplogroup DF27 appears to be split into two parts owing to the presence of two different haplotypes in the Y-STRs, DYS437/ DYS448 (Supplementary Figure S3). Both Y-STRs have low mutation rates, and they are therefore more robust in distinguishing Y-chr haplogroups or established phylogenetic splits within haplogroups. DYS448, aside from being the Y-STR with lower mutation rate[11,12] and a small variance V_L[19] in the Basque population, has a long hexanucleotide repeat unit, which gives even higher phylogenetic weight.[19] This may indicate the presence of different sub-haplogroups within DF27 in the Basque population, indicating that continuing the dissection of DF27 may contribute new information regarding the evolutionary history of this region.

Finally, a proper mutation rate was carefully selected for calculating TMRCAs, although the authors are aware of the lack of a definitive time scale for the Ychr; therefore, these calculations remain merely indicative. The classical mutation rate 6.9×10^{-4}/locus/25 years, established initially by Zhivotovsky *et al*,[15] was finally selected for being calibrated based on well-dated historical events and because its proper operation has been re-evaluated afterwards for the set of YSTRs analysed in this study.[16] Concretely, Shi *et al*[16] compared, in a very comprehensive study including a large panel of worldwide samples, the human male demographic inferences obtained with three different mutational rates: an observed mutation rate from the mutations counts in father–son pairs, the classical evolutionary mutation rate[15] and a recalibrated evolutionary mutation rate (rEMR) corrected for the differences in variance of different sets of YSTRs. For the set of YSTRs analysed here, the evolutionary mutation rate and the rEMR were equivalent. Shi *et al*[16] concluded that the rEMR provided the most comprehensive demographic inferences according to previous studies and actual geographical distributions.

The obtained coalescence times date the origin of haplogroup S116 in the native Basque region 11 673 ± 1962 ybp, and the origin of DF27 soon after, 10 468 ± 1831 ybp, which would place their origins after the last cold period of the Younger Dryas, that is, the early Holocene warm period, when weather conditions reached the current temperatures during the course of a few decades, encouraging population growth and expansion.[3]

These phylogenies and dates were confirmed also including non-native individuals in the network analysis, which allowed reaching identical conclusions (Supplementary Figure S4).

The spatial genetic patterns of the different haplogroups were deeply studied through sPCAs. Supplementary Figure S5 shows the sPCAs including the population data analysed here and data

compiled from the literature (and therefore at the same low level of resolution as the distribution maps of Supplementary Figure S1). The analysis detected four spatial patterns that explain most of the variance related to M529, S116*, U106 and U152 (Supplementary Figure S6). By increasing the level of resolution of the sPCAs, including only the more resolved Western European data from this study, a new spatial pattern was detected for DF27 in Iberia (Supplementary Figures S7–S9). Interestingly, the analysis finds strong affinity among all Iberian populations, with the exception of the Basque population, which shows little affinity with the populations both outside and inside the peninsula but appears to participate in the distribution patterns affecting both of those populations. This may indicate that the Basque country has been involved in the history of the different haplogroups that principally characterize both Western European regions (M529 and DF27), and this would support the previously proposed scenario (Figure 1), in which, first, S116 expands outside the refuge and, second, U152 and M529 originate outside the peninsula and DF27 inside.

MDS representation of Fst genetic distances between populations, calculated based on haplogroup frequencies, shows results consistent with those obtained in the sPCAs (Supplementary Figure S10, Supplementary Table S4).

In summary, this study provides new genetic evidence indicating the absence of diversity of M269 lineages over S116 in the current population of what once was the refuge, the maximum frequencies of S116, S116* and DF27 in the refuge area and their spatial distributions in Iberia and Western European coast. This is in addition to the evidence from previous studies: the homogeneity in Y-STR diversity within M269 in Europe[7] and the emergence of new sublineages such as L11 on the wave of the advance of M269 into Western Europe[9] consistent with the scenario proposed in Figure 1.

This scenario proposes an origin in the East for M269, in contrast to the classical theories.[2,3] The controversy in calculating TMRCAs makes it impossible to reliably date these evolutionary episodes, at least until the more complete Ychr allows more accurate time scales and/or until genotyped and firmly dated archaeological remains become available.

However, the authors believe that it is unlikely that an arrival to Europe of M269 during the Neolithic period has generated such a complex scenario of expansions for its sublineages, especially when genetic evidence of cultural diffusion has been found for Ychr in Anatolia[20,21] and for mtDNA in the refuge.[22] Thus, the spread of Neolithic culture would mean a lower demic movement. The theories that argue for an origin in the East and during the Neolithic period

assume a rapid expansion of M269 throughout Europe, replacing most of the previously settled haplogroups, which would be compatible with a main scenario of demic diffusion.

The scenario proposed here would be most compatible with an arrival of M269 from the East occurring in Palaeolithic times. The Wurm glaciation had numerous ups and downs in temperature that would have led to the existence of multiple glacial refugia, which has been proposed both for mtDNA and Ychr.[5,23] Improved weather conditions would allow colonization of more northern territories from all refuges simultaneously. Similarly, the mtDNA-H and Ychr-R lineages that evolved in the East from Palaeolithic times, could have expanded westwards during the Neolithic period, thereby mixing with other H and R lineages that arrived to Western Europe in Paleolithic times and evolved independently in these western territories. This may be one reason for the complexity of interpreting the results, in addition to the assumption that post-Neolithic movements may be masking and confounding the oldest traces.

In this context, the genetic evidence found for the sister haplogroup of M269 in the maternal line, haplogroup H, has been helpful for complementing and giving clues about M269 history (Supplementary Box 5).

In sum, this study demonstrates the importance of continuing the dissection of the M269 lineage in different European populations because the discovery and study of new sublineages can adjust or even completely rewrite the theories about European peopling, as has been the case with the place of origin of M269. Similarly, the future availability of complete sequences of the Ychr and of desirable Palaeolithic aDNA data may definitively reveal the complete and true history of this major lineage.

Conflict of interest

JA and SC are employed by the commercial company Progenika Biopharma SA (Grifols, Derio, Spain). This does not alter the authors' adherence to all the policies on sharing data and materials. The remaining authors declare no conflict of interest.

Acknowledgements

Funds were provided by the Basque Government (IT-424-07). LV received a postdoctoral grant from the Basque Government (DKR-2012-439), applied through the Ikerbasque Foundation. The authors are deeply indebted to Emmanuelle Génin, Dan Bradley, Kevin Kenna, the Basque Foundation of Science (BIOEF), the Spanish National DNA Bank, SGIker (UPV/EHU, MICINN, GV/EJ, ERDF and ESF) and to all the people who voluntarily participated in this study.

Author contributions

LV and MMP conceived the study. JA, SC, KR, CF, OH, SJ, MFP, BMJ, JO, NM and MMP contributed samples/materials. LV, MJI, PV, AMG, AG and JO analysed the samples. LV analysed the data. LV and MMP wrote the manuscript. All co-authors reviewed the manuscript before submission.

References

1 Pinhasi R, Thomas MG, Hofreiter M, Currat M, Burger J: The genetic history of Europeans. Trends Genet 2012; 28: 496–505.

2 Semino O, Passarino G, Oefner PJ et al: The genetic legacy of Paleolithic Homo sapiens sapiens in extant Europeans: a Y chromosome perspective. Science 2000; 290: 1155–1159.

3 Soares P, Achilli A, Semino O et al: The archaeogenetics of Europe. Curr Biol 2010; 4: R174–R183.

4 Cardoso S, Valverde L, Alfonso-Sánchez MA et al: The expanded mtDNA phylogeny of the Franco-Cantabrian region upholds the pre-neolithic genetic substrate of Basques. PLoS One 2013; 8: e67835.

5 Cinnioğlu C, King R, Kivisild T et al: Excavating Y chromosome haplotype strata in Anatolia. Hum Genet 2004; 114: 127–148.

6 Balaresque P, Bowden GR, Adams SM et al: A predominantly neolithic origin for European paternal lineages. PLoS Biol 2010; 8: e1000285.

7 Busby GB, Brisighelli F, Sánchez-Diz P et al: The peopling of Europe and the cautionary tale of Y chromosome lineage R-M269. Proc Biol Sci 2012; 279: 884–892.

8 Cruciani F, Trombetta B, Antonelli C et al: Strong intra- and inter-continental differentiation revealed by Y chromosome SNPs M269, U106 and U152. Forensic Sci Int Genet 2011; 5: e49–e52.

9 Myres NM, Rootsi S, Lin AA et al: A major Y-chromosome haplogroup R1b Holocene era founder effect in Central and Western Europe. Eur J Hum Genet 2011; 19: 95–101.

10 Klyosov A: Ancient history of the Arbins, bearers of haplogroup R1b, from central Asia to Europe, 16,000 to 1500 years before present. Advances in Anthropology 2012; 2: 87–105.

11 Goedbloed M, Vermeulen M, Fang RN et al: Comprehensive mutation analysis of 17 Y-chromosomal short tandem repeat polymorphisms included in the AmpFlSTR Yfiler PCR amplification kit. Int J Legal Med 2009; 123: 471–482.

12 Ballantyne KN, Goedbloed M, Fang O et al: Mutability of Y-chromosomal microsatellites: rates, characteristics, molecular bases, and forensic implications. Am J Hum Genet 2010; 87: 341–353.

13 Kayser M, Krawczak M, Excoffier L et al: An extensive analysis of Y-chromosomal microsatellite haplotypes in globally dispersed human populations. Am J Hum Genet 2001; 68: 990–1018.

14 Willuweit S, Roewer L, International Forensic Y Chromosome User Group: Y chromosome haplotype reference database (YHRD): update. Forensic Sci Int Genet 2007; 1: 83–87.

15 Zhivotovsky LA, Underhill PA, Cinnioğlu C et al: The effective mutation rate at Y chromosome short tandem repeats, with application to human population- divergence time. Am J Hum Genet 2004; 74: 50–61.

16 Shi W, Ayub Q, Vermeulen M et al: A worldwide survey of human male demographic history based on Y-SNP and Y-STR data from the HGDP-CEPH populations. Mol Biol Evol 2010; 27: 385–393.

17 Larmuseau MH, Vanderheyden N, Jacobs M, Coomans M, Larno L, Decorte R: Micro-geographic distribution of Y-chromosomal variation in the Central-Western European region Brabant. Forensic Sci Int Genet 2011; 5: 95–99.

18 Encarta Online Encyclopedia. Spain 2007, http://www.webcitation.org/5kwqnGivb.

19 Järve M, Zhivotovsky LA, Rootsi S et al: Decreased rate of evolution in Y chromosome STR loci of increased size of the repeat unit. PLoS One 2009; 4: e7276.

20 Mirabal S, Varljen T, Gayden T et al: Human Y-chromosome short tandem repeats: a tale of acculturation and migrations as mechanisms for the diffusion of agriculture in the Balkan Peninsula. Am J Phys Anthropol 2010; 142: 380–390.

21 Morelli L, Contu D, Santoni F, Whalen MB, Francalacci P, Cucca F: A comparison of Y-chromosome variation in Sardinia and Anatolia is more consistent with cultural rather than demic diffusion of agriculture. PLoS One 2010; 5: e10419.

22 Hervella M, Izagirre N, Alonso S et al: Ancient DNA from hunter-gatherer and farmer groups from northern Spain supports a random dispersion model for the Neolithic expansion into Europe. PLoS One 2012; 7: e34417.

23 Pala M, Olivieri A, Achilli A et al: Mitochondrial DNA signals of late glacial recolonization of Europe from near eastern refugia. Am J Hum Genet 2012; 90: 915–924.

Supplementary Information accompanies this paper on European Journal of Human Genetics website (http://www.nature.com/ejhg).

Electronic Supplementary Material

Box 1: Actual controversy about the origin and expansion of maternal haplogroup H (hg H)

The most accepted theories for mitochondrial DNA (mtDNA) haplogroup H support its origin in the Franco-Cantabrian refuge and its postglacial expansion [1].

However, there has recently been much controversy with the new adjustments of the mitochondrial time-scales based on the information from complete mitochondrial genomes. Soares *et al*. [2] have proposed the departure dates for the H1 and H3 subgroups from the refuge after the last cold period, the Younger Dryas, i.e., the early Mesolithic. Fu *et al*. [3] obtained a sharp increase in the population size of hg H at approximately 5,000-9,000 ybp, and they associated these results with an expansion during the Neolithic, based on previous analyses showing that Neolithic remains have a high frequency of hg H, while this hg was absent in pre-Neolithic remains, something that is no longer considered true since hg H has been detected in Upper Palaeolithic remains from the Franco-Cantabrian refuge [4]. Also interesting is that these dates would be Neolithic if the expansion is from Eastern Europe but Mesolithic if it is from Western Europe. In another possibility, Pala *et al*. [5] not only support the Palaeolithic origin of hg H in Western Europe, but their results point to a Palaeolithic entry age from east of the mtDNA haplogroups J and T, hitherto considered Neolithic.

Other studies have recalculated the mitochondrial timescale by analysing complete mitochondrial genomes of firmly dated Neolithic remains [6,7]. Both studies obtained much higher mutation rates than previously estimated, indicating that all events so far dated for mtDNA seem to be younger. Thus, these authors are more supportive of a Neolithic expansion for hg H. Nevertheless, the authors are aware that their results do not rule out the theory of the postglacial expansion of hg H, as some of them concluded the year before in a study of the Basque population [8], in which

the results suggested the presence of hg H in Basques since pre-Neolithic times, something currently confirmed by aDNA analysis in the Franco-Cantabrian refuge [4,9].

References

1. Torroni A, Bandelt HJ, D'Urbano L et al: mtDNA analysis reveals a major late Paleolithic population expansion from southwestern to northeastern Europe. Am J Hum Genet 1998; 62: 1137-1152.
2. Soares P, Achilli A, Semino O et al: The archaeogenetics of Europe. Curr Biol 2010; 4: R174-183.
3. Fu Q, Rudan P, Pääbo S, Krause J: Complete mitochondrial genomes reveal neolithic expansion into Europe. PLoS One 2012; 7: e32473.
4. Hervella M, Izagirre N, Alonso S, Fregel R, Alonso A, Cabrera VM, de la Rúa C: Ancient DNA from hunter-gatherer and farmer groups from northern Spain supports a random dispersion model for the Neolithic expansion into Europe. PLoS One 2012; 7: e34417.
5. Pala M, Olivieri A, Achilli A et al: Mitochondrial DNA signals of late glacial recolonization of Europe from near eastern refugia. Am J Hum Genet 2012; 90: 915-924.
6. Fu Q, Mittnik A, Johnson PL et al: A revised timescale for human evolution based on ancient mitochondrial genomes. Curr Biol 2013; 23: 553-559.
7. Brotherton P, Haak W, Templeton J et al: Neolithic mitochondrial haplogroup H genomes and the genetic origins of Europeans. Nat Commun 2013; 4: 1764.
8. Behar DM, Harmant C, Manry J et al: The Basque paradigm: genetic evidence of a maternal continuity in the Franco-Cantabrian region since pre-Neolithic times. Am J Hum Genet 2012; 90: 486-493.
9. Lacan M, Keyser C, Crubézy E, Ludes B: Ancestry of modern Europeans: contributions of ancient DNA. Cell Mol Life Sci 2013; 70: 2473-2487.

Box 2: Materials & Methods

Population

A total of 1,560 healthy, unrelated males from the Iberian Peninsula (Galicia, Asturias, Cantabria, Basque Country, Barcelona, Alicante, Andalucía, Madrid, Portugal) and Atlantic Europe (Brittany (Brest), Ireland, Denmark) were studied (Table S1). All participants provided written informed consent. The procedures were in accordance with the ethical principles of the Helsinki Declaration of 1975, as revised in 2000.

Y-SNP analysis

The Y-SNP M269 was analysed using a TaqMan® predesigned assay (Applied Biosystems) for rs9786153, following the manufacturer's guidelines. Allelic discrimination analysis was performed with a 7000 Real-Time PCR System (Applied Biosystems).

The Y-SNPs L11, U106, S116, U152, M529, DF27, DF19 and L238 [1,2] were analysed by High Resolution Melting. YSNP characteristics and the primers used for the amplification of each Y-SNP are shown in Table S2. Y-SNPs were amplified with 0.5 μL of each primer (1 μM), 2.5 μL of SsoFast EvaGreen Supermix (BioRad) and 1 ng of DNA in a final volume of 5 μL. Amplification and melting were done in a C1000 thermocycler equipped with a CFX96 optic module (BioRad) under the following conditions: 98°C 10 sec; 35 cycles at 98°C 5 sec, corresponding annealing temperature (see Table S2) 20 sec; 95°C 30 sec, 60°C 2 min and finally the melting cycle from 65°C to 95°C with an increase of 0.2°C/sec, for detecting the different allelic variants. Data interpretation was performed using Precision Melt Analysis software (BioRad). Only high-quality amplification and melting curves with a cluster assignment over 95% of confidence were considered. The assignment of the corresponding allelic variants of every cluster was performed by using positive and negative controls previously detected by sequencing.

Danish males were typed for M269, S116 and U106 using custom-designed TaqMan Assays (Thermo Fisher) and analysed on a 7900HT Fast Real-Time PCR System (Thermo Fisher).

Problematic samples were reanalysed or sequenced when necessary.

Amplifications for the sequencing of each Y-SNP were done with 2.5 μL of KAPA2GTM Fast HotStart Ready Mix (2X) (KAPA Biosystems), 0.5 μL of each primer at 1 μM (see Table S2) and 1 ng of DNA in a final volume of 5 μL. Amplification was conducted with the same conditions of amplification as previously described (without the melting cycle) in a C1000 thermocycler (BioRad). Sequencing reactions were carried out with BigDye Terminator v.3.1 Cycle Sequencing Kit (Applied Biosystems) following the manufacturer's guidelines.

Y-STR analysis

Individuals from Basque Country were genotyped for a set of 17 Y-STR loci using the AmpFlSTR®YfilerTM kit (Applied Biosystems), following the recommendations of the manufacturer. Capillary electrophoresis took place in an ABI Prism 3130 Genetic Analyser, and fragment sizes were assigned using GeneMapper® v. 4.0 software. The nomenclature used is that of the latest recommendations for the DNA Commission of the International Society of Forensic Genetics, except for locus Y GATA H4, which was named on the basis of the allelic ladder supplied with the AmpFlSTR ® YfilerTM kit.

Data analysis

The maps of haplogroup frequency distribution were constructed using Surfer Golden Software v. 10.0.500 by the kriging method.

The spatial genetic patterns were studied through spatial principal component analyses (sPCA), implemented using the algorithm provided in the R software package *adegenet* [3-6]. This method calculates the components based on the genetic variance between populations and their spatial autocorrelation. The components can be positive or negative. The most informative components are those with the absolute highest eigenvalues, i.e., the most positive (associated with positive spatial autocorrelation, global structure) and the most negative (associated with negative spatial autocorrelation, local structure). A global structure implies that each sampling location is genetically closer to its neighbours than randomly chosen locations, as occurs with spatial groups, clines or intermediate states. In contrast, a stronger genetic differentiation among neighbours than among random pairs of populations characterizes a local structure.

Genetic distances (Fst) between populations based on haplogroup frequencies were calculated with the Arlequin v 3.1 software [7] with 10,000 permutations. They were plotted in Multidimensional Scaling graphs using PAST software [8].

The phylogenetic relationships of Y-STR haplotypes were estimated by median joining networks using NETWORK v 4.5.1.6 [9]. Higher phylogenetic weight was allocated to the loci with lower mutation rate [10,11], lower variance [V_L, 12] and higher linearity [D, 13; calculated with the actual range published in YHRD, 14; Supplementary Box 4]. Coalescent times were estimated using Network software and the re-calibrated evolutionary STR mutation rate 6.9×10^{-4}/locus/25 years revised for this set of 17 Y-STRs [15,16].

References

1. Rocca RA, Magoon G, Reynolds DF, Krahn T, Tilroe VO, Op den Velde Boots PM, Grierson AJ: Discovery of Western European R1b1a2 Y chromosome variants in 1000 genomes project data: an online community approach. PLoS One 2012; 7: e41634.
2. International Society of Genetic Genealogy 2014. Y-DNA Haplogroup Tree 2014, Version: 9.71, Date: 21 July 2014, http://www.isogg.org/tree/.
3. Jombart T: adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 2008; 24: 1403-1405.
4. Jombart T, Devillard S, Dufour AB, Pontier D: Revealing cryptic spatial patterns in genetic variability by a new multivariate method. Heredity (Edinb) 2008; 101: 92-103.
5. Montano V, Ferri G, Marcari V, Batini C, Anyaele O, Destro-Bisol G, Comas D: The Bantu expansion revisited: a new analysis of Y chromosome variation in Central Western Africa. Mol Ecol 2011; 20: 2693-2708.
6. R Core Team: R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2013. URL http://www.R-project.org/.
7. Excoffier L, Laval G, Schneider S: Arlequin (version 3.0): an integrated software package for population genetics data analysis. Evol Bioinform Online 2007; 1: 47-50.

8. Hammer O, Harper DAT, Ryan PD: PAST: Paleontological statistics software package for education and data analysis. Palaeontologia Electronica 2001; 4: 9.
9. Bandelt HJ, Forster P, Röhl A: Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol 1999; 16: 37-48.
10. Goedbloed M, Vermeulen M, Fang RN et al: Comprehensive mutation analysis of 17 Y-chromosomal short tandem repeat polymorphisms included in the AmpFlSTR Yfiler PCR amplification kit. Int J Legal Med 2009; 123: 471-482.
11. Ballantyne KN, Goedbloed M, Fang O et al: Mutability of Y-chromosomal microsatellites: rates, characteristics, molecular bases, and forensic implications. Am J Hum Genet 2010; 87: 341-353.
12. Kayser M, Krawczak M, Excoffier L et al: An extensive analysis of Y-chromosomal microsatellite haplotypes in globally dispersed human populations. Am J Hum Genet 2001; 68: 990-1018.
13. Busby GB, Brisighelli F, Sánchez-Diz P et al: The peopling of Europe and the cautionary tale of Y chromosome lineage R-M269. Proc Biol Sci 2012; 279: 884-892.
14. Willuweit S, Roewer L: International Forensic Y Chromosome User Group. Y chromosome haplotype reference database (YHRD): update. Forensic Sci Int Genet 2007; 1: 83-87.
15. Zhivotovsky LA, Underhill PA, Cinnioğlu C et al: The effective mutation rate at Y chromosome short tandem repeats, with application to human population-divergence time. Am J Hum Genet 2004; 74: 50-61.
16. Shi W, Ayub Q, Vermeulen M et al: A worldwide survey of human male demographic history based on Y-SNP and Y-STR data from the HGDP-CEPH populations. Mol Biol Evol 2010; 27: 385-393.

Box 3: Special features of Basque population

The Basque region has been historically subjected to genetic isolation and is therefore a possible stronghold of potentially ancient lineages. Moreover, the native language of the Basque region allows the differentiation of indigenous people by their surnames [1], so in this regard, the Basque population provides a unique opportunity to explore the oldest Y-chromosome genetic substratum of the population. Individuals arrived in the Basque Country during the last century with the Industrial Revolution, who are non-native Basques, can be recognized because of their non-Basque surnames and then be removed from the population sample for statistical purposes [2,3].

Table S1 shows how the haplogroup frequencies would vary depending on whether the total Basque population (native and non-native) or only the native population is assessed. The frequencies of S116 and DF27 are slightly higher in the native population than in the total population. However, in both cases, the frequencies of these haplogroups are the highest in Europe, which supports the evolutionary inferences made about DF27 and S116 in this population.

However, the strong isolation of the Basque territory could also have caused a loss in the diversity of lineages, due to phenomena such as genetic drift or bottlenecks between the populations of the valleys of the complex Basque orography. However, the analysis has shown that haplogroup frequencies of the populations adjacent to Basque Country show a logical continuation of the pattern of frequencies, which suggests that, indeed, S116 and DF27 are ancient lineages that originated in this region.

Although the native Basque population offers a special opportunity for studying potential ancient lineages from the Franco-Cantabrian refuge (and has therefore been used for various phylogenetic approaches), in the statistical analysis, including population comparisons, the total Basque population sample (native and non-native Basques) has been assessed. This has been done because individual selection based on autochthonous surnames is difficult or impossible to do in other European populations, so the comparison between actual populations and only the native individuals of the Basque population would introduce bias into statistical calculations.

References

1. Valverde L, Rosique M, Köhnemann S et al: Y-STR variation in the Basque diaspora in the Western USA: evolutionary and forensic perspectives. Int J Legal Med 2012; 126: 293-298.
2. Peña JA, Garcia-Obregon S, Perez-Miranda AM, De Pancorbo MM, Alfonso-Sanchez MA: Gene flow in the Iberian Peninsula determined from Y-chromosome STR loci. Am J Hum Biol 2006; 18: 532-539.
3. Valverde L, Köhnemann S, Rosique M, Cardoso S, Zarrabeitia M, Pfeiffer H, de Pancorbo MM: 17 Y-STR haplotype data for a population sample of Residents in the Basque Country. Forensic Sci Int Genet 2012; 6: e109-111.

Box 4: Settings for constructing Y-STR phylogeny by median joining networks

For constructing the phylogeny, native Basque individuals were selected. Moreover, the properties of the Y-STRs were carefully assessed. Busby *et al*. [1] warned that the attributes of the Y-STRs are rarely considered in phylogenetic reconstructions and calculations of TMRCA, altering the precision of the results. Here, we have prioritised the Y-STRs with higher phylogenetic weight. Thus, DYS385 and DYS389b were discarded for lacking phylogenetic interest, because it is not possible to assign a specific allele to each locus of DYS385 with the genotyping method used here and because DYS389b has a very complex and repetitive structure and may then have several allelic variants [2,3]. In contrast, we gave the highest phylogenetic weight to Y-STRs DYS390, DYS392, DYS393, DYS437, DYS438 and DYS448 for having the lower mutation rates of the Y-STRs analysed [4,5]. The Y-STRs DYS19, DYS391 and DYS635 have higher mutation rates than the

previous, but the same phylogenetic weight was applied because they have a very low variance V_L in Basque population [6], and DYS635 also has a high linearity D [1], which gives them greater phylogenetic weight in this population. Finally, a minimum weight of 1 was given to the Y-STRs DYS389I, DYS439, DYS456, DYS458 and GATA H4 because they have much higher mutation rates.

References

1. Busby GB, Brisighelli F, Sánchez-Diz P et al: The peopling of Europe and the cautionary tale of Y chromosome lineage R-M269. Proc Biol Sci 2012; 279: 884-892.
2. Forster P, Röhl A, Lünnemann P, Brinkmann C, Zerjal T, Tyler-Smith C, Brinkmann B: A short tandem repeat-based phylogeny for the human Y chromosome. Am J Hum Genet 2000; 67: 182-196.
3. Zhivotovsky LA, Underhill PA, Cinnioğlu C et al: The effective mutation rate at Y chromosome short tandem repeats, with application to human population-divergence time. Am J Hum Genet 2004; 74: 50-61.
4. Goedbloed M, Vermeulen M, Fang RN et al: Comprehensive mutation analysis of 17 Y-chromosomal short tandem repeat polymorphisms included in the AmpFlSTR Yfiler PCR amplification kit. Int J Legal Med 2009; 123: 471-482.
5. Ballantyne KN, Goedbloed M, Fang O et al: Mutability of Y-chromosomal microsatellites: rates, characteristics, molecular bases, and forensic implications. Am J Hum Genet 2010; 87: 341-353.
6. Kayser M, Krawczak M, Excoffier L et al: An extensive analysis of Y-chromosomal microsatellite haplotypes in globally dispersed human populations. Am J Hum Genet 2001; 68: 990-1018.

Box 5: Inferences combining genetic evidence found for Ychr hg R-M269 and mtDNA hg H.

The genetic evidence found for the sister haplogroup of M269 in the maternal line, hg H, could be helpful for giving clues about M269 history, although with cautiousness because non-contemporaneous histories have also been proposed for these haplogroups [1]. However, it seems reasonable to consider that both haplogroups have coexisted at some point in their long evolutionary trees. For example, our results and other published data would allow for the coexistence of paternal S116 or DF27 haplogroups and maternal H1 and H3 in the Franco-Cantabrian refuge [2,3].

Recently, Brotherton *et al*. [4] have shown that some subgroups of hg H seem to have different geographical locations in Europe. This differential distribution for subhaplogroups emulates the distribution of subgroups of hg R, although it is too soon to determine whether some of them share a geographic location.

Moreover, Brotherton *et al*. [4] analysed remains from the early, mid and late Neolithic (ENE, MNE and LNE, respectively), concluding that the remains from the ENE show genetic discontinuity with

MNE/LNE remains. In fact, the authors report similarities between ENE remains and current populations from Eastern Europe and between MNE/LNE remains and current Central/SW Europe, respectively. This east-west genetic discontinuity could be interpreted as demic diffusion not reaching the western part of the continent. That is, the presence of hg H in Palaeolithic remains of the Franco-Cantabrian refuge would indicate the arrival of this haplogroup in Western Europe before the Neolithic. The Neolithic wave could bring early farmers belonging to subgroups of hg H that evolved independently in the East and different to those present in Europe in pre-Neolithic times. The demic diffusion would have been short in expansion because it was soon superseded by cultural diffusion. Thus, the Western European Palaeolithic populations were neolithised mainly by culture diffusion, and now the genetic substrate mainly present in Western and Central Europe would correspond with the Palaeolithic genetic substrate.

Brotherton *et al*. [4] relates the dominant maternal gene pool of current Western Europe with the expansion of the Neolithic culture Bell Beaker from Iberia in the LNE, as Klyosov [5] does for Ychr. We consider that this would also be consistent with the scenario proposed here. Bell Beaker Culture is believed to have emerged from the megalithic cultures, and it is believed that the Atlantic megalithic cultures arose from the ancient inhabitants of the European Atlantic coast [6-7]. The apogee of the megalithism has been linked to the arrival of new models of social organization or even to newcomers, which produced a sense of territoriality in the original inhabitants. This led them to build huge stone monuments. The ancient clans of hunter-gatherer-fishers, who inhabited the Atlantic coast from the Upper Palaeolithic, were spread across the Portuguese coast, Cantabrian Sea, western and northern coast of Europe, islands and even Baltic Sea coast. It is believed that they were the source of megalithic cultures. These ancient individuals could be carriers of L11 lineages. Evidence of this could be the actual maximum frequencies of L11* in these same Atlantic territories. This would imply a genetic continuity in SW Europe from Palaeolithic times, with a minor influence of Neolithic lineages arrived from the East.

The dates of origin and expansion of the U106 and S116 subtypes originated from these L11 individuals remain uncertain. Our calculations, which were made including all precautions reported so far, point to an origin and expansion at the beginning of the Holocene, as suggested previously by Myres *et al*. [8] and Soares *et al*. [2] for mtDNA. This would make sense because the improved weather conditions would have led to a large enough population explosion to allow its expansion and generation of new variability, illustrated by the M529, U152 and DF27 sublineages.

Thus, the presence in the network of undiscovered DF27 variability suggests that there may exist still more expansion events and unknown histories.

References

1. Boattini A, Martinez-Cruz B, Sarno S et al: Uniparental markers in Italy reveal a sex-biased genetic structure and different historical strata. PLoS One 2013; 8: e65441.
2. Soares P, Achilli A, Semino O et al: The archaeogenetics of Europe. Curr Biol 2010; 4: R174-183.
3. Cardoso S, Valverde L, Alfonso-Sánchez MA et al: The expanded mtDNA phylogeny of the Franco-Cantabrian region upholds the pre-neolithic genetic substrate of Basques. PLoS One 2013; 8: e67835.
4. Brotherton P, Haak W, Templeton J et al: Neolithic mitochondrial haplogroup H genomes and the genetic origins of Europeans. Nat Commun 2013; 4: 1764.
5. Klyosov A: Ancient history of the Arbins, bearers of haplogroup R1b, from central Asia to Europe, 16,000 to 1500 years before present. Advances in Anthropology 2012; 2: 87-105.
6. Fernández-Martínez VM: Prehistoria. El largo camino de la humanidad. Alianza Editorial, Madrid, 2007.
7. Barandiarán I, Martí B, Del Rincón MA, Maya JL: Prehistoria de la Península Ibérica. Editorial Ariel, Barcelona, Spain, 2012.
8. Myres NM, Rootsi S, Lin AA et al: A major Y-chromosome haplogroup R1b Holocene era founder effect in Central and Western Europe. Eur J Hum Genet 2011; 19: 95-101.

Supplementary Tables (collected in the Supplementary Excel File)

Table S1. Y-SNP frequencies (%) in the analysed samples of population. For each haplogroup/column, the higher the frequency, the more intense the colour. Below, detailed characteristics of the Basque population sample.

	xM269	M269	U106	S116	U152	M529	L238	DF19	DF27	L11[a]	S116*	N
Northern Europe, Atlantic Coast												
Denmark	62,64	37,36	17,82	16,67	·	·	·	·	·	·	·	174
Western Europe, Atlantic Coast												
Ireland	18,49	81,51	6,16	74,66	2,05	54,11	0,00	0,00	0,68	81,51	17,81	146
Brest (Brittany, France)	13,10	86,90	4,14	80,69	4,14	52,41	0,69	0,00	17,24	·	6,21	145
Iberian Peninsula, Portugal												
Portugal	37,27	62,73	2,73	50,91	3,64	2,73	0,00	0,00	40,91	·	3,64	110
Iberian Peninsula, Spain												
Alicante	35,34	64,66	4,31	56,90	6,03	0,00	0,00	0,00	43,10	·	7,76	116
Andalucía	37,00	63,00	3,00	59,00	4,00	0,00	0,00	0,00	46,00	·	9,00	100
Asturias	42,86	57,14	0,00	57,14	7,94	6,35	0,00	0,00	42,86	·	0,00	63
Barcelona	31,00	69,00	2,00	65,00	6,00	1,00	0,00	0,00	48,00	·	10,00	100
Cantabria	31,25	68,75	2,08	61,46	4,17	6,25	0,00	0,00	44,79	·	6,25	96
Galicia	38,57	61,43	4,29	55,71	8,57	7,14	0,00	0,00	40,00	·	0,00	70
Madrid	31,31	68,69	2,02	60,61	4,04	1,01	0,00	0,00	48,48	·	7,07	99
Basque Country	17,60	82,40	1,47	80,06	2,05	2,05	0,00	0,00	63,34	82,40	12,61	341

Details of the Basque sample of population

In relation to the place of sampling												
Rural Basques	7,77	92,23	0,52	91,71	2,59	1,55	0,00	0,00	71,50	92,23	16,06	193
Urban Basques	30,41	69,59	2,70	64,86	1,35	2,70	0,00	0,00	52,70	69,59	8,11	148
In relation to the Basque surnames												
Native Basques (with Basque surname, found both in rural and urban sampling)	7,83	92,17	1,30	90,87	2,17	2,17	0,00	0,00	70,87	92,17	15,65	230
Non native Basques (without Basque surname, found only in urban sampling)	37,84	62,16	1,80	57,66	1,80	1,80	0,00	0,00	47,75	62,16	6,31	111

[a]L11 was analyzed only in two populations for confirming it follows the last proposed phylogeny [1,2].

References

1. Rocca RA, Magoon G, Reynolds DF, Krahn T, Tilroe VO, Op den Velde Boots PM, Grierson AJ: Discovery of Western European R1b1a2 Y chromosome variants in 1000 genomes project data: an online community approach. PLoS One 2012; 7: e41634.
2. International Society of Genetic Genealogy 2014. Y-DNA Haplogroup Tree 2014, Version: 9.71, Date: 21 July 2014, http://www.isogg.org/tree/

Table S2. Y-SNP characteristics, primer sequences and analysis conditions.

Y-SNP	Primers (5´- 3´)	Method*	Annealing Temperature	Amplicon Size	Mutation (anc/der)	rs SNP ID	Y chr position GRCh37/hg19
M269	FW: ACATGGTATCACAATAGAAGGG RV: TTTCACCATGTTAGCCTGGA	Seq Seq	60.5	216	T/C	rs9786153	22739367
L11	FW: GTGTGATGTCTTTCTCCACC RV: GCAAGGATTGTCTCTTAGAACAG	HRM & Seq HRM & Seq	60.5	85	T/C	rs9786076	17844018
U106	FW: TTCCTGAATAGCAAATCCCA RV: GCTGTATGTGTCTTCCTGTG	HRM & Seq HRM & Seq	60.5	96	C/T	rs16981293	8796078
S116	FW: TCCTGCTAATGTATCTGCTG RV: CTCATTTATCACCTCAGTGC	HRM & Seq HRM & Seq	61	115	C/A	rs34276300	22157311
U152	FW: AGAAACATTCCACGCTTGAG RV: ATGGTAGTTTAATGGGAGTAGC	HRM & Seq HRM & Seq	60.5	103	G/A	rs1236440	15333149
M529	FW: TAAACCCTCCTCAGCAACAG RV: GGAAGCATTCAGAAGCAGGT	HRM & Seq HRM & Seq	60.5	150	C/G	rs11799226	15654428
L238	FW: AAGAAATGTCAACGGTACAGAG RV: CATACACATTCACAGCAGGT	HRM & Seq HRM & Seq	60.5	125	A/G	rs35199432	21253443
DF19	FW: AAAGGGCACTCTTGATAGGAC RV: TCCCTATTCGCCATCTTAGC	HRM & Seq HRM & Seq	60.5	95	G/T	_	14301499
DF27	FW*: TTGGCTGGATATGAAATTCTGGA RV*: GGAAGCCCATCAGATTAACAGA	HRM & Seq HRM & Seq	61	124	G/A	_	21380200

* These primers are nested primers of a longer amplicon amplified with the primers FW:GGGAATTTGATCCTGTCGTTG and RV: GAACAAAGCCTCCAAGAAATATGAGG [1] with the same annealing temperature.

** Seq: Sequencing, HRM: High Resolution Melting

References

1. Rocca RA, Magoon G, Reynolds DF, Krahn T, Tilroe VO, Op den Velde Boots PM, Grierson AJ: Discovery of Western European R1b1a2 Y chromosome variants in 1000 genomes project data: an online community approach. PLoS One 2012; 7: e41634.
2. van Oven M, Van Geystelen A, Kayser M, Decorte R, Larmuseau MH: Seeing the wood for the trees: a minimal reference phylogeny for the human Y chromosome. Hum Mutat 2014; 35(2):187-191. http://www.phylotree.org/Y/
3. International Society of Genetic Genealogy ISOGG 2014. Y-DNA Haplogroup Tree 2014, Version: 9.71, Date: 21 July 2014, http://www.isogg.org/tree/.
4. FTDNA. https://www.familytreedna.com/
5. Genographic Project. https://www.familytreedna.com/projects.aspx
6. NCBI dbSNP. http://www.ncbi.nlm.nih.gov/SNP/

Table S3. 17 YSTR-YSNP haplotype data from the Basque sample of population.

Corresponds to Attached Table 1 in Appendix section.

Table S4. Genetic Fst distances based on Y-SNP haplogroup frequencies (above diagonal) and p values (below diagonal). Statistically significant values after Bonferroni correction are shaded in blue.

	Madrid	Barcelona	Andalucía	Galicia	Asturias	Portugal	Brest	Ireland	Alicante	Cantabria	Basques
Madrid	*	-0.00896	-0.00673	0.00856	0.00817	-0.00084	0.24527	0.31059	-0.00442	-0.00699	0.03029
Barcelona	0.93832+-0.0025	*	-0.00657	0.00883	0.01040	0.00170	0.23246	0.29363	-0.00539	-0.00670	0.03109
Andalucía	0.75369+-0.0039	0.76299+-0.0039	*	0.00267	0.00195	-0.00614	0.24210	0.29736	-0.00867	-0.00516	0.05344
Galicia	0.15860+-0.0040	0.15177+-0.0040	0.26987+-0.0042	*	-0.01233	-0.00605	0.19592	0.25614	-0.00077	-0.00158	0.09160
Asturias	0.17295+-0.0040	0.14058+-0.0034	0.29819+-0.0045	0.88714+-0.0026	*	-0.00827	0.21871	0.27967	0.00066	0.00024	0.09597
Portugal	0.37462+-0.0055	0.28789+-0.0051	0.72102+-0.0040	0.64954+-0.0049	0.74349+-0.0045	*	0.23391	0.29142	-0.00548	-0.00355	0.07448
Brest	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	*	0.02754	0.23392	0.20177	0.29326
Ireland	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00297+-0.0005	*	0.28829	0.26648	0.37065
Alicante	0.61123+-0.0046	0.70755+-0.0045	0.97070+-0.0016	0.38828+-0.0048	0.33195+-0.0043	0.70399+-0.0045	0.00000+-0.0000	0.00000+-0.0000	*	-0.00437	0.05841
Cantabria	0.77339+-0.0042	0.77606+-0.0042	0.64182+-0.0050	0.41501+-0.0045	0.34125+-0.0047	0.53876+-0.0054	0.00000+-0.0000	0.00000+-0.0000	0.61845+-0.0047	*	0.04150
Basques	0.00347+-0.0006	0.00297+-0.0005	0.00000+-0.0000	0.00010+-0.0001	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00010+-0.0001	0.00099+-0.0003	*

Supplementary Figures

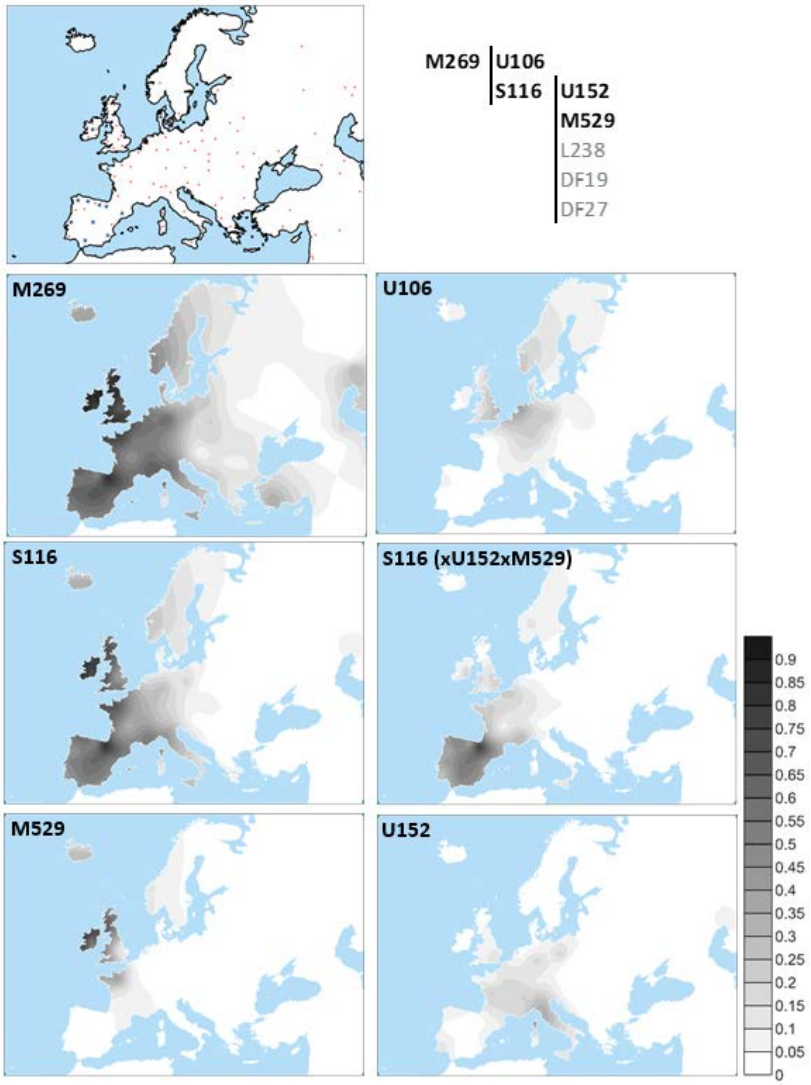


Fig. S1. Frequency distribution maps of the data compiled in this study (blue stars) and the data from Myres *et al*. (2011), Larmuseau *et al*. (2011) and Busby *et al*. (2012) (red points). This Fig. S1 represents the comparisons performed at a lower level of tree resolution than in Fig. S2 (exclusively data from present study), because no higher resolution data is available in the literature and a broadly geographical overview of European continent was intended in this 1st representation. The Y-SNPs used for the construction of these Fig. S1 maps are highlighted in bold in the upper right tree.

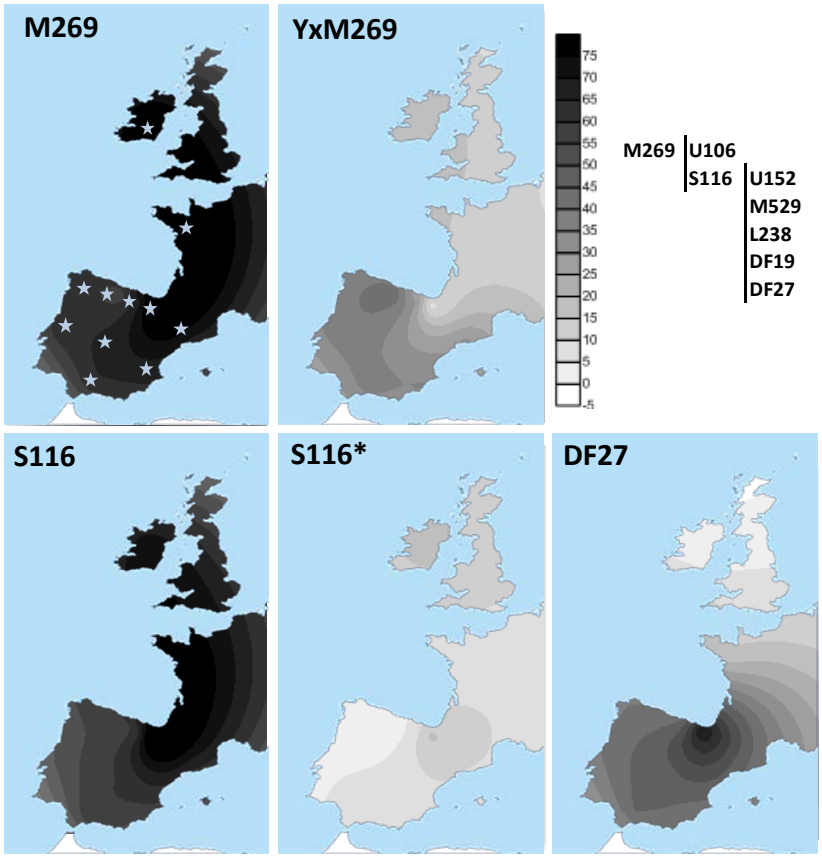


Fig. S2. Frequency distribution maps of M269, S116 and DF27 in the Atlantic Coast and Iberian Peninsula. The stars in M269 map indicate the samples of population analysed. The upper right tree includes the Y-SNPs used for constructing the distribution maps.

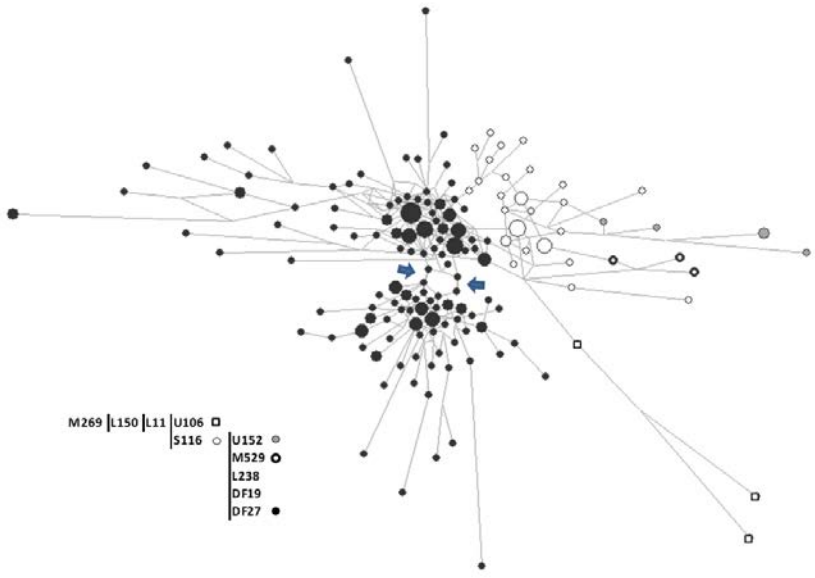


Fig. S3. Median joining network of the M269 haplogroup in the Basque native population (bearing Basque surnames). The blue arrows indicate a phylogenetic split of DF27 haplogroup into two groups bearing the alleles 14/18 and 15/19 in the Y-STR haplotype DYS437/DYS448.

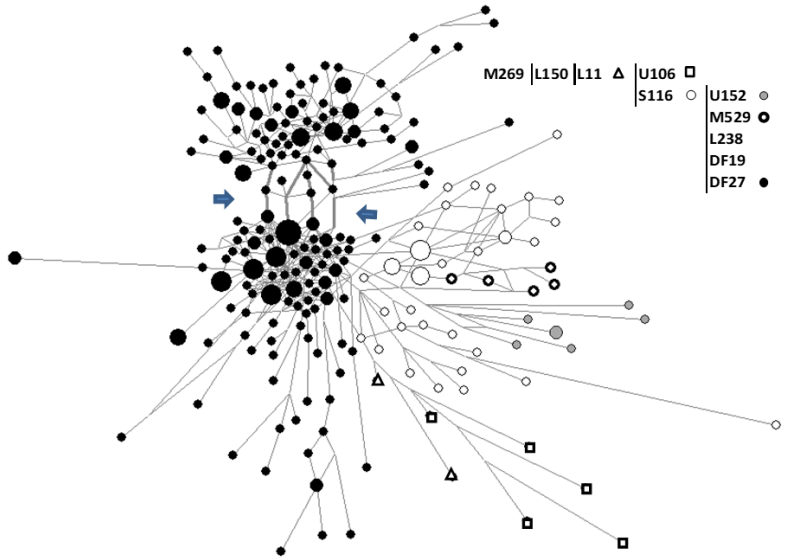


Fig. S4. Median joining network of the total Basque population, including both native and non-native individuals.

The network from Fig. S3 was assembled only for native individuals, with the aim of studying the ancestral gene pool of the population, in this case M269 ancestral lineages.

It is well known that Basque population is a genetic isolate. This may have caused a loss of diversity of lineages, which may affect the calculation of coalescence times and introduce errors in inferences.

In addition, the native Basque sample has been selected on the basis of the Basque surnames. This way of selection could remove part of the gene flow occurred during recent years, which may further reduce the diversity and alter the calculations.

To ensure the reliability of the calculations done with the native sample of population, a comparison of TMRCA results was done between the phylogeny constructed in Fig. S3 (including only native individuals) and a parallel phylogeny constructed including the total current Basque population (native and non native males, i.e. a random actual sampling in Basque Country without selection of individuals by surnames). This phylogeny of the actual Basque population is presented in Fig. S4.

The results in both cases were similar and identical conclusions were reached with both sets of population samples. So, the inclusion of non-native individuals did not alter the structure of the S116 and DF27 haplogroups. The blue arrows indicate the phylogenetic split of the DF27 haplogroup into two groups bearing the alleles 14/18 and 15/19 in the Y-STR haplotype DYS437/DYS448.

The dates obtained (S116: 10659.31 +/- 1511 YBP; DF27: 9988 +/- 1374.YBP) were only slightly lower, and they do not modify the prehistoric window period inferred in with the native Basque population.

These analyses demonstrate the reliability and robustness of the results obtained in the native sample of population, and state that the genetic isolation of Basque Country and/or the sampling strategy have not altered the demographic inferences.

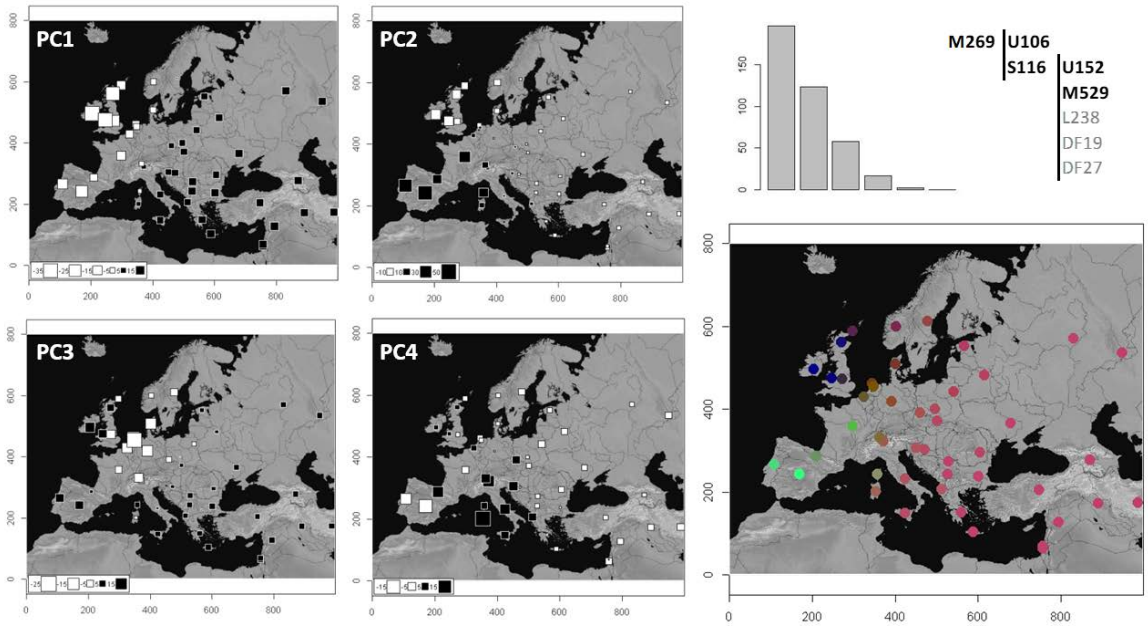


Fig. S5. Spatial PCAs based on haplogroup frequencies of the analysed populations and data compiled from Myres *et al*. (2011), Larmuseau *et al*. (2011) and Busby *et al*. (2012). Here, the level of resolution of the analysis is lower because S116 is not completely dissected in the literature data. The Y-SNPs used for the analysis are marked in bold in the tree. All the components of the analysis have positive eigenvalues (global structures). The spatial analyses of the most representative 4 principal components are presented. The colour plot corresponds to the spatial representations of the 2 principal components that explain the maximum variance. The colours make easier the identification of the different haplogroup spatial patterns found by the analysis.

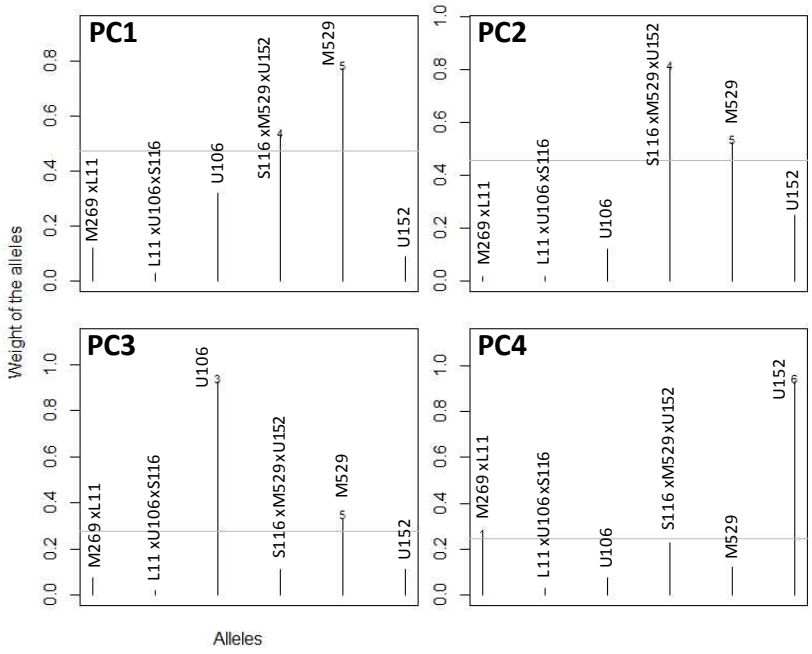


Fig. S6. Contributions of the alleles to the principal components 1, 2, 3 and 4 (PC1, PC2, PC3 and PC4, respectively) of sPCA of Fig. S5. The order of the Y-SNPs in the graph: (1) M269 (xL11), (2) L11 (xU106 xS116), (3) U106, (4) S116 (xM529 xU152), (5) M529 and (6) U152.

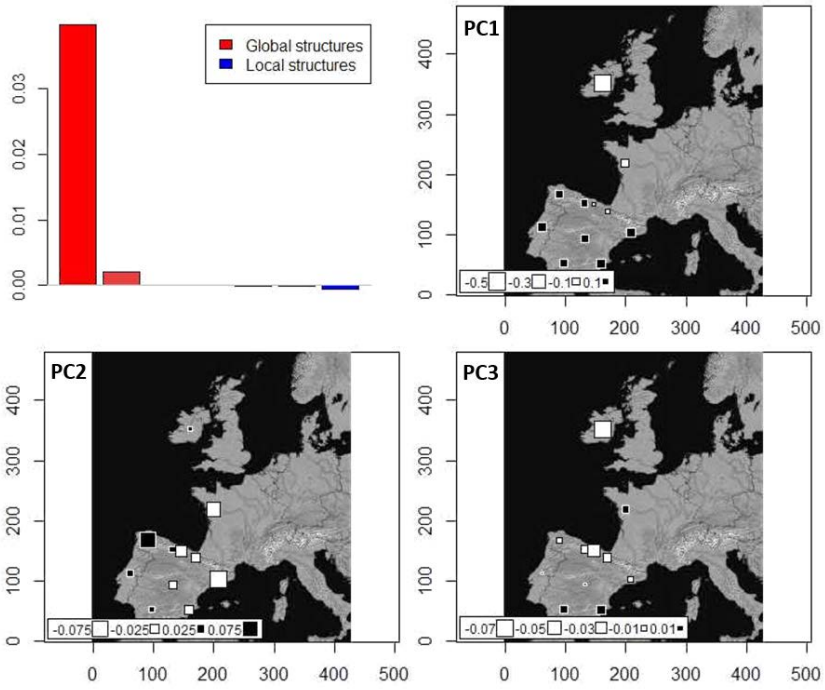


Fig. S7. Spatial PCAs based on haplogroup frequencies of the analysed populations. The bar plot indicates the eigenvalues obtained for every component. Single-population scores of the 2 positive eigenvalues (red) (PC1 and PC2) and the negative eigenvalues (blue) (PC3) are represented with black/white squares, associated with positive/negative values, respectively. Square size is proportional to the absolute value, indicating the degree of differentiation.

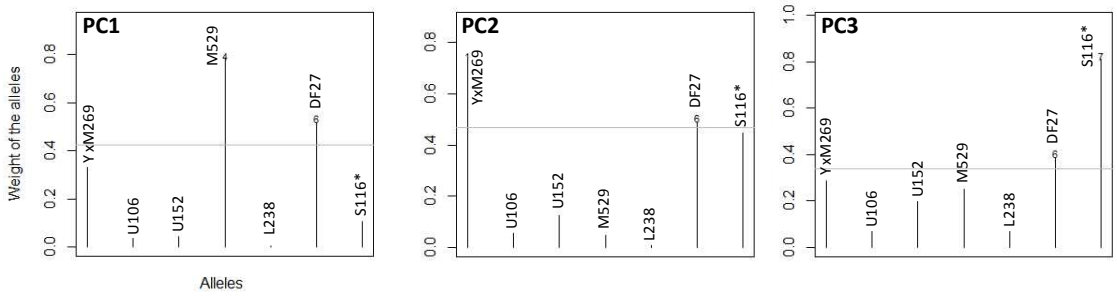


Fig. S8. Contribution of the alleles to the principal components 1, 2 and 3 (PC1, PC2 and PC3, respectively) of the sPCA of Fig. S7. The order of the Y-SNPs in the graph: (1) xM269, (2) U106, (3) U152, (4) M529, (5) L238, (6) DF27 and (7) S116*.

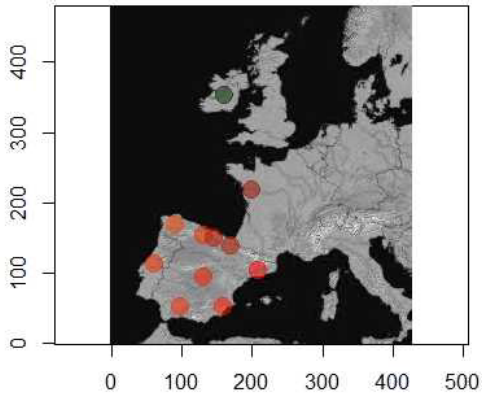


Fig. S9. Colour plot of the 2 principal positive components of the sPCA from Fig. S7. The colours make easier the identification of the different haplogroup spatial patterns found by the analysis. In this case, the red-orange dots identify the spatial pattern found for DF27 in Iberia, and the green dot for M529 in Ireland.

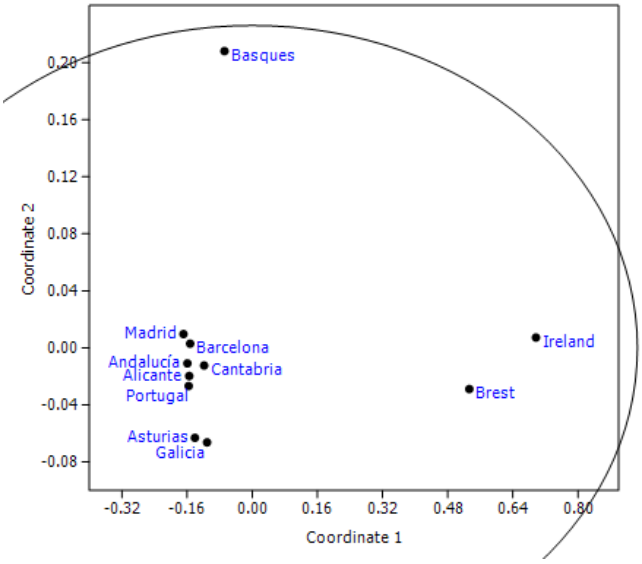


Fig. S10. Multidimensional scaling of genetic Fst distances calculated on the basis of Y-SNP haplogroup frequencies. Stress 0.048.

Iberian populations appear more clustered due to the absence of statistically significant differences between them, with the exception of the Basque population, which statistically differs both from Iberia (with the exception of the neighbouring Cantabria population and the cosmopolitan cities Madrid and Barcelona) and from Brest and Ireland (see Table S4).

4.2 Study Number 2

'Characterization of the Iberian Y chromosome haplogroup R-DF27 in Northern Spain; Dissection of the DF27 paternal lineage'

The Study Number 2 corresponds to the attainment of the first part of the objective 2 of the present doctoral thesis: *To characterize the structure and spatial distribution of the Iberian near-specific paternal lineage R1b-DF27 in Southwest European populations through the dissection in its sublineages, with the aim to estimate its time of origin, as well as to model its expansion in the phylogenetic context and the related demographic events.*

The dissection of R1b-M269 lineage in its sublineages and their analysis in Southwest European populations has revealed that the sublineage R1b-P312 (also known as S116) is split into geographically localized subhaplogroups. Among them, DF27 turned out to be near-specific of the Iberian Peninsula, where it most likely originated. Given the forensic interest of inferring the bio-geographical origin of a sample, knowing the distribution of DF27 lineage, as well as its structure, would be helpful for forensic casework. Despite the high interest DF27 created in the community, very few academic publications could be found in the literature, and only concerning a couple of its subhaplogroups in some populations. For that reason, a detailed dissection of this lineage was still necessary.

The objective of the present study was the characterization of DF27 lineage through the dissection in its sublineages. For that purpose, we analyzed the Y-SNPs DF27, Z195, Z196, L617, L881, Z220, Z278, M153, L176.2, S68, M167, and DF17 in 591 individuals from the population that previously displayed the highest frequencies for DF27 (Basque Country), along with other three surrounding populations (Asturias, Cantabria and Aragón) located in the North of the Iberian Peninsula. Additionally, we also collected frequency data from the reference populations in the 1,000 Genomes Project. We calculated Pairwise F_{ST} distances between the four populations and estimated the phylogenetic relationships of the related Y-STR haplotypes (extracted from previously reported data) by median joining networks. Time to the most Recent Common Ancestor (TMRCA) was estimated from 15 Y-STRs using the algorithms Rho (ρ) and the average square distance (ASD) with a mean germline mutation rate.

Our results revealed high frequencies of DF27 (34-70%) and its subhaplogroups in the analyzed populations. Accordingly, similar frequencies to those of the Iberian Peninsula were also observed in some of the populations from America extracted from the 1,000 Genomes database. The Y-STR haplotypes disclosed a phylogenetic split of Z196-Z220 from the bulk of DF27 due to the differing

haplotypes for DYS437-DYS448 related to the presence of sublineage Z220. Strikingly, despite the high subhaplogroup diversity displayed by DF27, the age estimated by TMRCA points to a recent origin approximately 4,000 years ago, in the early Bronze Age.

In view of the obtained findings, the need for further analysis including more coverage of the Iberian Peninsula in addition to other Southwest European populations is clear, since it will shed light on the origin and expansion of DF27.

This study has resulted in:

1. An international publication in the journal *Forensic Science International: Genetics* under the heading '*Characterization of the Iberian Y chromosome haplogroup R-DF27 in Northern Spain'* in March 2017. Q1, IP: 5.637.
2. A publication reporting the first stages of this study in the journal *Forensic Science International: Genetics Supplement Series* in September 2015 under the heading '*Dissection of the DF27 paternal lineage'*.

The above-mentioned publications are shown below.

Forensic Science International: Genetics 27 (2017) 142–148

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsig

Article

Characterization of the Iberian Y chromosome haplogroup R-DF27 in Northern Spain

Patricia Villaescusa[1], María José Illescas[1], Laura Valverde[1], Miriam Baeta[1], Carolina Nuñez[1], Begoña Martínez Jarreta[2], Maria Teresa Zarrabeitia[3], Francesc Calafell[4], Marian M. de Pancorbo[1,*]

[1]BIOMICs Research Group, Lascaray Research Center, University of the Basque Country UPV/EHU, Vitoria-Gasteiz, Spain.
[2]Laboratory of Genetics and Genetic Identification, University of Zaragoza, Spain.
[3]Unit of Legal Medicine, University of Cantabria. Av. Herrera Oria, s/n. 39011 Santander, Cantabria (Spain) and Instituto de Investigación Marqués de Valdecilla (IDIVAL). Avda. Cardenal Herrera Oria s/n. 39011 Santander, Cantabria, Spain.
[4]Departament de Ciències Experimentals i de la Salut, Institute of Evolutionary Biology (CSICUniversitat Pompeu Fabra), Universitat Pompeu Fabra, CEXS-UPF-PRBB, Barcelona, Catalonia, Spain.
[*] Corresponding Author

Received 26 June 2016, Received in revised form 22 December 2016, Accepted 29 December 2016, Available online 29 December 2016.

Abstract

The European paternal lineage R-DF27 has been proposed as an haplogroup of Iberian origin due to its maximum frequencies in the Iberian Peninsula. In this study, the distribution and structure of DF27 were characterized in 591 unrelated male individuals from four key populations of the north area of the Iberian Peninsula through the analysis of 12 Y-SNPs that define DF27 main sublineages. Additionally, Y-SNP allele frequencies were also gathered from the reference populations in the 1000 Genomes Project to compare and obtain a better landscape of the distribution of DF27. Our results reveal frequencies over 35% of DF27 haplogroup in the four North Iberian populations analyzed and high frequencies for its subhaplogroups. Considering the low frequency of DF27 and its sublineages in most populations outside of the Iberian Peninsula, this haplogroup seems to have geographical significance; thus, indicating a possible Iberian patrilineal origin of vestiges bearing this haplogroup. The dataset presented here contributes with new data

to better understand the complex genetic variability of the Y chromosome in the Iberian Peninsula, that can be applied in Forensic Genetics.

Keywords: Y chromosome; R haplogroup; Iberia; DF27; Z195.

1. Introduction

The inference of bio-geographical ancestry using markers with population-differentiated variation can provide investigative leads in forensic cases when eyewitness testimony or a database hit are not available [1]. A first step in forensic ancestry inference may be the analysis of the Y chromosome or the mitochondrial DNA (mtDNA), as they are uniparental markers differentiated geographically. Both lineages are not affected by recombination and correlate with continental regions [1,2]. Y chromosome and mtDNA, along with autosomal ancestry informative markers (AIMs) can be used for inferring the bio-geographical origin of a vestige, as they are tools that complement each other.

The analysis of Y chromosome SNPs (Y-SNPs) has revealed the existence of specific lineages in human populations at continental and regional levels [3]. The Y chromosome diversity analysis performed in multiple European populations disclosed the existence of significant frequency clines in the major patrilineal lineages [4]. The most frequent paternal lineage in Europe is R1b [5], being haplogroup R-M269 the most common in Central and Western Europe [6,7], with frequencies ≈0.4 in Italy and Germany, ≈0.6 in Britain, France, and the Iberian Peninsula; and up to >0.8 in Ireland and the Basque Country [7].

The origin of R-M269 has been the subject of great controversy [6–8], as it was originally believed to have originated in the Palaeolithic [9,10]. More recent analysis [11,12] suggested that this lineage had a Neolithic origin, but this claim was challenged [7] due to the Y-STR choice for computing the coalescence times and sample ascertainment. The last studies involving next-generation sequencing (NGS) of the Y-chromosome [8,13] and the analysis of ancient DNA [14] bring light to the debate, as they support more recent origin and continentwide expansion of the main European patrilineages, including R-M269 (≈5 KYA, middle Neolithic).

R-M269 is split into geographically localized sublineages, being the main branches U106 (more frequent in Central-Northern Europe) and P312 (Western and South Western Europe) [6,7]. The latter, in turn, trifurcates into U152, M529, and DF27. U152 is common in northern Italy and the Alpine region, while M529 is nearly restricted to the British Isles and Britanny [6,7,12]. DF27, instead, shows its maximum frequencies in the Basque Country, from where it decreases gradually

into the rest of the Iberian Peninsula; elsewhere, it is much rarer. Therefore, it has been suggested that DF27 is a distinctive Iberian ancestry haplogroup, where it likely first originated [6].

Given the forensic interest that derives from the near-specificity of DF27 in Iberia, we have characterized the structure and distribution of the DF27 lineage through the dissection in its sublineages [15–17]. For that purpose, we have analyzed the population that showed the highest frequency for DF27 [6], along with other four surrounding populations from the north of the Iberian Peninsula. In order to reach this aim, the Y-SNPs defining DF27 and its sublineages, Z195, Z196, L617, L881, Z220, Z278, M153, L176.2, S68, M167, and DF17 were genotyped. Additionally, Y-STR data from the Northern Iberia populations here analyzed were compiled from other studies [6,18–20].

2. Materials and methods

2.1 Sample collection

A total of 591 healthy unrelated males from four different populations from Northern Iberia (from West to East: Asturias, Cantabria, Basque Country, and Aragon) were obtained. Informed consent was obtained from all individuals participating in the study. Human DNA samples were extracted from saliva (Asturias and Basque Country), peripheral blood (Cantabria and Basque Country) or blood stains collected on FTA paper (Aragon). Samples from Cantabria and Aragon were provided by the collection of the University of Cantabria and from the University of Zaragoza, respectively. Samples from Asturias were provided by the Spanish National DNA Bank-Carlos III (BNADN). Favorable ethical reports were obtained (Faculty of Pharmacy UPV/EHU, September 26th 2008; CEISH/119/2012, BNADN Ref. 12/0031). Fig. 1 shows the geographic location of the selected populations.

The Basque population was treated as two separate groups, native and resident Basques. Significant genetic differences were found between the two Basque groups, so it is not recommended to treat these populations as a single sample [21]. The inclusion criteria used in order to define natives was the Basque origin of surnames and birthplaces of the individuals and their ancestors going back at least three generations. The resident group corresponds to those individuals that live in the Basque Country but whose paternal ancestors are not native Basques, coming from elsewhere in Spain.

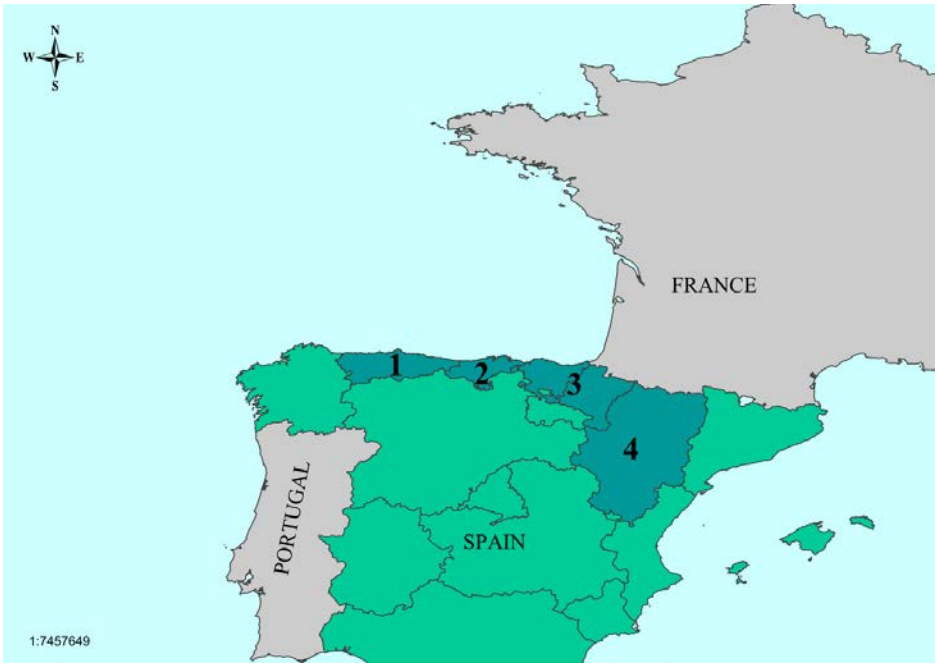


Fig. 1. Geographic location of the studied populations in Northern Spain. 1: Asturias; 2: Cantabria; 3: Basque Country; 4: Aragon.

2.2 Y-SNP analysis

The samples were genotyped in a hierarchical manner for the following Y-SNPs within the RDF27 haplogroup: DF27, Z195, Z196, Z220, Z278, M153, L176.2, M167 (also known as SRY2627 [16]), S68, DF17, L617, and L881 [15,22]. More details about the phylogeny are represented in Fig. 2. The Y-SNPs selected in this study correspond to the diagnostic positions that determine the main sublineages of DF27 haplogroup. These positions were chosen following the minimal reference phylogeny for the human Y chromosome [22], supplemented when necessary with the more detailed tree maintained by the International Society of Genetic Genealogy [23].

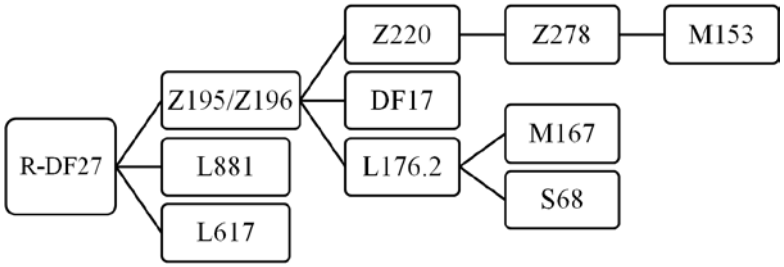


Fig. 2. Phylogenetic tree of R-DF27 paternal lineage. The assignment of haplogroups follows the minimal reference phylogeny for the human Y chromosome [22], supplemented when necessary with the more detailed tree maintained by the International Society of Genetic Genealogy [23].

SNPs were genotyped with High Resolution Melting (HRM) technology, Sanger sequencing and pyrosequencing. Y-SNP details, primers used for amplification and further technical details on techniques are shown in Table S1.

Y-SNPs DF27, Z195, Z196, Z220, Z278, L176.2, M167, S68, and L881 were analyzed by HRM, while M153 and DF17 were analyzed by Sanger sequencing. The amplification, melting and sequencing conditions were previously described [6].

SNP L617 was genotyped through pyrosequencing. PCR amplification was performed by using Hot Start Taq®Plus DNA Polymerase (Qiagen), 7.5 µM biotinylated primer, 15 µM nonbiotinylated primer and 1 ng DNA. The quality and quantity of the PCR product was confirmed with agarose gel (1.5%) electrophoresis. Pyrosequencing was carried out using the PyroMark Gold Q96 reagents (Qiagen) on a PyroMark 96MD Pyrosequencer (Qiagen). Output data was interpreted with PyroMark Q96 MD Software (Qiagen).

2.3 Y-STR data for Northern Iberian populations

In order to further study the genetic variability of the Northern Iberian populations, data from 16 Y-STRs (DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS438, DYS439, DYS437, DYS448, DYS456, DYS458, DYS635, DYS385ab, and YGATA H4) for a selected sub-population of 203 Basques (154 native and 49 residents), 30 Cantabrian, 26 Asturian and 29 Aragonese were collected from previously reported data [6,18,19].

2.4 Statistical analysis

The absolute and relative frequencies for each SNP were manually counted. Genetic distances F_{ST}, haplotype diversity (HD) and haplogroup diversity (HGD) were calculated using Arlequin v 3.5 software [24]. Significance p values were obtained after the Bonferroni correction ($\alpha = 0.05/\{[(1 + n)/2] \times n\}$; n = number of populations) [25]. The phylogenetic relationships of Y-STR haplotypes were estimated with median joining networks using Network v 5.0.0.0 [26]. Phylogenetic weight was assigned to each locus proportionally to the inverse of the repeat size variance. Anti-reticulation options implemented within the software, such as the shortest trees option, were performed if necessary. For the network construction, alleles at DYS389I were subtracted from those at DYS389II and DYS385ab was excluded.

Time to the Most Recent Common Ancestor (TMRCA) was estimated using Rho, as implemented within Network v 5.0.0.0 [26]; and average square distance (ASD) [27,28], by using the Kilin-Klyosov TMRCA Calculator [29] with some modifications. Mean germ-line mutation rates for the Y-STRs ($\mu = 1.37 \times 10^{-3}$ per locus per generation) were obtained from the compilation in the YHRD

(Y-STR Haplotype Reference Database) database (www.yhrd.org, accessed on Feb. 17th 2016) given the mutation rate of a set of 14 Y-STR; with a generation time for men of 30 years [30], this yields one mutation per haplotype per 728 years. Limitations discussed in [7] were considered when using Y-STRs for the estimation of TMRCA.

The Y-STR dataset used in this study can be consulted online on the yhrd.org website [31] with the accession number YA003887 and YA003047 for Asturias; YA004015 for Cantabria; YA003672-3677, YA003718 and YA004016-4018 for the Basque Country; and YA003046 for Aragon. New Y-SNP results were updated for each population dataset.

3. Results and discussion

This is the first study that presents the dissection of the Y chromosome haplogroup DF27 for its immediately known subhaplogroups: L617, L881, and Z196 and their corresponding subhaplogroups, in a representative sample of populations from the North Iberian Peninsula where DF27 haplogroup shows a peak of maximum frequency. SNP Z195 was also genotyped, but according to its phylogenetic position [22] it should be redundant with SNP Z196; indeed, Z195 and Z196 results are consistent in all samples.

The resulting Y-SNP frequencies and Y-SNP/STR haplotypes are shown in Table 1 and Supplementary Tables S2-S3. The DF27 lineage reaches frequencies over 40% in all the populations analyzed except Aragon, where it is slightly lower (35%). The frequency peak displayed in native Basques (70%) is particularly striking. This could be due to the Basque Country being the place of origin of DF27 but, on the other hand, it could also be explained by the effect of genetic drift due to the geographical and cultural isolation of this population. L617 was not represented in Northern Iberian populations except in native Basque population, where it was rare (2%). L881 was absent in all of the 591 individuals assayed. In contrast, the Z196 derived allele was present in frequencies between 38-13% in all the populations analyzed.

The frequent Z196 haplogroup was further studied for their subhaplogroups Z220, L176.2, and DF17; significant differences in their frequency distribution along the North of the Iberian Peninsula were observed. DF17 was found only in extremely low frequencies in Aragon and native Basques (0-1%). However, the Y-SNPs Z220 and L176.2 virtually resolved almost all the variability within Z196, resulting in low frequencies of the paragroup Z196* in all populations. SNP Z220, specifically, reaches maximum frequencies in Basques and Cantabria (>28% in native Basques and 17% in Cantabria), while L176.2 reaches its maximum frequency in Aragon (11%). Similarly, the frequencies of the analyzed Z220 and L176.2 subhaplogroups maintain their maximum

frequencies in the same geographical area as their parent haplogroup. The Z220-Z278 subhaplogroup nearly resolves all the variability within Z220, whereas Z220-M153 shows a peak frequency in native Basques, being almost restricted to this population. SNPs L176.2-M167 and L176.2-S68 show the highest frequencies in Aragon and Basques, being completely absent in the western population of Asturias (Table 1).

Table 1. Y-SNP frequencies in the analyzed samples of population. For each haplogroup/column, the higher the frequency, the more intense the color.

Frequencies (%)	N	DF27	DF27*	L617	L881	Z196	Z196*	Z220	Z220*	Z278	M153	DF17	L176.2	L176.2*	M167	S68
Asturias	63	42.86	30.16	0.00	0.00	12.70	7.94	4.76	1.59	3.17	0.00	0.00	0.00	0.00	0.00	0.00
Cantabria	96	44.79	20.83	0.00	0.00	21.88	2.08	16.67	2.08	12.50	0.00	0.00	3.13	1.04	2.08	0.00
Native Basques	229	70.74	31.00	1.75	0.00	37.99	3.06	28.38	7.86	20.52	6.55	0.44	6.11	2.18	3.06	0.87
Resident Basques	111	47.75	24.32	0.00	0.00	23.42	4.50	11.71	1.80	9.91	0.90	0.00	7.21	0.90	6.31	0.00
Aragon	92	34.78	15.22	0.00	0.00	19.57	0.00	7.61	4.35	3.26	1.09	1.09	10.87	5.43	4.35	1.09

The remaining DF27 individuals not belonging to Z196 (with the exception of the few L617 Basques) were within the paragroup DF27*, reaching maximum frequencies in native Basques and Asturias (30%), and a frequency over 15% in the rest of Northern Iberian populations analyzed.

We also compared the frequency of the DF27 lineage and its sublineages of the populations we analyzed with 15 other populations extracted from the literature [32], including the Iberian Peninsula (Table S2). We selected populations from Europe, America, Africa, and Asia to obtain a better landscape of the distribution of this haplogroup over the world. The frequency of DF27 in the Iberian Peninsula (44%) is similar to what we obtained in our analyses. However, in other European populations (Britain, Italy and Finland) the frequency is much lower (0-13%), which suggests that the DF27 haplogroup decreases away from Iberia; this decrease is even steeper when subhaplogroups of DF27 are considered, as in the case of Z220, with frequencies 18% in Iberia and 0-2% in the other European populations. In the Americas, DF27 displays similar frequencies to those of the Iberian Peninsula in Colombia and Puerto Rico, and shows lower frequencies in Mexico and Peru. These areas have been a destination of the historically known Spanish migration [33–36], therefore the frequency of this haplogroup may actually be an indication of the degree of patrilineal Spanish versus Native American admixture. Given the dispersion of the Spanish population throughout history, it would be of great interest to study the presence of DF27 in other areas of the world that have been former Spanish European possessions (such as Southern Italy, Sardinia, Sicily, and Flanders) or overseas colonies (such as the Philippines). SNP DF27 is absent in populations from Africa (Sierra Leone and Kenya) and Asia (Japan and Vietnam), and it is present in low frequencies (2-3%) in African Caribbean from Barbados and Americans of African ancestry in Southwest United States (USA) where they could signal limited periods of Spanish admixture.

The high frequencies of DF27 and its subhaplogroups in the Iberian Peninsula, together with the low frequencies observed outside of this area point out the geographical distinctiveness of DF27; thus, individuals bearing this haplogroup would have a high probability of having a patrilineal Iberian origin. This is especially useful in forensic genetics, since a vast majority of crimes where DNA evidence is helpful involve males as perpetrators, and determining the lineage or likely geographical origin of a male individual may be helpful [37]. Y-SNPs are also quite informative for evolutionary studies and kinship analyses because of the lack of recombination and their low mutation rate. They are usually employed in forensics for missing person cases or mass disaster identifications, especially in instances where the reference sample(s) and the evidence sample are separated by several generations [2]. SNP DF27 could be useful for contributing to the information of the origin of individuals from the Iberian Peninsula or of Hispanic origins. In particular, data are available for four populations sampled in the US: European Americans from Utah, African Americans from the Southwest, Mexicans from Los Angeles, and Puerto Ricans. While the persistence of the Native American substrate in Mexicans seems to have brought down the DF27 frequency to levels similar to those in European Americans, it is 3.5 times higher in Puerto Ricans; conversely, it is 2.6 times smaller in African Americans (Table S2). Thus, DF27 and its subhaplogroups show some potential for suggesting an origin of a sample of unknown source in some Hispanic subpopulations. Nevertheless, complementing the information with other markers such as Ancestry AIMs may be needed in order to ascertain the origin of a vestige.

Haplogroup diversity based on Y-SNPs was obtained (Table S4). The populations from Basque Country and Cantabria displayed the highest diversity values, 0.789 ± 0.015 (native Basques); 0.659 ± 0.037 (residents Basques) and 0.609 ± 0.046, respectively. Asturias and Aragon showed lower diversities, 0.584 ± 0.047 and 0.550 ± 0.057 respectively. This could be explained by the absence of some sublineages of DF27, like in the case of Asturias (e.g. L176.2). If the main YSNPs (other than DF27) are added, higher diversities are observed, in a range from 0.85 to 0.91 (Table S4). The diversities for these same populations obtained by Y-STRs [18,19] are higher than the ones obtained by Y-SNPs, with values between 0.99 and 1, as it would be expected mainly due to the larger allelic range and higher mutation rate of the Y-STRs [38].

In order to study the genetic relationships between the populations analyzed in this study, genetic distances (F_{ST}) and their corresponding p-values were obtained (Table S5). Significant differences ($p<0.005$) were found between native Basques and the rest of the populations studied. The native Basque population is composed only by autochthonous individuals and has been subjected to a long isolation that has minimized the contribution of non Basque Y chromosomes to their lineages; therefore, this kind of differentiation is to be expected since most of their patrilineages may have

a Basque origin. The distinctiveness of the Basque population has also been observed in other studies based on Y-SNPs and Y-STRs [6,18,21], X chromosome STRs [39] and mitochondrial DNA [40,41]. The rest of the analyzed populations did not show statistically significant differences between each other. These results are consistent with other studies that show a similar population structure of Y-SNP and Y-STR at the level of M269 lineage for other populations from the Iberian Peninsula [6,18,42].

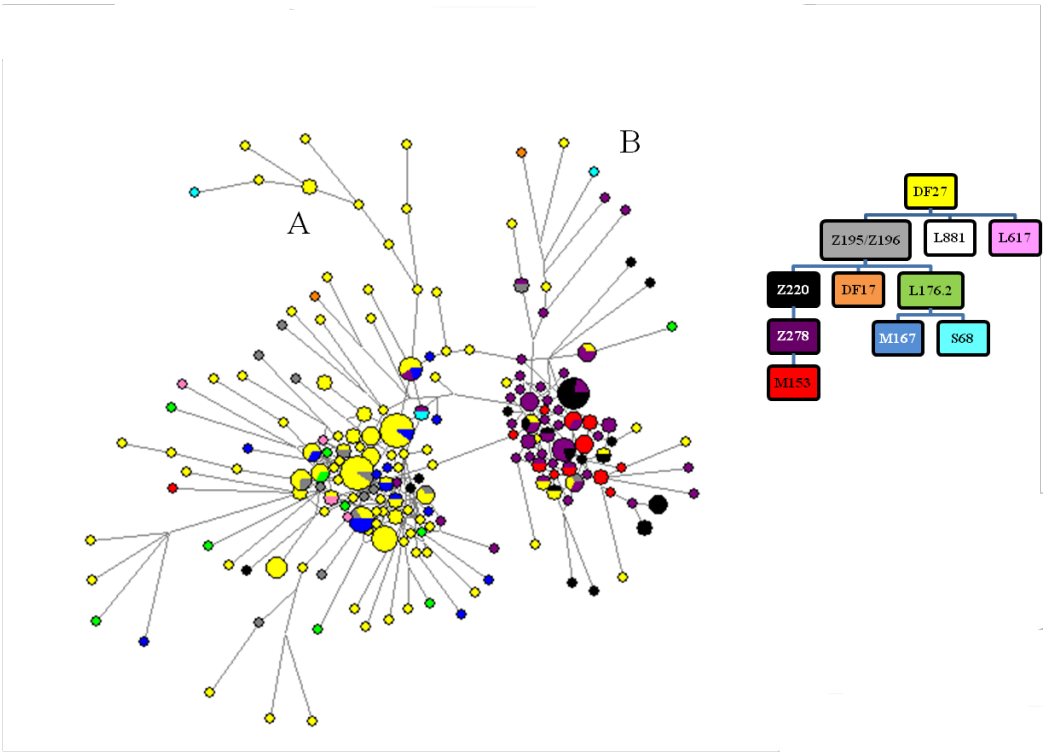


Fig. 3. Median joining network of DF27 haplogroup in the populations of Asturias, Cantabria, native Basques, resident Basques, and Aragon. The phylogenetic split for DF27 haplogroup is due to differing haplotypes for DYS437/DYS448 Y-STRs.

Additionally, median joining networks were built with 14 Y-STR haplotypes to further characterize the structure of the DF27 lineage (Fig. 3 and Supplementary Figs. S1-S3). The phylogenetic split of Z196-Z220 from the bulk of DF27 Y-STR haplotypes was previously detected by network analyses [6], but now it can be confirmed by Z220 subhaplogroup (Fig. 3). The phylogenetic split is due to differing haplotypes for DYS437/DYS448 Y-STRs. The right-hand part of the split in the network included principally Z278, M153 and Z220* individuals and bears haplotype 14/18, whereas the left-hand part of the split, which includes the remaining DF27 subhaplogroups, bears the 15/19 haplotype, which is also the most prevalent in the rest of R1b-M269 chromosomes [42]. The present network has also showed divergent branches from DF27 node groups, which includes almost exclusively native Basque individuals (Fig. 3, A). The Z220 node groups also show divergent branches, whereas this branch contains individuals from Cantabria and native Basques (Fig. 3, B).

This could indicate some degree of independent evolution of some of the paternal lineages from each of the regions, probably due to the isolation of the populations in the mountain area of Northern Iberia.

The DF27 paragroup was further studied in order to determine possible patterns of internal variability (Figs. S1-S3). The network (Fig. S1) showed a non homogeneous star-like structure, where the core appears highly reticulated and is composed mainly of native Basque individuals. Two short branches (Fig. S1, A and B) appear to diverge from the core haplotype. Branch A is composed mainly by native Basque individuals (except a few Cantabrian, Asturian and resident Basque individuals) and can be traced by the Y-STR alleles DYS391*10 and DYS393*12. Branch B is composed by the few DF27* individuals that share the haplotype DYS437/DYS448 14/18, which is more distinctive of the Z220 lineage and its sublineages. Those samples can also be traced by the Y-STR alleles DYS391*11 and DYS393*13. A more simplified version of DF27* network (Fig. S2), made by using the anti-reticulation options included within the Network software, shows a more clear structure of the core and some divergent branches, which could be due to the intrinsic variability of the Y-STR loci or to a greater divergence, which may have resulted in the presence of yet undiscovered SNPs. Nonetheless, the artificial simplification of the structure of DF27 paragroup by the Network software should be also considered. Two main branches (Figs. S2, A and B) are distinguished by the DYS456 15 and 16 alleles, respectively. Since the core of DF27* network is composed mainly of native Basque individuals, a separate network of DF27* lineage was constructed for the native Basque population (Fig. S3), which conserved a similar structure to the previous network and, interestingly, practically has no median vectors. The absence of median vectors could be due to the exhaustive sampling of such a young paragroup. The presence of such variability of DF27* Y-STR haplotypes in the analyzed populations and the fact that the core of the network is composed mainly by native Basque individuals could be evidence for the origin of DF27 in the northern area of the Iberian Peninsula.

Finally, we estimated the TMRCAs of the DF27 haplogroup and its subhaplogroups Z196, Z220, L176.2, Z278, M153, and M167. The ages of M269 lineage and its sublineages have been the subject of controversy [6–8,14]; however, recent studies [8,13,43] place the age of M269 in the Middle Neolithic (5 KYA). Using a “pedigree” Y-STR mutation rate based on direct detection of mutations in father-son pairs [43], our results for the age of DF27 and its subhaplogroups are consistent with this period. DF27 (4176 ± 696 years) seems to have originated in the early Bronze Age probably somewhere in the Iberian Peninsula. Its main sublineage Z196 (3173 ± 502) arose early after DF27, and the same pattern is observed with the rest of its sublineages Z220 (2904 ± 474), L176.2 (2935 ± 514), Z278 (2817 ± 515), M153 (1627 ± 535), and M167 (2141 ± 437). Similar

results are obtained when estimating the TMRCAs using ASD (DF27: 3862 ± 372; Z196: 3575 ± 436; Z220: 3155 ± 464; L176.2: 3352 ± 631; Z278: 2767 ± 411; M153: 1260 ± 232; and M167: 2595 ± 744).

A further analysis of the Iberian lineages together with other lineages located in different areas of Europe and/or the world may be helpful in order to establish more accurately the origin of DF27 haplogroup.

4. Conclusion

The analysis of 591 individuals from Northern Iberia reveals a high frequency of the DF27 haplogroup and its subhaplogroups in four North Iberian populations and increase the available information of the structure and distribution of this lineage. The results of the dissection of DF27 here performed suggest a recent origin of this lineage. A deeper knowledge of the variability within DF27 paragroups and subhaplogroups would provide new data about even younger subhaplogroups related to the Iberian people that inhabited the Peninsula during more recent times, which would be of great value to elucidate the complex population movements and mixtures along the history of the Iberian Peninsula.

The data presented in this research will be valuable in the field of the forensic genetics due to the usefulness of DF27 and its subhaplogroups for indicating the Iberian origin of forensic evidences.

Quality control

This paper follows the guidelines for publication of population data requested by the journal [44].

Conflict of interest

Authors declare no competing interest in the content of this manuscript.

Acknowledgements

Funds were provided by the Basque Government (IT-424-07). PV received a PhD grant from the University of the Basque Country UPV/EHU. The authors are deeply indebted to the Basque Foundation of Science (BIOEF), the Spanish National DNA Bank Carlos III (BNADN), SGIker (UPV/EHU, MICINN, GV/EJ, ERDF and ESF) and to all the people who voluntarily participated in this study.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.fsigen.2016.12.013.

References

[1] C. Phillips, Forensic genetic analysis of bio-geographical ancestry, Forensic Sci. Int. Genet. 18 (2015) 49–65, doi:http://dx.doi.org/10.1016/j.fsigen.2015.05.012.

[2] B. Budowle, A. Van Daal, Forensically relevant SNP classes, Biotechniques 44 (2008) 603–610, doi:http://dx.doi.org/10.2144/000112806.

[3] P.A. Underhill, P. Shen, A.A. Lin, L. Jin, G. Passarino, W.H. Yang, et al., Y chromosome sequence variation and the history of human populations, Nat. Genet. 26 (2000) 358–361, doi:http://dx.doi.org/10.1038/81685.

[4] Z.H. Rosser, T. Zerjal, M.E. Hurles, M. Adojaan, D. Alavantic, A. Amorim, et al., Ychromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language, Am. J. Hum. Genet. 67 (2000) 1526–1543, doi:http://dx.doi.org/10.1086/316890.

[5] O. Lao, T.T. Lu, M. Nothnagel, O. Junge, S. Freitag-Wolf, A. Caliebe, et al., Correlation between genetic and geographic structure in europe, Curr. Biol. 18 (2008) 1241–1248, doi:http://dx.doi.org/10.1016/j.cub.2008.07.049.

[6] L. Valverde, M.J. Illescas, P. Villaescusa, A.M. Gotor, A. García, S. Cardoso, et al., New clues to the evolutionary history of the main European paternal lineage M269: dissection of the Y-SNP S116 in Atlantic Europe and Iberia, Eur. J. Hum. Genet. 24 (2016) 437–441, doi:http://dx.doi.org/10.1038/ejhg.2015.114.

[7] G.B.J. Busby, F. Brisighelli, P. Sánchez-Diz, E. Ramos-Luis, C. Martinez-Cadenas, M.G. Thomas, et al., The peopling of Europe and the cautionary tale of Y chromosome lineage R-M269, Proc. Biol. Sci. 279 (2012) 884–892, doi:http://dx.doi.org/10.1098/rspb.2011.1044.

[8] C. Batini, P. Hallast, D. Zadik, P.M. Delser, A. Benazzo, S. Ghirotto, et al., Largescale recent expansion of European patrilineages shown by population resequencing, Nat. Commun. Commun. (2015) 7152, doi:http://dx.doi.org/10.1038/ncomms8152.

[9] O. Semino, G. Passarino, P.J. Oefner, A.A. Lin, S. Arbuzova, L.E. Beckman, et al., The genetic legacy of Paleolithic Homo sapiens in extant Europeans: a Y chromosome perspective, Science 290 (2000) 1155–1159, doi:http://dx.doi.org/10.1126/science.290.5494.1155.

[10] P. Soares, A. Achilli, O. Semino, W. Davies, V. Macaulay, H.J. Bandelt, et al., The archaeogenetics of europe, Curr. Biol. 20 (2010), doi:http://dx.doi.org/10.1016/j.cub.2009.11.054.

[11] P. Balaresque, G.R. Bowden, S.M. Adams, H.Y. Leung, T.E. King, Z.H. Rosser, et al., A predominantly neolithic origin for European paternal lineages, PLoS Biol. 8 (2010), doi:http://dx.doi.org/10.1371/journal.pbio.1000285.

[12] N.M. Myres, S. Rootsi, A.A. Lin, M. Järve, R.J. King, I. Kutuev, et al., A major Ychromosome haplogroup R1b holocene era founder effect in central and western europe, Eur. J. Hum. Genet. 19 (2011) 95–101, doi:http://dx.doi.org/10.1038/ejhg.2010.146.

[13] G.D. Poznik, Y. Xue, F.L. Mendez, T.F. Willems, A. Massaia, M.A. Wilson Sayres, et al., Punctuated bursts in human male demography inferred from 1,244 worldwide Y-chromosome sequences, Nat Genet. Adv. On. 48 (2016) 593–599, doi:http://dx.doi.org/10.1038/ng.3559.

[14] W. Haak, I. Lazaridis, N. Patterson, N. Rohland, S. Mallick, B. Llamas, et al., Massive migration from the steppe was a source for Indo-European languages in Europe, Nature 522 (2015) 207–211, doi:http://dx.doi.org/10.1038/nature14317.

[15] R.A. Rocca, G. Magoon, D.F. Reynolds, T. Krahn, V.O. Tilroe, P.M. Op den Velde Boots, et al., Discovery of Western European R1b1a2 Y chromosome variants in 1000 genomes project data: an online community approach, PLoS One 7 (2012), doi:http://dx.doi.org/10.1371/journal.pone.0041634.

[16] M.E. Hurles, R. Veitia, E. Arroyo, M. Armenteros, J. Bertranpetit, A. Pérez-Lezaun, et al., Recent male-mediated gene flow over a linguistic barrier in Iberia, suggested by analysis of a Y-chromosomal DNA polymorphism, Am. J. Hum. Genet. 65 (1999) 1437–1448, doi:http://dx.doi.org/10.1086/302617.

[17] P.S. Underhill, P. Shen, A.A. Lin, L. Jin, G. Passarino, W.H. Yang, et al., Y chromosome sequence variation and the history of human populations, Nat. Genet. 26 (2000) 358–361, doi:http://dx.doi.org/10.1038/81685.

[18] C. Nuñez, M. Baeta, M. Fernández, M. Zarrabeitia, B. Martinez-Jarreta, M.M. de Pancorbo, Highly discriminatory capacity of the PowerPlex (®) Y23 System for the study of isolated populations, Forensic Sci. Int. Genet. 17 (2015) 104–107, doi:http://dx.doi.org/10.1016/j.fsigen.2015.04.005.

[19] J. Purps, S. Siegert, S. Willuweit, M. Nagy, C. Alves, R. Salazar, et al., A global analysis of Y-chromosomal haplotype diversity for 23 STR loci, Forensic Sci. Int. Genet. 12 (2014) 12–23, doi:http://dx.doi.org/10.1016/j.fsigen.2014.04.008.

[20] L. Valverde, S. Köhnemann, M. Rosique, S. Cardoso, M. Zarrabeitia, H. Pfeiffer, et al., 17 Y-STR haplotype data for a population sample of Residents in the Basque Country, Forensic Sci. Int. Genet. 6 (2012), doi:http://dx.doi.org/10.1016/j.fsigen.2012.01.006.

[21] L. Valverde, M. Rosique, S. Köhnemann, S. Cardoso, A. García, A. Odriozola, et al., Y-STR variation in the Basque diaspora in the Western USA: evolutionary and forensic perspectives, Int. J. Legal Med. 126 (2012) 293–298, doi:http://dx.doi.org/10.1007/s00414-011-0644-8.

[22] M. Van Oven, A. Van Geystelen, M. Kayser, R. Decorte, M.H. Larmuseau, Seeing the wood for the trees: a minimal reference phylogeny for the human Y chromosome, Hum. Mutat. 35 (2014) 187–191, doi:http://dx.doi.org/10.1002/humu.22468.

[23] International Society of Genetic Genealogy (2016). Y-DNA Haplogroup Tree 2016, Version 10.01, (n.d.). http://isogg.org/tree/ (accessed September 1, 2015).

[24] L. Excoffier, H.E.L. Lischer, Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows, Mol. Ecol. Resour. 10 (2010) 564–567, doi:http://dx.doi.org/10.1111/j.1755-0998.2010.02847.x.

[25] Z. Zeng, R. Garcia-Bertrand, S. Calderon, L. Li, M. Zhong, R.J. Herrera, Extreme genetic heterogeneity among the nine major tribal Taiwanese island populations detected with a new generation Y23 STR system, Forensic Sci. Int. Genet. 12 (2014) 100–106, doi:http://dx.doi.org/10.1016/j.fsigen.2014.05.004.

[26] H.J. Bandelt, P. Forster, A. Röhl, Median-joining networks for inferring intraspecific phylogenies, Mol. Biol. Evol. 16 (1999) 37–48, doi:http://dx.doi.org/10.1093/oxfordjournals.molbev.a026036.

[27] D.B. Goldstein, A.R. Linares, L.L. Cavalli-Sforza, M.W. Feldman, An evaluation of genetic distances for use with microsatellite loci, Genetics 139 (1995) 463–471.

[28] D.B. Goldstein, A. Ruiz Linares, L.L. Cavalli-Sforza, M.W. Feldman, Genetic absolute dating based on microsatellites and the origin of modern humans, Proc. Natl. Acad. Sci. U. S. A. 92 (1995) 6723–6727, doi:http://dx.doi.org/10.1073/pnas.92.15.6723.

[29] A.A. Klyosov, V.V. Kilin, Kilin-Klyosov TMRCA calculator for time spans up to millions of years, Adv. Anthropol. 6 (2016) 51–71.

[30] A. Helgason, A.W. Einarsson, V.B. GuÐmundsdóttir, Á. SigurÐsson, E.D. Gunnarsdóttir, A. Jagadeesan, et al., The Y-chromosome point mutation rate in humans, Nat. Genet. 47 (2015) 453–457, doi:http://dx.doi.org/10.1038/ng.3171.

[31] S. Willuweit, L. Roewer, Y chromosome haplotype reference database (YHRD): update, Forensic Sci. Int. Genet.1 (2007) 83–87, doi:http://dx.doi.org/10.1016/j.fsigen.2007.01.017.

[32] 1000 Genomes Project Consortium, A. Auton, L.D. Brooks, R.M. Durbin, E.P. Garrison, H.M. Kang, et al., A global reference for human genetic variation, Nature 526 (2015) 68–74, doi:http://dx.doi.org/10.1038/nature15393.

[33] L.A. Alonso, W. Usaquén, Y-chromosome and surname analysis of the native islanders of San Andrés and Providencia (Colombia), HOMO- J. Comp. Hum. Biol. 64 (2013) 71–84, doi:http://dx.doi.org/10.1016/j.jchb.2012.11.006.

[34] M.C. Noguera, A. Schwegler, V. Gomes, I. Briceño, L. Alvarez, D. Uricoechea, et al., Colombia's racial crucible: y chromosome evidence from six admixed communities in the Department of Bolivar, Ann. Hum. Biol. 41 (2013) 453–459, doi:http://dx.doi.org/10.3109/03014460.2013.852244.

[35] C.O. Sauer, The Early Spanish Main, Cambridge University Press, 2008.

[36] F.M. Padron, Historia Del Decubrimiento Y Conquista De América, Gredos,1990.

[37] J.M. Butler, Forensic DNA Typing, 2nd edition, Academic Press, 2005.

[38] M.H.D. Larmuseau, N. Vanderheyden, A. Van Geystelen, M. Van Oven, P. De Knijff, R. Decorte, Recent radiation within Y-chromosomal haplogroup R-M269 resulted in high Y-STR haplotype resemblance, Ann. Hum. Genet. 78 (2014) 92–103, doi:http://dx.doi.org/10.1111/ahg.12050.

[39] M.J. Illescas, A. Pérez, J.M. Aznar, L. Valverde, S. Cardoso, J. Algorta, et al., Population genetic data for 10 X-STR loci in autochthonous Basques from Navarre (Spain), Forensic Sci. Int. Genet. 6 (2012), doi:http://dx.doi.org/10.1016/j.fsigen.2012.02.014.

[40] S. Cardoso, M.A. Alfonso-Sánchez, L. Valverde, A. Odriozola, A.M. Pérez-Miranda, J.A. Peña, et al., The maternal legacy of Basques in northern navarre: new insights into the mitochondrial DNA diversity of the Franco-Cantabrian area, Am. J. Phys. Anthropol. 145 (2011) 480–488, doi:http://dx.doi.org/10.1002/ajpa.21532.

[41] M.A. Alfonso-Sánchez, S. Cardoso, C. Martínez-Bouzas, J.A. Peña, R.J. Herrera, A. Castro, et al., Mitochondrial DNA haplogroup diversity in basques: a reassessment based on HVI and HVII polymorphisms, Am. J. Hum. Biol. 20 (2008) 154–164, doi:http://dx.doi.org/10.1002/ajhb.20706.

[42] N. Solé-Morata, J. Bertranpetit, D. Comas, F. Calafell, Recent radiation within Y chromosomal haplogroup R-M269 resulted in high Y-STR haplotype resemblance, Ann. Hum. Genet. 78 (2014) 253–254, doi:http://dx.doi.org/10.1111/ahg.12066.

[43] P. Hallast, C. Batini, D. Zadik, P.M. Delser, J.H. Wetton, E. Arroyo-Pardo, et al., The Y-chromosome tree bursts into leaf: 13 000 high-confidence SNPs covering the majority of known clades, Mol. Biol. Evol. 32 (2014) 661–673, doi: http://dx.doi.org/10.1093/molbev/msu327.

[44] Á. Carracedo, J.M. Butler, L. Gusmão, Á. Linacre, W. Parson, L. Roewer, et al., New guidelines for the publication of genetic population data, Forensic Sci. Int. Genet. 7 (2013) 217–220, doi:http://dx.doi.org/10.1016/j.fsigen.2013.01.001.

Electronic supplementary material

Supplementary Tables

Table S1. Y-SNP characteristics, primer sequences and analysis conditions.

Y-SNP	Primers (5´-3´)	Method**	Annealing Temperature	Amplicon Size	Mutation (anc/der)	db SNP ID	Y chr position GRCh37
DF27	FW*: TTGGCTGGATATGAAATTCTGGA RV*: GGAAGCCCATCAGATTAACAGA	HRM & Seq HRM & Seq	61	124	G/A	rs577478344	21380200
Z196	FW: AACTGTAAGTCTATGCTGCT RV: ACAGACTGGTTCTGCTTATGT	HRM & Seq HRM & Seq	60.5	HRM: 106 ; Seq: 104	AT/del	–	21033704..21033705
Z195	FW: CATTGCAAGGCTCCAACCAC RV: CGGCATAACCCTGATTTCAACC	HRM & Seq HRM & Seq	61	144	G/A	rs568477247	17922066
Z220	FW: TCTCTAACTTCTGGCTTCAAGTG RV: TGGAATGATATCAGCTTCCATGTC	HRM & Seq HRM & Seq	60	102	G/A	rs538725564	16310705
Z278	FW: AAGGAGAATGATTGCATCAGTC RV: GGAATGTTTGTTTACAATGTGGT	HRM & Seq HRM & Seq	58	156	A/G	rs1469371	18167479
M153	FW: ATTGTCTCCTTTAAGTGGGT RV: TTAATCTGACTTGGAAAGGG	Seq Seq	60.5	113	A/T	rs375151448	21706360
L176.2	FW: ACCCAGTGTTAATTACCCGT RV: GAGCCTCAGGATTCAAAGGA RV2: CTATCATTATTGAGGGCTGGA	HRM & Seq HRM Seq	HRM: 63 ; Seq: 60.5	HRM: 79/84 ; Seq: 109/114	6AAAAC/7AAAAC	–	21779256..21779257
M167	FW: GGAGTGACAACCAAGAAGAG RV: TTTCAAGCTCTGGTTCTGTG	HRM & Seq HRM & Seq	60.5	229	G/A	rs1800865	2658271
S68	FW: TGTCAGATGCTTAATTGTGTTTC RV: CAGGAGTTATGTGAGGACCC FW2: TGCTTGAAACCGAGTTTGTA	HRM HRM & Seq Seq	HRM: 63 ; Seq: 60.5	HRM: 62 ; Seq: 141	C/T	rs775040950	21930315
DF17	FW: ATTAGCCAACTGTAATCTTGGTTA RV: TCTTATTCCATCACCCAGGC	Seq Seq	60	125	T/G	rs754186919	6631746
L617	FW*: /5Biosg/TCATCTTCACCCTTGAGAAGC RV*: TAATGGCAGAAACCATGACAAA S: CCAAGGTGGAGAAGTG	Pyroseq	60/45	84	G/A	–	8466862
L881	FW: TGGCTGTGGCTTTACTTTCTG RV: GCAGGACAACCCTTCTTTGA	HRM & Seq HRM & Seq	60	211	A/G	–	21842521

* These primers are nested primers of a longer amplicon. DF27 was amplified with the primers FW:GGGAATTTGATCCTGTCGTTG and RV:GAACAAAGCCTCCAAGAAATATGAGG [1] with the same annealing temperature. L617 was amplified with the primers FW: AATGGTCTGGTGTTGAAGTGG and RV: GGTTGCGTGAAATAATTGGGT with the same temperature of annealing.

**Seq: Sequencing, HRM: High Resolution Melting, Pyroseq: Pyrosequencing

Table S2. Y-SNP frequencies in the analyzed samples of population. For each haplogroup/column, the higher the frequency, the more intense the colour. * These populations are extracted from the 1000 Genomes Project [19]

Absolute frequencies	N	DF27	DF27*	L617	L881	Z196	Z196*	Z220	Z220*	Z278	M153	DF17	L176.2	L176.2*	M167	S68
Asturias	63	27	19	0	0	8	5	3	1	2	0	0	0	0	0	0
Cantabria	96	43	22	0	0	21	2	16	4	12	0	0	3	1	2	0
Native Basques	229	162	71	4	0	87	7	65	18	47	15	1	14	5	7	2
Resident Basques	111	53	27	0	0	26	5	13	2	11	1	0	8	1	7	0
Aragon	92	32	14	0	0	18	0	7	4	3	1	1	10	5	4	1

Frequencies (%)	N	DF27	DF27*	L617	L881	Z196	Z196*	Z220	Z220*	Z278	M153	DF17	L176.2	L176.2*	M167	S68
Asturias	63	42,86	30,16	0,00	0,00	12,70	7,94	4,76	1,59	3,17	0,00	0,00	0,00	0,00	0,00	0,00
Cantabria	96	44,79	22,92	0,00	0,00	21,88	2,08	16,67	4,17	12,50	0,00	0,00	3,13	1,04	2,08	0,00
Native Basques	229	70,74	31,00	1,75	0,00	37,99	3,06	28,38	7,86	20,52	6,55	0,44	6,11	2,18	3,06	0,87
Resident Basques	111	47,75	24,32	0,00	0,00	23,42	4,50	11,71	1,80	9,91	0,90	0,00	7,21	0,90	6,31	0,00
Aragon	92	34,78	15,22	0,00	0,00	19,57	0,00	7,61	4,35	3,26	1,09	1,09	10,87	5,43	4,35	1,09
Iberian Peninsula*	54	44,44	20,37	0,00	0,00	24,07	0,00	18,52	5,56	12,96	1,85	1,85	3,70	3,70	0,00	0,00
Britain*	46	13,04	6,52	0,00	0,00	6,52	0,00	0,00	0,00	0,00	0,00	2,17	4,35	0,00	4,35	0,00
Toscani in Italy*	53	7,84	2,18	0,00	0,00	5,66	0,00	1,89	1,89	0,00	0,00	3,77	0,00	0,00	0,00	0,00
Finland*	38	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Utah residents (CEPH) with North and Western European ancestry*	49	10,20	6,12	0,00	0,00	4,08	0,00	2,04	2,04	0,00	0,00	0,00	2,04	0,00	2,04	0,00
Colombia*	43	44,19	18,61	0,00	0,00	25,58	0,00	13,95	2,32	11,63	2,33	0,00	11,63	2,33	9,30	0,00
Mexico*	32	9,38	0,00	0,00	0,00	9,38	0,00	3,13	0,00	3,13	0,00	0,00	6,25	0,00	6,25	0,00
Peru*	41	7,32	4,88	0,00	0,00	2,44	0,00	2,44	0,00	2,44	2,44	0,00	0,00	0,00	0,00	0,00
Puerto Rico*	54	35,19	24,08	0,00	0,00	11,11	0,00	7,41	0,00	7,41	1,85	1,85	1,85	1,85	0,00	0,00
Japan*	56	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Vietnam*	46	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Kenya*	44	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Sierra Leona*	42	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
African Caribbean in Barbados*	47	2,13	2,13	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Americans of African ancestry in SW USA*	26	3,85	0,00	0,00	0,00	3,85	0,00	0,00	0,00	0,00	0,00	0,00	3,85	0,00	3,85	0,00

Table S3. Y-SNP and Y-STR haplotypes for all the analyzed samples of population.

Corresponds to Attached Table 2 in Appendix section.

Table S4. Sample size (N), number of haplotypes (K), and haplogroup diversity (HGD) in the populations analyzed. *For HGD*, besides DF27 and its subhaplogroups, M2, M35, P15, M253, P37, M223, M267, M410, M12, M184, M420, M343, M269, U106, P312, U152 and M529 were considered (data not shown).

Population	N	K	HGD	HGD*
Asturias	63	5	0.5842±0.0474	0.8652±0.0291
Cantabria	96	7	0.6086±0.0460	0.8803±0.0247
Native Basques	229	11	0.7890±0.0149	0.8477±0.0137
Resident Basques	111	8	0.6591±0.0373	0.9093±0.0177
Aragon	92	9	0.5499±0.0567	0.8970±0.0175

Table S5. Genetic Fst distances based on Y-SNP haplotypes and p-values. Statistically significant differences after Bonferroni correction are shaded in yellow. Bonferroni: 0.005.

Fst values

	Asturias	Cantabria	Native Basques	Resident Basques	Aragon
Asturias	0.00000				
Cantabria	0.01387	0.00000			
Native Basques	0.11407	0.06055	0.00000		
Resident Basques	0.01476	-0.00384	0.05174	0.00000	
Aragon	0.00817	0.00644	0.10425	0.00569	0.00000

p values

	Asturias	Cantabria	Native Basques	Resident Basques	Aragon
Asturias	*				
Cantabria	0.11454+-0.0034	*			
Native Basques	0.00000+-0.0000	0.00030+-0.0002	*		
Resident Basques	0.09940+-0.0029	0.55697+-0.0054	0.00020+-0.0001	*	
Aragon	0.17127+-0.0037	0.17681+-0.0035	0.00000+-0.0000	0.17563+-0.0044	*

Supplementary Figures

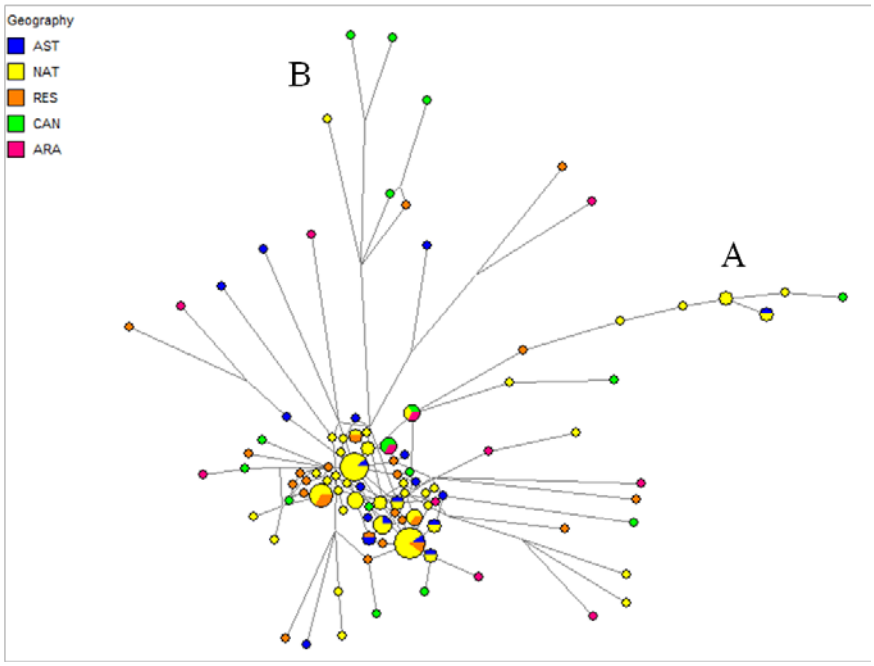


Figure S1. Median Joining Network of individuals belonging to DF27*.

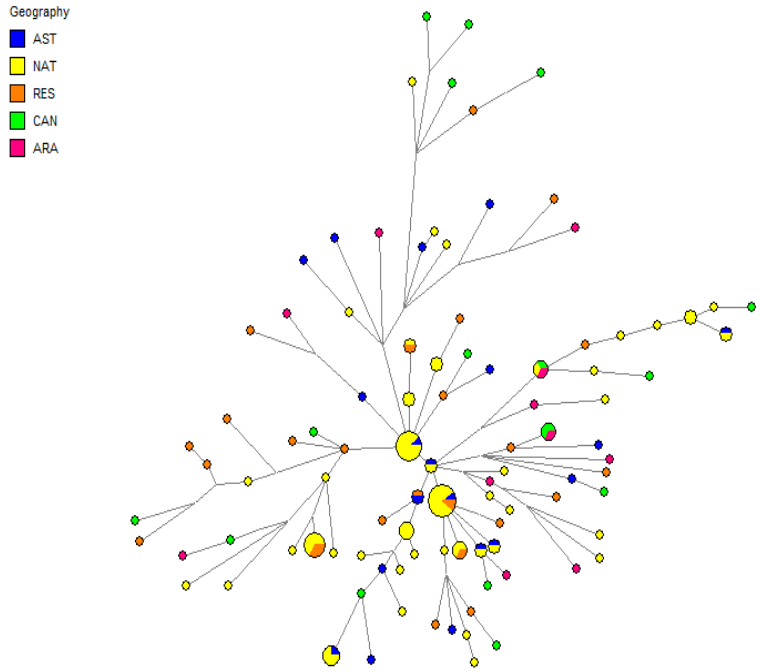


Figure S2. Simplified median joining network of individuals belonging to DF27*.

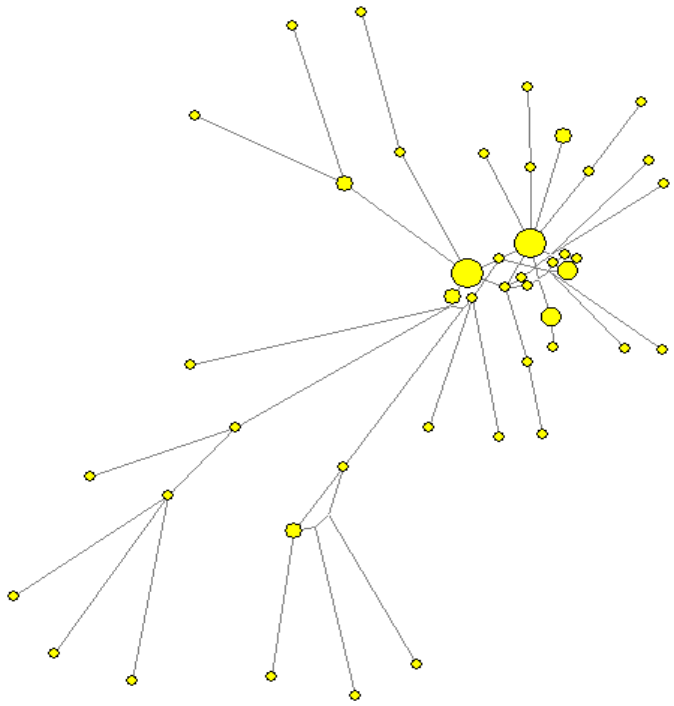


Figure S3. Median joining network of DF27* in the native Basque population.

Forensic Science International: Genetics Supplement Series 5 (2015) e433–e435

Contents lists available at ScienceDirect

Forensic Science International: Genetics Supplement Series

journal homepage: www.elsevier.com/locate/FSIGSS

Supplement series

Dissection of the DF27 paternal lineage

P. Villaescusa, L. Valverde, M.J. Illescas, M.M. de Pancorbo*

BIOMICs Research Group, Lascaray Research Center, University of Basque Country UPV/EHU, Vitoria-Gasteiz, Spain.

*Corresponding Author

Received 29 August 2015, Accepted 20 September 2015, Available online 28 September 2015.

Abstract

The genetic evidence provided by the analysis of the Y chromosome is a valuable tool for the study of the evolution of paternal lineages. The dissection of S116, the major M269 subhaplogroup in Western and South-Western Europe uncovered an outstanding frequency of DF27 sublineage in the Basque region. In this study, a dissection of DF27 haplogroup was performed to the highest resolution to date in 340 individuals from the Basque Country. Our results describe frequency distribution patterns for some DF27 sublineages for the first time, and reveal a possible substructure of its paragroups.

Keywords

Y-SNPs; Paternal lineages; DF27.

1. Introduction

The analysis of Y chromosome SNPs (Y-SNPs) reveals the existence of specific lineages in human population at continental and regional level [1]. Currently a large majority of individuals in Central and Western Europe (40–90%) belong to a single lineage: R-M269 [2].

The dissection of the lineage M269 in European populations has shown a broad range of areas that possess geographically located subhaplogroup expansions. S116 is distributed in western and south-western Europe. Regarding S116 sublineages, DF27 show a high frequency in the Iberian Peninsula while M529 is more common in the British Isles and U152 in northern Italy and the

Alpine region. DF19 and L238 show very low frequencies in Western Europe [2–3]. This provides useful information in Forensic Genetics.

The objective of our study is to dissect the DF27 sublineage in the Basque population, where the DF27 frequency is the highest found to date in Europe. With this aim, the Y-SNPs Z195, Z196, L176.2, M167, S68, S356, M153, DF17, L881 and L617 were analyzed. These new data contribute to know the frequencies of these sublineages and it may be of interest in forensic casework.

2. Material and methods

2.1 Population sample

Blood or saliva samples from autochthonous (N = 229) and resident (N = 111) men were taken in the Basque Country after informed consent.

2.2 Molecular analysis

DNA concentration was adjusted to 1 ng/µL.

The Y-SNPs DF27, Z195, Z196, L176.2, M167, S68, S356 and L881 were analyzed by high resolution melting (HRM). M153 and DF17 were typed by Sanger sequencing. Finally, L617 was analyzed by pyrosequencing. The amplification, melting, sequencing and pyrosequencing conditions are described in www.BiomicsResearchGroup.net.

2.3 Statistical analysis

The absolute and relative frequencies for each SNP were manually estimated. Frequencies of Y-STR haplotypes [4] were calculated using Arlequin v 3.1 software [5]. The phylogenetic relationships of Y-STR haplotypes were estimated using Network v 4.6.1.3 [6]. Phylogenetic weights were assigned in a manner inversely proportional to observed mutations.

3. Results and discussion

The observed frequencies are shown in Table 1.

The dissection of this haplogroup reveals Z196 as the main sublineage of DF27 in Basque population, with frequencies up to 38%. In autochthonous Basques the frequency was quite higher (38%) than in non autochthonous population (23%), being a common trend the differences in frequency between the two samples of this population. This is consistent with the previous data of Valverde et al. [7], which places the origin of S116 in the Iberian Peninsula.

Z196 is divided in the subhaplogroups S356 (the most frequent), L176.2 and DF17. Compared to previous studies the frequencies of M153 and M167 sublineages of S356 are lower than in other Pyrenean populations [8].

Table 1

Frequencies of Y-SNP lineages (%) in the analyzed population samples.

Y-SNPs	Autochthonous Basques $N = 229$	Residents non autochthonous Basques $N = 111$
DF27	70.74	47.75
DF27*	31.00	24.32
L617	1.75	0.00
L881	0.00	0.00
Z196	37.99	23.42
Z196*	3.06	4.50
S356	28.38	11.71
S356*	21.83	10.81
M153	6.55	0.90
DF17	0.44	0.00
Z196 (xL176.2)	31.88	16.22
L176.2	6.11	7.21
L176.2*	2.18	0.90
M167	3.06	6.31
S68	0.87	0.00

The frequency of DF27* was quite higher in the Basque region, showing frequencies between 31% and 24% in the two population samples. In the same sense, the paragroup of S356 (S356*) draws attention because of its striking frequency. The high frequencies of DF27* and S116* paragroups indicate a probable existence of new subhaplogroups supporting the subdivision of both of them.

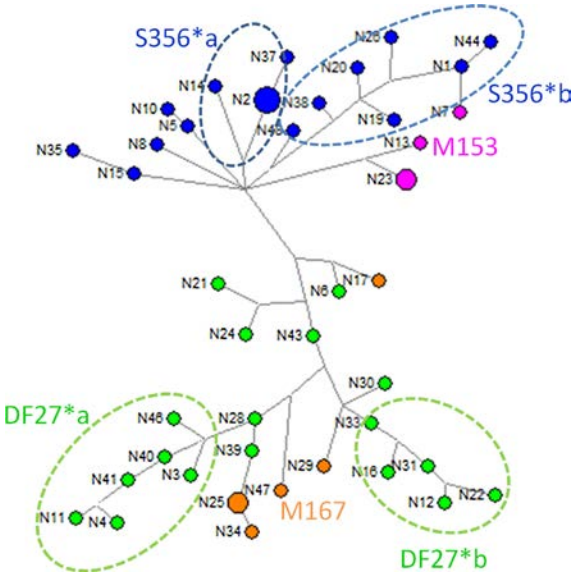


Fig. 1. MJN of the DF27 haplogroup in the autochthonous Basque population sample. Two phylogenetic splits into DF27* and S356* paragroups can be observed.

This hypothesis was supported by the structure of the median joining network (MJN) (Fig. 1). The MJN showed two main groups corresponding the individuals belonging to DF27* and S356* haplogroups characterized by the Y-STR haplotypes DYS437/DYS448 bearing the alleles 15/19 and 14/18, respectively. Moreover, DF27 lineage appears to be split in two groups that differ in the

locus DYS391. DF27*a samples bear the allele 10 whereas DF27*b samples share the allele 11. On the other hand, S356* shows two principal groups, S356*a and S356*b that bear the alleles 13/12 and 14/13 in the haplotype DYS393/DYS438.

4. Conclusion

Our results reveal a high frequency of DF27 lineage in Basque population that may be an indicator or a possible Iberian origin of this lineage and its sublineages. The high proportion of individuals in the paragroups DF27* and S356*, as well as the subdivision of DF27* and S356* haplotypes, makes clear the need to continue the searching of new Y-SNPs in order to attain a better resolution of their respective subhaplogroup.

Funding

Funds were provided by the Basque Government (Grupo Consolidado IT833-13). MJ and PV received predoctoral grants from the University of the Basque Country.

Conflict of interest

None

Acknowledgements

The authors are grateful to the Basque Foundation of Science (BIOEF) for samples and to PhD Maite Alvarez for her technical and human support provided by the DNA Bank Service (SGIker—UPV/EHU) and to all the people who voluntarily participated in this study.

References

[1] P.A. Underhill, P. Shen, A.A. Lin, et al., Y chromosome sequence variation and the history of human populations, Nat. Genet. 26 (2000) 358–361.

[2] G.B.J. Busby, F. Brisighelli, P. Sánchez-Diz, et al., The peopling of Europe and the cautionary tale of Y chromosome lineage R-M269, Proc. Biol. Sci. 279 (2012) 884–892.

[3] L. Valverde, M.J. Illescas, P. Villaescusa, et al., New clues to the evolutionary history of the main European paternal lineage M269: dissection of the Y-SNP S116 in Atlantic Europe and Iberia, Eur. J. Hum. Genet. (June) (2015), doi:http://dx.doi.org/10.1038/ejhg.2015.114 (Epub ahead of print).

[4] C. Nuñez, M. Baeta, M. Fernández, et al., Highly discriminatory capacity of the PowerPlex® Y23 System for the study of isolated populations, Forensic Sci. Int. Genet. 17 (2015) 104–107.

[5] L. Excoffier, G. Laval, S. Schneider, Arlequin (version 3.0): an integrated software package for population genetics data analysis, Evol. Bioinform. Online 1 (2005) 47–50.

[6] H.J. Bandelt, P. Forster, A. Röhl, Median-joining networks for inferring intraspecific phylogenies, Mol. Biol. Evol. 16 (1999) 37–48.

[7] L. Valverde, M. Rosique, S. Köhnemann, et al., Y-STR variation in the Basque diaspora in the Western USA: evolutionary and forensic perspectives, Int. J. Legal Med. 126 (2012) 293–298.

[8] A.M. López-Oarra, L. Gusmão, L. Tavares, et al., In search of the pre- and postneolithic genetic substrates in Iberia: evidence from Y-chromosome in Pyrenean populations, Ann. Hum. Genet. 73 (2009) 42–53.

4.3 Study Number 3

'Analysis of the R1b-DF27 haplogroup shows that a large fraction of the Iberian Y-chromosome lineages originated recently in situ'

The third study of the present work corresponds to the attainment of the second part of the objective 2: *To characterize the structure and spatial distribution of the Iberian near-specific paternal lineage R1b-DF27 in Southwest European populations through the dissection in its sublineages, with the aim to estimate its time of origin, as well as to model its expansion in the phylogenetic context and the related demographic events.*

To properly characterize the distribution and structure of a lineage it is required to analyze a large number of individuals from several locations, which is of utmost importance in the field of forensic genetics where the ancestry of an evidence could be linked to a geographic location. In the case of the paternal lineage R1b-DF27, the preceding findings unveiled the interest in improving the coverage of the Iberian Peninsula in addition to other Southwest European populations to get a detailed view of its distribution, structure and origin.

The present work has a dual aim. On the one hand, to further characterize the special distribution of DF27 haplogroup by analyzing more populations. On the other hand, to estimate its time of origin and model its expansion considering both its phylogenetic context and the demographic events that influenced the genetic variability of West Europe around 4,500 years ago. For that purpose, a total of 1,072 male individuals were genotyped in 26 populations from Spain, Portugal, and France for the Y-SNPs M269, S116, DF27, Z195, L176.2, M167, Z220, Z278, and M153. Basic descriptive statistics, principal component analysis (PCA), AMOVA, and *times to the most recent common ancestor* (TMRCA) were calculated. We used Approximate Bayesian Computing (ABC) to test alternative models of demographic expansion for DF27 and to estimate their parameters.

The findings of this study complete the distribution landscape of DF27, reconfirming its presence in Iberian populations with frequencies around 40%, as previously observed, and adding information regarding France, where the frequencies drop to 6-20%. The analysis of DF27 sublineages reveal certain degree of geographic structure, since the subhaplogroup L176.2 is more frequent in East Iberia and Z220, in turn, peaks in North-Central Iberia. These domains could be reminiscent of previous partitions of the Iberian region, such as the pre-roman Iberian/Celtic division of the Iberian Peninsula, or of the Christian Kingdoms in the Middle Ages. By and large, the age of DF27 was estimated at 4,200 years ago, at the transition between the Neolithic and the Bronze Age. Regarding the place of origin of DF27, North-East Iberia (Basque Country, Aragon) is

the most likely option considering the frequencies and the Y-STR internal diversity. Finally, the Approximate Bayesian Computing (ABC) results suggest that the demography of DF27 is more compatible with population growth than with stationarity, being the start of this growth closer to the TMRCA of the haplogroup.

In conclusion, the present study has contributed valuable genetic data for the understanding of the phylogeography of DF27. However, a global characterization of the whole sequence diversity of this haplogroup as well as sampling more locations in Atlantic Iberia would be desirable to obtain a more accurate picture of DF27 haplogroup.

This study has resulted in an international publication in the journal *Scientific Reports* under the heading '*Analysis of the R1b-DF27 haplogroup shows that a large fraction of the Iberian Y-chromosome lineages originated recently in situ*' in August 2017. Q1, IP: 4.122. The publication is shown below.

www.nature.com/scientificreports

SCIENTIFIC REPORTS

SCIENTIFIC REPORTS | 7: 7341 | DOI:10.1038/s41598-017-07710-x

Article

Analysis of the R1b-DF27 haplogroup shows that a large fraction of Iberian Y-chromosome lineages originated recently in situ

Neus Solé-Morata[1], Patricia Villaescusa[2], Carla García-Fernández[1], Neus Font-Porterias[1], María José Illescas[2], Laura Valverde[2], Francesca Tassi[3], Silvia Ghirotto[3], Claude Férec[4,5,6,7], Karen Rouault[4,5], Susana Jiménez-Moreno[8], Begoña Martínez-Jarreta[9], Maria Fátima Pinheiro[10], María T. Zarrabeitia[11], Ángel Carracedo[12,13], Marian M. de Pancorbo[2] & Francesc Calafell[1]

[1]Institut de Biologia Evolutiva (CSIC-UPF), Departament de Ciències Experimentals i de la Salut, Universitat Pompeu Fabra, Barcelona, Catalonia, Spain.

[2]BIOMICs Research Group, Lascaray Research Center, University of the Basque Country UPV/EHU, Vitoria-Gasteiz, Spain.

[3]Dipartimento di Scienze della Vita e Biotecnologie, Università di Ferrara, Ferrara, Italy.

[4]Inserm, UMR 1078, Brest, France.

[5]Laboratoire de Génétique Moléculaire, CHRU Brest, Hôpital Morvan, Brest, France.

[6]Université de Bretagne Occidentale, Brest, France.

[7]Etablissement Français du Sang-Bretagne, Brest, France.

[8]Forensic and Legal Medicine Area, Department of Pathology and Surgery, University Miguel Hernández, Elche, Spain.

[9]Department of Forensic Medicine, University of Zaragoza, Zaragoza, Spain.

[10]Forensic Genetics Department, National Institute of Legal Medicine and Forensic Sciences, Porto, Portugal.

[11]Unit of Legal Medicine, University of Cantabria, Santander, Spain. 12Genomic Medicine Group, CIBERER- University of Santiago de Compostela, Galician Foundation of Genomic Medicine (SERGAS), Santiago de Compostela, Spain.

[13]Center of Excellence in Genomic Medicine Research, King Abdulaziz University, Jeddah, Saudi Arabia.

Correspondence and requests for materials should be addressed to F.C.

Received 6 April 2017, Accepted 28 June 2017, Published online 04 August 2017.

Abstract

Haplogroup R1b-M269 comprises most Western European Y chromosomes; of its main branches, R1b-DF27 is by far the least known, and it appears to be highly prevalent only in Iberia. We have genotyped 1072 R1b-DF27 chromosomes for six additional SNPs and 17 Y-STRs in population samples from Spain, Portugal and France in order to further characterize this lineage and, in particular, to ascertain the time and place where it originated, as well as its subsequent dynamics. We found that R1b-DF27 is present in frequencies ~40% in Iberian populations and up to 70% in Basques, but it drops quickly to 6–20% in France. Overall, the age of R1b-DF27 is estimated at ~4,200 years ago, at the transition between the Neolithic and the Bronze Age, when the Y chromosome landscape of W Europe was thoroughly remodeled. In spite of its high frequency in Basques, Y-STR internal diversity of R1b-DF27 is lower there, and results in more recent age estimates; NE Iberia is the most likely place of origin of DF27. Subhaplogroup frequencies within R1b-DF27 are geographically structured, and show domains that are reminiscent of the pre-Roman Celtic/Iberian division, or of the medieval Christian kingdoms.

Introduction

Although it contains ~1% of the genome length in a human male cell, the lack of recombination along most of the Y chromosome makes constructing phylogenies for genetic variation relatively easy. Coupled with a robust geographic differentiation, this trait has provided a comprehensive phylogeography of Y chromosome haplotypes (usually referred to as haplogroups), that has been thoroughly characterized. Thus, the origin, dispersal, and geographic spread of many haplogroups are known. Moreover, both the genotyping of fast-mutating short tandem repeats (STRs) in the non-recombining region of the Y chromosome (NRY), and the recent availability of ascertainment-bias-free whole sequences of the NRY have reliably added a temporal scale to the deployment of the Y-chromosome diversity. One of the most salient features of the recent evolutionary history of human Y chromosomes is that it seems to have happened in bursts, with haplogroups rising to high frequency in the wake of major lifestyle shifts and technological innovations such as the advent of the Neolithic or the recently acknowledged demographic upheaval caused by the Bronze Age in Europe[1, 2].

The most frequent Y-chromosome haplogroup in W Europe is R1b-M269, with frequencies ranging from 41% (Germany) to 83% (Ireland)[3]. Precisely, the higher frequency of this haplogroup in W Europe rather than in E Europe or W Asia led previous authors to believe it had a post-glacial Palaeolithic origin[4, 5]; however, a larger STR variance in SE European and W Asian R1b-M269 chromosomes and direct TMRCA dating pointed to R1b-M269 having surfed the Neolithic wave of

advance[3, 6], the evidence for which other authors did not find conclusive[7]. Finally, direct dating from NRY sequence variation puts the origin of R1b-M269 in the Early Bronze Age, ~4500 years ago (ya)[1, 8], consistent with the growing ancient DNA record, where a surge in R1b-M269 is indeed seen at that time[2, 9]. Note, though, that R1b-M415, a branch ancestral to R1b-M269, was found as early as 14,000 ya in Italy[10] and 7,000 ya in Spain[2]. Moreover, lack of structure of STR variation within R1b-M269[11, 12] points also to an explosive growth.

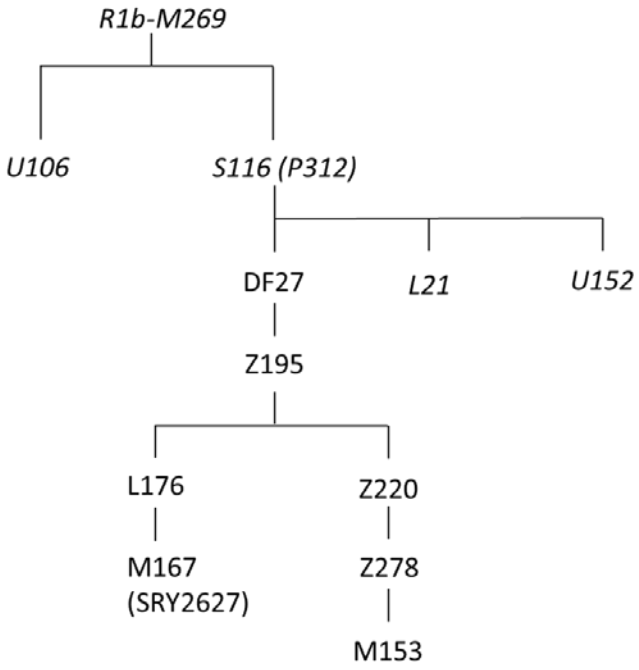


Figure 1. Simplified phylogenetic tree of the R1b-M269 haplogroup. SNPs in italics were not analyzed in this manuscript.

The most important branches of R1b-M269 are R1b-U106, particularly frequent in the Low Countries and NW Germany[3, 13], and R1b-S116 (also known as R1b-P312), which is common throughout W Europe[3]. The latter trifurcates in turn into U152 (frequent in N Italy and Switzerland[13]), L21 (also known as M529, abundant in the British Isles[7]), and DF27 (Fig. 1; Supplementary Figure 1). DF27 was first discovered by citizen scientists[14] and, although among the burgeoning amateur genetic genealogy it is known to be frequent in Iberian populations and their overseas offshoots, few academic publications have been devoted to it. It was found in the 1000 Genome Project populations at a frequency of 49% in Iberians, 6% in Tuscans, 7% in British, and it was absent elsewhere except for admixed populations in the Americas: Colombia (40%), Puerto Rico (36%), Mexico (10%), Perú (8%), African-Americans (4%) and Afro-Caribbeans (2%)[14, 15]. It was first genotyped specifically in a few Iberian populations, Brittany and Ireland as part of a study on R1b-S116[16], which indeed confirmed that R1b-DF27 is present at frequencies >40% in Spain and Portugal. Subsequently, 12 SNPs within DF27 were genotyped in four N Spanish

populations[17], confirming its high frequency and hinting at some substructure within the Iberian Peninsula. As for its presence elsewhere, the frequency of R1b-S116 (xL21, U152) can be used as an upper bound for the frequency of R1b-DF27. R1b-S116 (xL21, U152) was found at frequencies 0–10% in Germany[3, 18], 7% in the Netherlands[3], 8–12% in Flanders[19], 6–12% in Switzerland[3], and 1–12% in Italy[3, 20].

Here, by extending population sampling to cover France as well as by improving coverage in the Iberian Peninsula, we aim to i) further characterize the spatial distribution of R1b-DF27, ii) estimate its time of origin, and iii) explicitly model its expansion in relation both to its phylogenetic context, and to the demographic events that thoroughly reshaped the genetic diversity of W Europe around 4500 ya.

Results

Over one thousand individuals carrying DF27 were typed for six additional SNPs (Table 1, Fig. 1) and 17 Y-STRs. DF27 itself was found at frequencies 0.3–0.5 in Iberia (with a mean of 0.42), with the notable exception of native Basques, where it reached 0.74 (for this and all subsequent frequency values, see Fig. 2 and Supplementary Table 1). In France, it dropped to a range of 0.06–0.20 and a mean of 0.11. Elsewhere, it was 0.15 in Britain (but <0.01 in Ireland) and 0.08 in Tuscany. Most (50–100%, with a proportion that dropped from East to West) DF27 Y chromosomes were also derived for Z195; thus, the highest frequencies of Z195 (0.29–0.41) were reached both in the Basque Country and in E Iberia (Catalonia, Valencia), and it becomes as rare in Portugal as it is in France. Conversely, the highest frequencies of R1b-DF27* (xZ195) are found in Native Basques and Western Iberian populations such as Asturias, Portugal and Galicia, which may harbor yet unknown branches of R1b-DF27. In turn, Z195 splits into two branches, namely L176.2 and Z220 (Fig. 1). Note that L176 is a recurrent mutation that defines two clades in the Y phylogeny: L176.1 within R1a, and L176.2 under R1b-DF27; throughout this manuscript, we will refer exclusively to the latter. L176.2 and Z220 peak, respectively, in E Iberia and the Basque Country. L176.2 is further subdivided into M167 (SRY2627, ref. 21), with the highest frequencies in Catalonia and the lands settled from Catalonia in the 13th century (Valencia, the Balearics). This marker had been typed in a number of Iberian and other European populations[4, 18–20, 22–25], and the overall frequency pattern found (Supplementary Figure 1) confirms a distribution centered in the eastern half of Iberia, although with higher frequencies (up to 0.16) in the upper Ebro river valley and the Pyrenees. As mentioned above, Z220 is most frequent in the Basque Country (0.28), and a similar pattern is found for its successive nested clades, namely Z278 and M153. For the latter, available additional data[22, 23, 25] showed it confined to the Iberian Peninsula, with frequencies

0.06–0.40 among Basque subpopulations, but rarely above 0.01 elsewhere (Supplementary Figure 1).

Population	abbr.	Region	lon	lat	Total N	Source	N DF27 typed	Additional R1b (1)	Additional P312 (2)	Additional Z195 (3)	SNP typing	STR typing
Alacant	ALA	Valencia	-0.56	38.36	142	(113)[16], (29)[27]	22	0	0	0	this work[26, 27]	[27], [54]
Alsace	ALS	France	7.75	48.58	80	[39]	6	0	0	0	this work[39]	[39]
Andalucía	AND	NA	-6	37.5	100	[16]	47	0	0	0	this work[16]	NA
Aragón	ARA	Aragón	-0.86	41.63	92	[17]	34	0	0	0	[17]	[26]
Asturias	AST	N. Central Spain	-5.87	43.34	63	[16]	27	0	0	0	[17]	[26]
Auvergne	AUV	France	3.09	45.78	89	[39]	5	0	0	0	this work[39]	[39]
Barcelona	BCN	Catalonia	2.13	41.4	571	(99)[16], (472)[27]	245	11	8	6	this work[16, 27]	[26], [27]
Brittany	BRI	NA	-4.49	48.39	145	[16]	28	0	0	0	this work[16]	NA
Cantabria	CAN	N. Central Spain	-3.83	43.35	96	[16]	43	0	0	0	[16], [17]	[41]
Castelló	CAS	Valencia	-0.06	40	49	[27]	22	2	0	1	this work[27]	[27]
Galicia	GAL	NA	-8.55	42.88	70	[16]	28	0	0	0	this work[16]	NA
Girona	GIR	Catalonia	2.81	41.99	131	[27]	42	2	3	0	this work[27]	[27]
Île-de-France	IDF	France	2.35	48.86	91	[39]	9	2	0	0	this work[39]	[39]
Ireland	IRE	NA	-8	53	146	[16]	1	0	0	0	this work[16]	NA
Lleida	LLE	Catalonia	0.62	41.62	104	[27]	52	2	0	0	this work[27]	[27]
Madrid	MAD	NA	-3.72	40.42	99	[16]	49	0	0	0	this work[16]	NA
Mallorca	MAL	Mallorca	3	39.63	48	[27]	24	0	2	0	this work[27]	[27]
Midi-Pyrénées	MPY	France	1.45	43.61	67	[39]	7	1	0	0	this work[39]	[39]
Native Basque	NBA	Basques	-2.47	43.17	229	[16]	169	0	0	0	[16], [17]	[16]
Nord-Pas-de-Calais	NPC	France	3.04	50.63	68	[39]	8	3	0	0	this work[39]	[39]
Portugal	POR	NA	-8.5	39.5	109	[16]	44	0	0	0	this work[16]	NA
Provence–Alpes-Côte d'Azur	PAC	France	5.45	43.29	45	[39]	5	3	0	0	this work[39]	[39]
Pyrenees	PYR	Catalonia	1.12	42.49	46	[27]	24	0	1	0	this work[27]	[27]
Resident Basque	RBA	N. Central Spain	-3.69	42.35	111	[16]	53	0	0	0	[16], [17]	[16]
Tarragona	TAR	Catalonia	1.17	41.12	120	[27]	47	0	1	1	this work[27]	[27]
Valencia	VAL	Valencia	-0.4	39.48	79	[27]	31	3	2	1	this work[27]	[27]
Total					2990		1072	29	17	9		

Table 1. Populations sampled and their original sources. Region: higher population grouping used in some analyses. (1): Chromosomes predicted to be R1b but without further SNP typing; (2): Chromosomes known to be R1b-P312* (xZ195, U152, L21) but without further typing; (3): Chromosomes known to be R1b-Z195 (xM167, Z220), but not typed for L176.

The subhaplogroup frequencies were summarized in a PC plot (Fig. 3). The first PC separated the Iberian populations (save for the three westernmost samples, namely Portugal, Galicia, and Asturias) from the rest, explained 68.6% of the total variation, and was positively correlated with DF27 and all of its subhaplogroups. On the contrary, PC2 (20.9%) was positively correlated with L176.2 and M167 and most negatively correlated with Z278 and M153, and separated most Eastern Iberian populations from the rest.

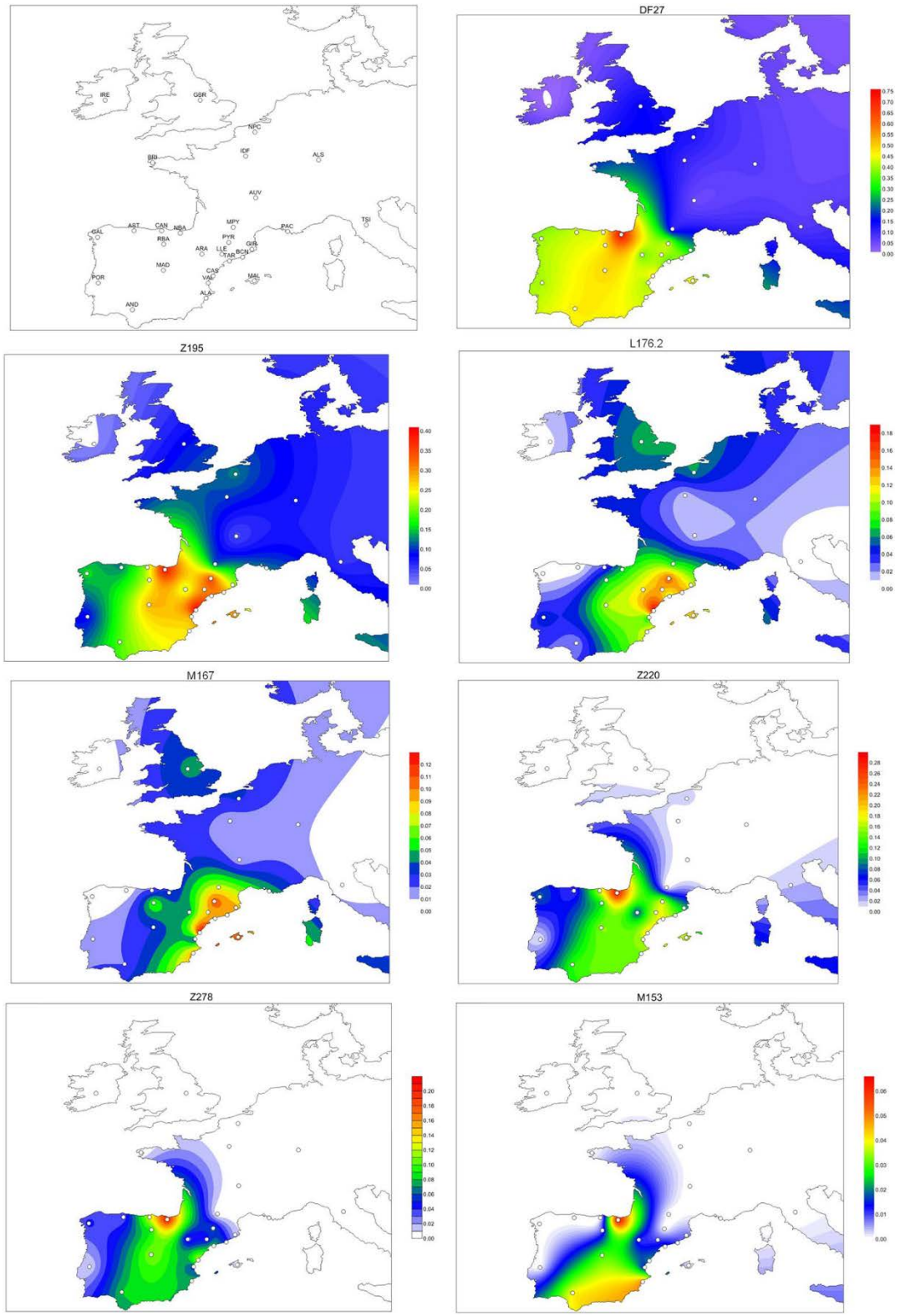


Figure 2. Contour maps of the derived allele frequencies of the SNPs analyzed in this manuscript. Population abbreviations as in Table 1. Maps were drawn with SURFER v. 12 (Golden Software, Golden CO, USA).

In order to quantify the structure of subhaplogroup frequencies, we performed AMOVA with several population groupings. Thus, if we compared Iberian populations vs. the rest, the proportion of the variance explained by the differences among these two groups (i.e., F_{CT}) was

12.40% ($p < 10^{-4}$), while the proportion of the variance found within groups (i.e., F_{SC}) was 3.20% ($p < 10^{-4}$). If the native Basques were split from the Iberians, then F_{CT} = 13.57% ($p < 10^{-4}$) and F_{SC} = 1.37% ($p < 10^{-4}$). Finally, if Eastern Iberians are also split from the rest of Iberians, then F_{CT} = 11.68% ($p < 10^{-4}$) and F_{SC} = 0.31% (p = 0.0106). In conclusion, the differences among the groups that are apparent in the PCA plot are highly statistically significant.

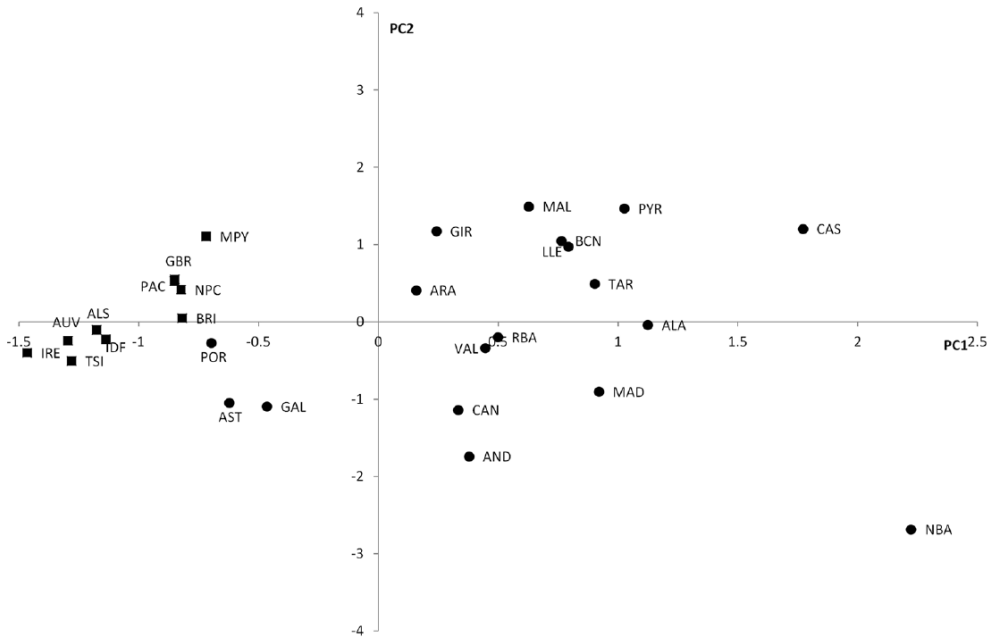


Figure 3. Principal component analysis of subhaplogroup frequencies. Population abbreviations as in Table 1. Circles: Iberian populations; squares: non-Iberian populations.

Haplotypes comprising 17 Y-STRs were available for 758 individuals (Table 2). AMOVA among this set of populations gave R_{ST} = 0.72% (p = 0.00386), while, for the same populations, subhaplogroup frequencies yielded F_{ST} = 8.33% ($p < 10^{-4}$). Thus, Y-STRs seem to capture much less phylogeographic structure than SNPs themselves, as described for R1b-M269[11, 12]. Still, some Y-STR structure may be present within R1b-DF27[12, 17]. Since a median-joining tree with 688 different haplotypes is unmanageable, we resorted to principal component analysis (PCA) among haplotypes (Fig. 4). The first PC explained 15.1% of the STR variation and correlated mostly with DYS437 (r = 0.865), DYS448 (r = 0.858), and YGATAH4 (r = 0.724), and separated haplotypes that carried the derived allele for Z220 (that is, belonging to R1b-Z220*, R1b-Z278* and R1b-M153) from the rest. PC1 coordinates were highly significantly different by subhaplogroup ($p \sim 10^{-150}$, ANOVA). The median haplotype for Z220-derived chromosomes was 11-14-18 at YGATAH4-DYS437-DYS448 while it was 12-15-19 for the rest of DF27 chromosomes (other STRs showed the same median allele). PC2 explained 8.6% of the STR variation and correlated with DYS390 (r = 0.647) and DYS456 (r = −0.544), and separated R1b-Z220* from R1b-Z278* chromosomes; overall,

R1b-DF27 subhaplogroups had significantly different PC2 coordinates ($p = 9.12 \times 10^{-4}$). The same PC results can also be analyzed by population (Fig. 4b): PC1 coordinates are statistically significantly different by population (ANOVA, p = 0.00512), with the French samples having higher, positive values in this PC, while PC2 is not significant across populations (p = 0.781). These results can be explained by the very low frequency of Z220-derived chromosomes outside of Iberia.

Population	N	K	Dhap	Var	sd (Var)
All	758	688	0.9996	0.330	0.215
Alacant	57	56	0.9994	0.377	0.400
Alsace	5	5	1	0.260	0.292
Aragón	29	29	1	0.372	0.218
Asturias	26	26	1	0.283	0.274
Auvergne	5	5	1	0.187	0.125
Native Basques	154	122	0.9951	0.282	0.174
Barcelona	184	178	0.9995	0.348	0.226
Cantabria	30	30	1	0.403	0.364
Castelló	23	23	1	0.294	0.213
Girona	33	33	1	0.306	0.183
Île-de-France	8	8	1	0.323	0.204
Lleida	39	39	1	0.314	0.187
Mallorca	21	21	1	0.341	0.280
Midi-Pyrénées	7	7	1	0.308	0.314
Nord-Pas-de-Calais	7	7	1	0.432	0.459
Provence–Alpes-Côte d'Azur	3	3	1	0.156	0.278
Pyrenees	17	17	1	0.443	0.328
Resident Basques	49	49	1	0.285	0.188
Tarragona	38	38	1	0.351	0.276
Valencia	23	23	1	0.337	0.152

Table 2. Diversity parameters for STR variation within R1b-DF27 chromosomes. K: number of different haplotypes; Dhap: haplotype diversity; Var: average STR allele repeat size variance; sd: standard deviation across loci of Var.

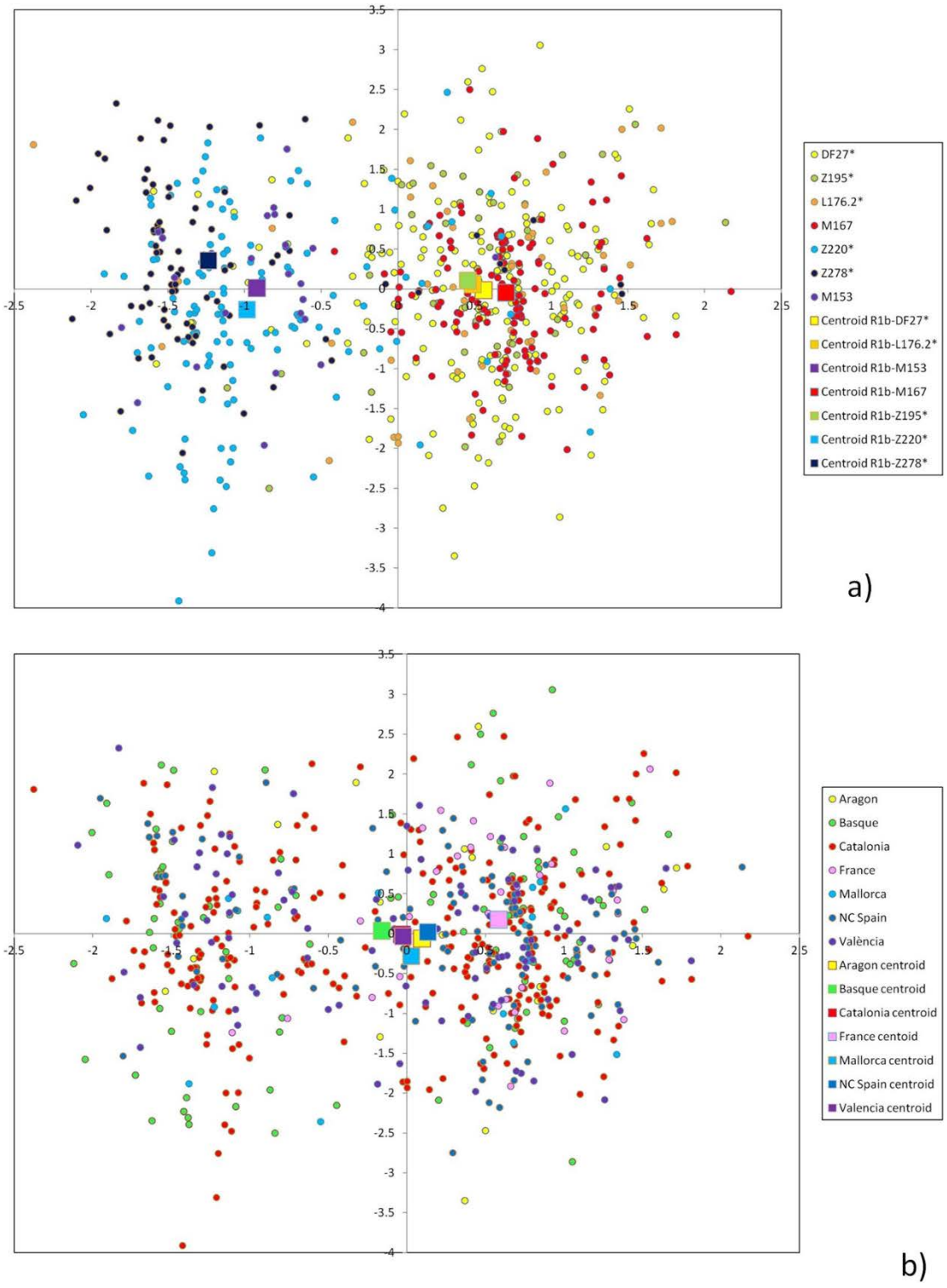


Figure 4. Principal component analysis of STR haplotypes. (a) Colored by subhaplogroup, (b) colored by population. Larger squares represent subhaplogroup or population centroids.

Internal diversity and ages of DF27 and its derived subhaplogroups

The average STR variance of DF27 and each subhaplogroup is presented in Suppl. Table 2. As expected, internal diversity was higher in the deeper, older branches of the phylogeny. If the same

diversity was divided by population, the most salient finding is that native Basques (Table 2) have a lower diversity than other populations, which contrasts with the fact that DF27 is notably more frequent in Basques than elsewhere in Iberia (Suppl. Table 1). Diversity can also be measured as pairwise differences distributions (Fig. 5). The distribution of mean pairwise differences within Z195 sits practically on top of that of DF27; L176.2 and Z220 have similar distributions, as M167 and Z278 have as well; finally, M153 shows the lowest pairwise distribution values. This pattern is likely to reflect the respective ages of the haplogroups, which we have estimated by a modified, weighted version of the ρ statistic (see Methods).

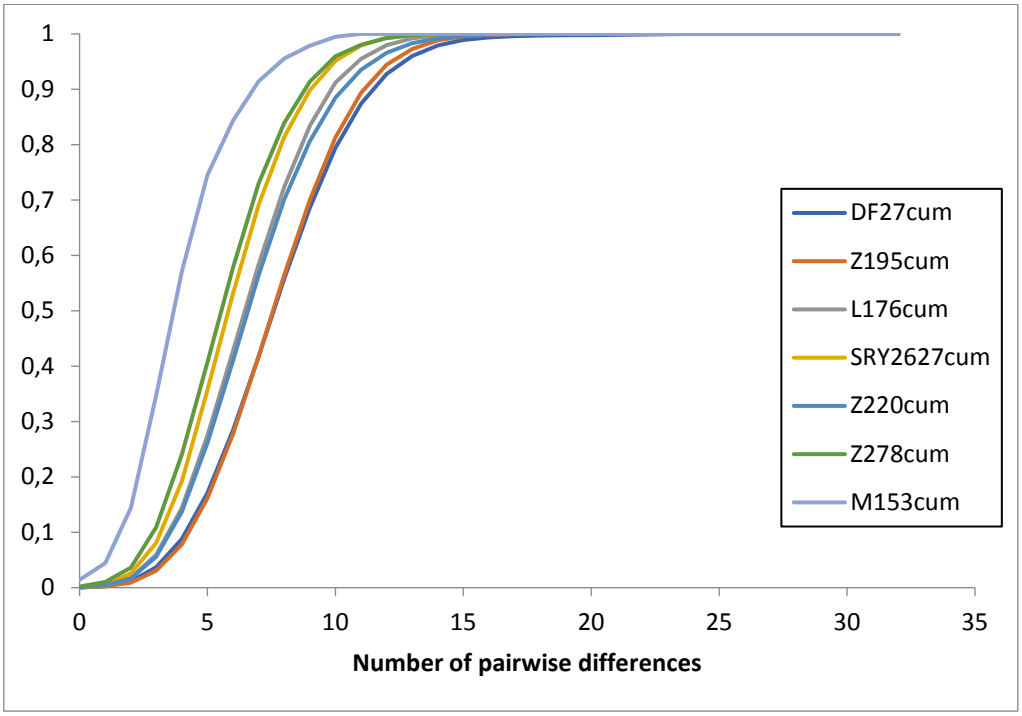


Figure 5. Cumulative distributions of the number of pairwise absolute differences in repeat size among individuals, by subhaplogroup.

We estimated an age of 4190 ± 140 ya for the whole of DF27. This figure is remarkably similar both to the estimate (4128 ± 71 ya) that can be produced from whole Y-chromosome sequence variability in the 88 DF27-derived individuals present overall in the 1000 genomes project dataset, and to the age estimated from 201 individuals in our dataset for which 21 non-duplicated Y-STRs from the Powerplex Y23 System were available[26] (3880 ± 165).

Z195 seems to have appeared almost simultaneously within DF27, since its estimated age is actually older (4570 ± 140 ya). Of the two branches stemming from Z195, L176.2 seems to be slightly younger than Z220 (2960 ± 230 ya vs. 3320 ± 200 ya), although the confidence intervals slightly overlap. M167 is clearly younger, at 2600 ± 250 ya, a similar age to that of Z278 (2740 ± 270 ya). Finally, M153 is estimated to have appeared just 1930 ± 470 ya.

Haplogroup ages		N	age	95% CI
DF27	total	758	4194	(4055, 4333)
	Aragón	29	4527	(3831, 5223)
	Basques	154	3931	(3617, 4245)
	Catalonia	311	4367	(4163, 4571)
	France	35	3428	(2911, 3945)
	Mallorca	21	4091	(3360, 4822)
	N. Central Spain	105	3614	(3293, 3935)
	Valencia	103	4525	(4166, 4884)
Z195	total	510	4569	(4398, 4740)
	Aragón	18	4177	(3293, 5061)
	Basques	83	3257	(2869, 3645)
	Catalonia	246	4584	(4349, 4819)
	France	34	3450	(2923, 3977)
	Mallorca	14	4028	(3081, 4975)
	N. Central Spain	44	3917	(3407, 4427)
	Valencia	71	4541	(4114, 4968)
L176.2	total	189	2964	(2737, 3191)
	Aragón	10	2759	(1626, 3892)
	Basques	14	3059	(2222, 3896)
	Catalonia	108	2931	(2651, 3211)
	France	15	3016	(2273, 3759)
	Mallorca	7	1465	(440, 2490)
	N. Central Spain	8	2365	(1465, 3265)
	Valencia	27	2933	(2370, 3496)
M167	total	137	2602	(2351, 2853)
	Aragón	4	3458	(1919, 4997)
	Basques	7	1221	(217, 2225)
	Catalonia	81	2597	(2289, 2905)
	France	12	2812	(2010, 3614)
	Mallorca	6	1626	(466, 2786)
	N. Central Spain	6	2050	(1082, 3018)
	Valencia	21	2404	(1818, 2990)
Z220	total	267	3318	(3114, 3522)
	Aragón	7	1956	(1080, 2832)
	Basques	62	2693	(2291, 3095)
	Catalonia	123	3380	(3094, 3666)
	Mallorca	7	3198	(2079, 4317)
	N. Central Spain	27	3718	(3105, 4331)
	Valencia	40	3683	(3166, 4200)
Z278	total	130	2745	(2475, 3015)
	Aragón	3	1716	(464, 2968)
	Basques	46	2298	(1877, 2719)
	Catalonia	41	2425	(1994, 2856)
	N. Central Spain	20	3417	(2733, 4101)
	Valencia	20	3084	(2369, 3799)
M153	total	34	1926	(1454, 2398)
	Basques	15	1582	(933, 2231)
	Catalonia	9	1202	(438, 1966)
	Valencia	8	2504	(1393, 3615)

Table 3. Haplogroup ages estimated from STR variation with the weighted ρ method. 95% CI: 95% confidence interval.

Haplogroup ages can also be estimated within each population, although they should be interpreted with caution (see Discussion). For the whole of DF27, (Table 3), the highest estimate was in Aragon (4530 ± 700 ya), and the lowest in France (3430 ± 520 ya); it was 3930 ± 310 ya in Basques. Z195 was apparently oldest in Catalonia (4580 ± 240 ya), and with France (3450 ± 269

ya) and the Basques (3260 ± 198 ya) having lower estimates. On the contrary, in the Z220 branch, the oldest estimates appear in North-Central Spain (3720 ± 313 ya for Z220, 3420 ± 349 ya for Z278). The Basques always produce lower estimates, even for M153, which is almost absent elsewhere.

R1b-DF27 and demography

We tested the dynamics of R1b-DF27 by means of Approximate Bayesian Computing (ABC). In particular, we compared two simple models: constant population size vs. growth since time Tstart. Both the rejection and the regression method undoubtedly favoured the growth model, with associated posterior probabilities that were never lower than 0.99. The principal component analysis (PCA) of the first 3,000 best simulations from each model (i.e. the 3,000 simulation closest to the observed dataset that are generated by each model) actually shows that the point corresponding to the observed data falls in the middle of the results obtained simulating growth, thus confirming that the growth model is also able to generate the observed variation (Suppl. Fig. 2). We then estimated the demographic parameters associated with the growth model. The median value for Tstart has been estimated at 103 generations (Table 4), with a 95% highest probability density (HPD) range of 50–287 generations; effective population size increased from 131 (95% HPD: 100–370) to 72,811 (95% HPD: 52,522–95,334). Considering patrilineal generation times of 30–35 years[27], our results indicate that R1b-DF27 started its expansion ~3,000–3,500 ya, shortly after its TMRCA.

As a reference, we applied the same analysis to the whole of R1b-S116, as well as to other common haplogroups such as G2a, I2, and J2a. Interestingly, all four haplogroups showed clear evidence of an expansion ($p > 0.99$ in all cases), all of them starting at the same time, ~50 generations ago (Table 4), and with similar estimated initial and final populations. Thus, these four haplogroups point to a common population expansion, even though I2 (TMRCA, weighted ρ, 7,800 ya) and J2a (TMRCA, 5,500 ya) are older than R1b-DF27. It is worth noting that the expansion of these haplogroups happened after the TMRCA of R1b-DF27.

R1b-DF27	Mean	Median	Mode	95% HPD-LowB	95% HPD-UppB
Tstart	128	103	75	50	287
Na	162	131	100	100	350
Nc	72685	72812	73624	52523	95334
GSM	0.047	0.046	0.044	0.004	0.089
R1b-S116					
Tstart	50	50	50	50	52
Na	182	134	110	100	370
Nc	99844	99854	99860	99690	100000

GSM	0.239	0.240	0.246	0.094	0.380
G2a					
Tstart	82	61	51	50	175
Na	145	122	100	100	340
Nc	10448	7984	5865	1549	25267
GSM	0.179	0.177	0.179	0.023	0.325
I2					
Tstart	82	67	58	50	136
Na	178	140	100	100	370
Nc	40622	37366	28135	9162	81946
GSM	0.036	0.027	0	0	0.099
J2a					
Tstart	55	52	51	50	61
Na	154	129	100	100	350
Nc	94264	95421	96953	85633	100000
GSM	0.022	0.017	0.000	0.000	0.069

Table 4. ABC results for R1b-DF27 and reference haplogroups. GSM: Generalized Stepwise Mutation; HPD: Highest Probability Density.

Discussion

We have characterized the geographical distribution and phylogenetic structure of haplogroup R1b-DF27 in W. Europe, particularly in Iberia, where it reaches its highest frequencies (40–70%). The age of this haplogroup appears clear: with independent samples (our samples vs. the 1000 genome project dataset) and independent methods (variation in 15 STRs vs. whole Y-chromosome sequences), the age of R1b-DF27 is firmly grounded around 4000–4500 ya, which coincides with the population upheaval in W. Europe at the transition between the Neolithic and the Bronze Age[2, 9]. Before this period, R1b-M269 was rare in the ancient DNA record, and during it the current frequencies were rapidly reached[2, 9, 10]. It is also one of the haplogroups (along with its daughter clades, R1b-U106 and R1b-S116) with a sequence structure that shows signs of a population explosion or burst[1]. STR diversity in our dataset is much more compatible with population growth than with stationarity, as shown by the ABC results, but, contrary to other haplogroups such as the whole of R1b-S116, G2a, I2 or J2a, the start of this growth is closer to the TMRCA of the haplogroup. Although the median time for the start of the expansion is older in R1b-DF27 than in other haplogroups, and could suggest the action of a different demographic process, all HPD intervals broadly overlap, and thus, a common demographic history may have affected the whole of the Y chromosome diversity in Iberia. The HPD intervals encompass a broad timeframe, and could reflect the post-Neolithic population expansions from the Bronze Age to the Roman Empire[28].

While when R1b-DF27 appeared seems clear, where it originated may be more difficult to pinpoint. If we extrapolated directly from haplogroup frequencies, then R1b-DF27 would have originated in the Basque Country; however, for R1b-DF27 and most of its subhaplogroups, internal diversity measures and age estimates are lower in Basques than in any other population. Then, the high frequencies of R1b-DF27 among Basques could be better explained by drift rather than by a local origin (except for the case of M153; see below), which could also have decreased the internal diversity of R1b-DF27 among Basques. An origin of R1b-DF27 outside the Iberian Peninsula could also be contemplated, and could mirror the external origin of R1b-M269, even if it reaches there its highest frequencies. However, the search for an external origin would be limited to France and Great Britain; R1b-DF27 seems to be rare or absent elsewhere: Y-STR data are available only for France, and point to a lower diversity and more recent ages than in Iberia (Table 3). Unlike in Basques, drift in a traditionally closed population seems an unlikely explanation for this pattern, and therefore, it does not seem probable that R1b-DF27 originated in France. Then, a local origin in Iberia seems the most plausible hypothesis. Within Iberia, Aragon shows the highest diversity and age estimates for R1b-DF27, Z195, and the L176.2 branch, although, given the small sample size, any conclusion should be taken cautiously. On the contrary, Z220 and Z278 are estimated to be older in North Central Spain (N Castile, Cantabria and Asturias). Finally, M153 is almost restricted to the Basque Country: it is rarely present at frequencies >1% elsewhere in Spain (although see the cases of Alacant, Andalusia and Madrid, Suppl. Table 1), and it was found at higher frequencies (10–17%) in several Basque regions[25]; a local origin seems plausible, but, given the scarcity of M153 chromosomes outside of the Basque Country, the diversity and age values cannot be compared.

Within its range, R1b-DF27 shows same geographical differentiation: Western Iberia (particularly, Asturias and Portugal), with low frequencies of R1b-Z195 derived chromosomes and relatively high values of R1b-DF27* (xZ195); North Central Spain is characterized by relatively high frequencies of the Z220 branch compared to the L176.2 branch; the latter is more abundant in Eastern Iberia. Taken together, these observations seem to match the East-West patterning that has occurred at least twice in the history of Iberia: i) in pre-Roman times, with Celtic-speaking peoples occupying the center and west of the Iberian Peninsula, while the non-Indoeuropean eponymous Iberians settled the Mediterranean coast and hinterland; and ii) in the Middle Ages, when Christian kingdoms in the North expanded gradually southwards and occupied territories held by Muslim fiefs.

The relevance and possible applications of R1b-DF27

Although R1b-DF27 as a whole has remained relatively obscure in the academic literature, two of the SNPs it contains, namely M167 (SRY2627) and M153 have accrued quite a number of studies. Thus, excluding this paper, M153 has been typed in 42 populations, for a total of 3,117 samples[22, 23, 25, 29, 30]; M167 has been typed in at least 113 populations and 10,379 individuals[4, 18–20, 22–25, 29–31]. It is not obvious then why both markers are absent from Y-phylotree (http://www.phylotree.org/Y/tree/index.htm, ref. 32), which is the current academic Y-chromosome haplogroup reference tree and which contains within DF27 a number of much more obscure SNPs.

Potentially, a SNP with relatively high frequencies in Iberian and Iberian-derived populations and rarer elsewhere could be applied in a forensic genetics setting to infer the biogeographic origin of an unknown contributor to a crime scene[33]. However, neither the specificity nor the sensitivity of such an application would guarantee significant investigative leads in most cases. When compared to the 1000 genomes CEU sample of European-Americans[15], R1b-DF27 is just 4.19 times more frequent in Iberians than in CEU, a ratio that raises to 6.82 for R1b-Z220 (which, though, has a frequency of only 13.9% in Iberians). Probably, other types of evidence of the involvement of a person of interest of Iberian descent would be needed to justify tying R1b-DF27.

R1b-DF27 may also be used to trace migratory events involving Spanish or Portuguese men, particularly outside of Western Europe; a clear example can be seen the Latin American populations (see the Introduction section), where R1b-DF27 seems to correlate with the amount of male-mediated Spanish admixture: it is clearly less frequent in the populations with a stronger Native American component, such as Mexico and Peru. Even within Europe, Y haplogroup frequencies have been used to detect short-range migration events, such as that from Northern France to Flanders[34]. Thus, the traces of the medieval expansion of the Aragon kingdom towards the Mediterranean in the 14th–15th centuries, or the Castilian occupation of Flanders in the 17th century may be traced through the male lineages, R1b-DF27 in particular.

Finally, the Y chromosome in often studied in connection with surnames, since the latter are also often transmitted through the male line[35]. For that, Y-STR haplotypes are analyzed, and, given the Y-STR mutation rates, similarity in Y-STR haplotypes between men sharing the same surname is taken as indicative of a shared genealogical origin[36, 37]. However, diversity in Y-STR haplotypes within the R1b-M269 branch is rather small[11, 12], and the sole use of Y-STRs may result in homoplasy, rather than shred origin, causing Y-STR haplotype convergence. Thus, particularly within Iberia, R1b-DF27 should be used when trying to ascertain the founding events of surnames.

No SNP deeper than R1b-M269 was typed in a survey of Spanish surnames[38], while some SNPs in the R1b-DF27 branch (Z195, Z220, Z278, M153 and M167) were used in a similar study[27].

Although we have contributed to the understanding of the phylogeography of R1b-DF27, which makes up a dominant fraction of Iberian (and Latin American) Y chromosomes, better tools and designs would be needed to solve some of the issues we discussed above. In particular, we genotyped pre-ascertained SNPs, and a global characterization of the whole sequence diversity of this haplogroup would allow more precise statistical analyses to be run. Also, a more comprehensive sampling scheme, including more information from Atlantic Iberia, would be desirable to obtain a more accurate picture of this haplogroup.

Methods

Samples/ethics

The population samples we analyzed comprised a total of 2990 individuals, of which 1072 carried the derived allele at the DF27 SNP. Additionally, 55 individuals with partial information were used to estimate subhaplogroup frequencies (see below). These samples cover the Iberian Peninsula and France, and were originally described in[16, 27, 39] (Table 1, Fig. 2). Also, subhaplogroup frequencies were estimated for the British (GBR) and Tuscan (TSI) samples of the 1000 genomes project[15]. Informed consent for study participation was obtained from all the subjects; Internal Review Board approval for this work was granted by Faculty of Pharmacy UPV/EHU, September 26th 2008; CEISH/119/2012, BNADN Ref. 12/0031; and CEIC-PSMAR ref. 2016/6723/I. This research was conducted under the principles of the Helsinki declaration.

SNP genotyping

All samples were typed for SNPs/indels M269, S116 (P312), DF27, Z195, L176, M167 (SRY2627), Z220 (S356), Z278, and M153 (Fig. 1). DF27, Z195, L176, M167, Z220, Z278, and M153 were typed in samples from Portugal, Andalusia, Galicia, Madrid, and part of the Alacant and Barcelona samples as described in ref. 17. The original genotyping of the French samples (except Brittany)[39] was supplemented with the SNPs in the Open Array panel described in ref. 27. Subsequently, the French samples plus others from Eastern Iberia (see Table 1) were genotyped for DF27 and L176 by Sanger sequencing, since these polymorphisms were not part of the original Open Array panel. Both were amplified using 2.5 µl buffer, 2 µl dNTPs, 1.25 µl each of forward/reverse primers, 1.5 µl MgCl2, 0.2 µl Taq polymerase, 1.5 µl DNA, and 14.8 µl H2O. DF27 was first amplified with a nested PCR to reduce non-specific amplifications. The nested PCR involves two sets of primers used in two successive PCR amplifications, namely outer DF27 forward:

GGGAATTTGATCCTGTCGTTG, outer DF27 reverse: GAACAAAGCCTCCAAGAAATATGAGG, M13F-tagged nested DF27 forward: TGTAAAACGACGGCCAGTTATTTTATTTCTCCTTCACTTATA, nested DF27 Reverse: ATCCAGGAGAACTTCCCCAATC. In the first PCR, 30 cycles were performed at 95 °C (30 sec), 60.5 °C (30 sec), and 72 °C (40 sec); in the second PCR, the annealing temperature was lowered to 59.2 °C. For L176, primers were L176 Forward: CAACAGGCCAGAAGGAACAG and L176 reverse: TTACAGGTGGAATGGGGTGT; the annealing temperature was 58.3 °C, and times and number of cycles were the same as in DF27.

Genotypes for 17 short tandem repeats (STRs) contained in the AmpFlSTR®YFiler® PCR Amplification kit (Applied Biosystems) were available for most populations (see Table 1)[16, 26, 27, 39–41]. The dataset generated during the current study is available from the corresponding author on reasonable request.

Statistics

For most populations, the frequencies of DF27 and its subhaplogroups were estimated by direct counting. However, in some populations, individuals with partial information were present: in some cases, no SNP information was available, but they could be inferred to carry R1b from their STR haplotypes[42, 43]; further subhaplogroup inference is precluded by the high homogeneity of STR haplotypes within R1b-M269[11, 12]. In other cases, individuals were known to be S116 (xZ195, L21, U152) or Z195 (xM167, Z220), but further genotyping for DF27 or L176 failed. The relative proportions of cases with full genotypes over R1b, S116 or Z195 were used to estimate the probabilities of each individual with missing genotypes to belong to each possible subhaplogroup. Using these probabilities as frequencies, the frequency of each subhaplogroup was estimated. Detailed formulas for each subhaplogroup are given in Supplementary note 1. Individuals with missing information were used only to refine the estimation of subhaplogroup frequencies.

Haplogroup frequency maps were drawn with SURFER v. 12 (Golden Software, Golden CO, USA) by krigging. Principal component analysis was performed with IBM SPSS Statistics v. 19. Basic descriptive statistics, as well as AMOVA, were computed with Arlequin 3.5[44]. Haplogroups were dated from STR variation with ρ_w, a weighted version of ρ[45] that leverages on the relatively precise knowledge of the mutation rate of each Y-STR. Thus, it takes into account that mutations at slow STRs take longer to accumulate than mutations at faster STRs. It is defined as

$$\rho_w = \frac{1}{N}\sum_{i=1}^{k} n_i \left(\sum_{j=1}^{S} \left(\left| X_{ji} - X_{jm} \right| \right) \cdot \frac{\bar{\mu}}{\mu_j} \right)$$

where *N* is the number of chromosomes, *k* is the number of different haplotypes, n_i is the absolute frequency of the *ith* haplotype, *S* is the number of different STRs, X_{ji} is the allelic state of the *ith* haplotype at the *jth* STR, X_{jm} is the median allele at the *jth* STR, $\bar{\mu}$ is the average mutation rate and μ_j is the mutation rate of the *jth* STR. The standard deviation of ρ_w is given by

$$sd(\rho_w) = \frac{1}{N}\sqrt{\sum_{i=1}^{k} n_i^2 \left(\sum_{j=1}^{S} \left(|X_{ji} - X_{jm}| \right) \cdot \frac{\bar{\mu}}{\mu_j} \right)}$$

and age, as in ref. 45, is estimated as

$$T = \rho_w \cdot \bar{\mu}$$

where $\bar{\mu}$ is now expressed in years per mutation. ρ_W was computed with an *ad hoc* R script, which is available in github (http://github.com/fcalafell/weighted_rho). Mutation rates were retrieved from the Y-Chromosome STR Haplotype Database (YHRD, www.yhrd.org) on Feb. 1, 2017. DYS385 was omitted from the calculations, and DYS389I was subtracted from DYS389II. Additionally, outlier individuals were detected and removed from the estimate as suggested in ref. 20.

Unweighted ρ was used to estimate the age of DF27 by using the whole Y chromosome sequences of the 88 unrelated individuals derived for this SNP and present in the 1000 genomes project dataset. The mutation rate considered was 0.888×10^{-9} per year[1, 46], which, taking into account the ~10.36 Mb of the Y chromosome amenable to short-read sequencing and SNP detection[1], translates to a rate of 108.7 years/mutation.

Approximate Bayesian Computing (ABC) was used to test alternative demographic models and to estimate their parameters. One million simulations were run with *fastsimcoal2*[47, 48], either with a constant population size (drawn from a lognormal distribution between 100 and 100,000), or with an exponential growth that started Tstart generation ago. In the growth model, the effective population sizes before (Na) and at the end (Nc) of the growth were drawn in the same fashion of the constant model, and conditioned to Na < Nc. Na refers to a time Tstart drawn from a uniform distribution between 50 and 350 generations. STR mutation rates were taken as fixed given the high precision with which they are known, but the value of the geometric parameter for the Generalized Stepwise Mutation model was sampled from a uniform distribution with limits (0; 0.8). To summarize the data, we calculated the mean and the standard deviation over loci of four statistics: the number of different haplotypes (K), the haplotype diversity (H), the allelic range and the Garza- Williamson's index. We calculated posterior probabilities of the models by means of the simple rejection algorithm[49] as well as of the weighted multinomial logistic regression[50],

evaluating different thresholds for both methods to check the stability of the results as in ref. 51. For parameter estimation, we calculated the Euclidian distance between the simulated and observed summary statistics and retained the 5% of the total simulations corresponding to the shortest distances. Posterior probability for each parameter was estimated using a weighted local regression[52], after a logtan transformation[53].

References

1. Poznik, G. D. et al. Punctuated bursts in human male demography inferred from 1,244 worldwide Y-chromosome sequences. Nat. Genet. 48, 593–599 (2016).

2. Haak, W. et al. Massive migration from the steppe was a source for Indo-European languages in Europe. Nature 522, 207–211 (2015).

3. Myres, N. M. et al. A major Y-chromosome haplogroup R1b Holocene era founder effect in Central and Western Europe. Eur. J. Hum. Genet. 19, 95–101 (2011).

4. Rosser, Z. H. et al. Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language. Am. J. Hum. Genet. 67, 1526–1543 (2000).

5. Semino, O. et al. The genetic legacy of Paleolithic Homo sapiens sapiens in extant Europeans: a Y chromosome perpective. 290, 1155–1159 (2000).

6. Balaresque, P. et al. A predominantly neolithic origin for European paternal lineages. PLoS Biol 8, e1000285 (2010).

7. Busby, G. B. J. et al. The peopling of Europe and the cautionary tale of Y chromosome lineage R-M269. Proc. R. Soc. B Biol. Sci. 279, 884–892 (2012).

8. Batini, C. et al. Large-scale recent expansion of European patrilineages shown by population resequencing. Nat. Commun. 6, 7152 (2015).

9. Allentoft, M. E. et al. Population genomics of Bronze Age Eurasia. Nature 522, 167–172 (2015).

10. Fu, Q. et al. The genetic history of Ice Age Europe. Nature 534, 200–5 (2016).

11. Larmuseau, M. H. D. et al. Recent Radiation within Y-chromosomal Haplogroup R-M269 Resulted in High Y-STR Haplotype Resemblance. Ann. Hum. Genet. 78, 92–103 (2014).

12. Solé-Morata, N., Bertranpetit, J., Comas, D. & Calafell, F. Recent radiation of R-M269 and high Y-STR haplotype resemblance confirmed. Ann. Hum. Genet. 78 (2014).

13. Cruciani, F. et al. Strong intra- and inter-continental differentiation revealed by Y chromosome SNPs M269, U106 and U152. Forensic Sci. Int. Genet. 5, e49–52 (2011).

14. Rocca, R. A. et al. Discovery of Western European R1b1a2 Y chromosome variants in 1000 genomes project data: an online community approach. PLoS One 7, e41634 (2012).

15. Auton, A. et al. A global reference for human genetic variation. Nature 526, 68–74 (2015).

16. Valverde, L. et al. New clues to the evolutionary history of the main European paternal lineage M269: dissection of the Y-SNP S116 in Atlantic Europe and Iberia. Eur. J. Hum. Genet. 24, 437–41 (2016).

17. Villaescusa, P. et al. Characterization of the Iberian Y chromosome haplogroup R-DF27 in Northern Spain. Forensic Sci. Int. Genet. 27, 142–148 (2017).

18. Rębała, K. et al. Contemporary paternal genetic landscape of Polish and German populations: from early medieval Slavic expansion to post-World War II resettlements. Eur. J. Hum. Genet. 21, 415–22 (2013).

19. Larmuseau, M. H. D. et al. Increasing phylogenetic resolution still informative for Y chromosomal studies on West-European populations. Forensic Sci. Int. Genet. 9, 179–185 (2014).

20. Boattini, A. et al. Uniparental markers in Italy reveal a sex-biased genetic structure and different historical strata. PLoS One 8, e65441 (2013).

21. Hurles, M. E. et al. Substantial recent male-mediated gene flow between Basque and Catalan populations suggested by analysis of a Y-chromosomal polymorphism. 65, 1437–1448 (1999).

22. Flores, C. et al. Reduced genetic structure of the Iberian peninsula revealed by Y-chromosome analysis: implications for population demography. Eur. J. Hum. Genet. 12, 855–63 (2004).

23. Alonso, S. et al. The place of the Basques in the European Y-chromosome diversity landscape. Eur. J. Hum. Genet. 13, 1293–1302 (2005).

24. Beleza, S. et al. Micro-phylogeographic and demographic history of Portuguese male lineages. Ann. Hum. Genet. 70, 181–94 (2006).

25. Martínez-Cruz, B. et al. Evidence of pre-Roman tribal genetic structure in Basques from uniparentally inherited markers. Mol. Biol. Evol. 29, 2211–22 (2012).

26. Purps, J. et al. A global analysis of Y-chromosomal haplotype diversity for 23 STR loci. Forensic Sci. Int. Genet. 12, 12–23 (2014).

27. Solé-Morata, N., Bertranpetit, J., Comas, D. & Calafell, F. Y-chromosome diversity in Catalan surname samples: insights into surname origin and frequency. Eur. J. Hum. Genet. 23, 1549–57 (2015).

28. McEvedy, C. & Jones, R. Atlas of world population history. (Penguin, 1978).

29. Adams, S. M. et al. The genetic legacy of religious diversity and intolerance: paternal lineages of Christians, Jews, and Muslims in the Iberian Peninsula. Am. J. Hum. Genet. 83, 725–736 (2008).

30. Bekada, A. et al. Introducing the Algerian mitochondrial DNA and Y-chromosome profiles into the North African landscape. PLoS One 8, e56775 (2013).

31. Varzari, A. et al. Paleo-Balkan and Slavic contributions to the genetic pool of Moldavians: insights from the Y chromosome. PLoS One 8, e53731 (2013).

32. van Oven, M., Van Geystelen, A., Kayser, M., Decorte, R. & Larmuseau, M. H. D. Seeing the wood for the trees: a minimal reference phylogeny for the human Y chromosome. Hum. Mutat. 35, 187–91 (2014).

33. Kayser, M. Forensic use of Y-chromosome DNA: a general overview. Hum. Genet. 136, 621–635 (2017).

34. Larmuseau, M. H. D. et al. In the name of the migrant father—Analysis of surname origins identifies genetic admixture events undetectable from genealogical records. Heredity (Edinb). 109, 90–95 (2012).

35. Calafell, F. & Larmuseau, M. H. D. The Y chromosome as the most popular marker in genetic genealogy benefits interdisciplinary research. Hum. Genet. 136, 559–573 (2017).

36. King, T. E. & Jobling, M. A. Founders, drift, and infidelity: the relationship between Y chromosome diversity and patrilineal surnames. Mol. Biol. Evol. 26, 1093–1102 (2009).

37. McEvoy, B. & Bradley, D. G. Y-chromosomes and the extent of patrilineal ancestry in Irish surnames. Hum. Genet. 119, 212–219 (2006).

38. Martinez-Cadenas, C. et al. The relationship between surname frequency and Y chromosome variation in Spain. Eur. J. Hum. Genet. 24, 120–128 (2016).

39. Ramos-Luis, E. et al. Y-chromosomal DNA analysis in French male lineages. Forensic Sci. Int. Genet. 9, 162–168 (2014).

40. Núñez, C. et al. Reconstructing the population history of Nicaragua by means of mtDNA, Y-chromosome STRs, and autosomal STR markers. Am. J. Phys. Anthropol. 143, 591–600 (2010).

41. Nuñez, C. et al. Highly discriminatory capacity of the PowerPlex® Y23 System for the study of isolated populations. Forensic Sci. Int. Genet. 17, 104–107 (2015).

42. Athey, T. W. Haplogroup prediction from Y-STR values using an allele-frequency approach. J. Genet. Geneal. 1, 1–7 (2005).

43. Athey, T. W. Haplogroup prediction from Y-STR values using a Bayesian allele frequency approach. J. Genet. Geneal. 2, 34–39 (2006).

44. Excoffier, L. & Lischer, H. E. L. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol. Ecol. Resour. 10, 564–567 (2010).

45. Saillard, J., Forster, P., Lynnerup, N., Bandelt, H. J. & Nørby, S. mtDNA variation among Greenland Eskimos: the edge of the Beringian expansion. Am. J. Hum. Genet. 67, 718–726 (2000).

46. Helgason, A. et al. The Y-chromosome point mutation rate in humans. Nat. Genet. 47, 453–7 (2015).

47. Excoffier, L. & Foll, M. fastsimcoal: a continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. Bioinformatics 27, 1332–1334 (2011).

48. Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C. & Foll, M. Robust demographic inference from genomic and SNP data. PLoS Genet. 9, e1003905 (2013).

49. Pritchard, J. K., Seielstad, M. T., Perez-Lezaun, A. & Feldman, M. W. Population growth of human Y chromosomes: a study of Y chromosome microsatellites. Mol. Biol. Evol. 16, 1791–8 (1999).

Acknowledgements

We want to thank the thousands of volunteers who made this work possible. Cristina de Vasconcelos superbly organized the logistics of collecting part of the samples used in this paper. Funding was provided by the Agencia Estatal de Investigación and Fondo Europeo de Desarollo Regional (FEDER) (grants CGL2013-44351-P, CGL2016-75389-P), by Agència de Gestió d'Ajuts Universitaris i de la Recerca (Generalitat de Catalunya) grant 2014 SGR 866, and by the Basque Government (IT-424-07). FT was supported by the ERC Advanced Grant Agreement No 295733, 'LanGeLin' project.

Author contributions

N.S.M., C.F., K.R., S.J.M., B.M.J., M.F.P., M.T.Z., A.C., M.M.P., and F.C. contributed samples. L.V., M.J.I., F.T., and S.G. participated in the design of the study and performed some of the analyses. N.S.M., P.V., C.G.F., N.F.P., and F.C. generated the dataset, performed most of the analyses, and contributed to the interpretation of the result. N.S.M., P.V., M.M.P., and F.C. designed the study and wrote the manuscript. All authors reviewed the manuscript.

Additional Information

Supplementary information accompanies this paper at doi:10.1038/s41598-017-07710-x

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Electronic supplementary material

Supplementary Dataset

Supplementary Table 1. Subhaplogroup frequencies.

Population	N	other	R1b-M269	R1b-P312	R1b-DF27	R1b-DF27*	R1b-Z195	R1b-Z195*	R1b-L176.2	R1b-L176.2*	R1b-M167	R1b-Z220	R1b-Z220*	R1b-Z278	R1b-Z278*	R1b-M153
Alacant	142	0,3451	0,6549	0,5845	0,4225	0,1620	0,2606	0,0000	0,1127	0,0141	0,0986	0,1479	0,0704	0,0775	0,0282	0,0493
Alsace	80	0,4125	0,5875	0,3875	0,0750	0,0000	0,0750	0,0500	0,0250	0,0125	0,0125	0,0000	0,0000	0,0000	0,0000	0,0000
Andalucía	100	0,3700	0,6300	0,6000	0,4700	0,2800	0,1900	0,0400	0,0200	0,0000	0,0200	0,1300	0,0500	0,0800	0,0400	0,0400
Aragón	92	0,3370	0,6630	0,6087	0,3696	0,1522	0,2174	0,0217	0,1196	0,0761	0,0435	0,0761	0,0435	0,0326	0,0217	0,0109
Asturias	63	0,4286	0,5714	0,5714	0,4286	0,3016	0,1270	0,0794	0,0000	0,0000	0,0000	0,0476	0,0159	0,0317	0,0317	0,0000
Auvergne	89	0,4719	0,5281	0,4944	0,0562	0,0112	0,0449	0,0337	0,0112	0,0000	0,0112	0,0000	0,0000	0,0000	0,0000	0,0000
Barcelona	571	0,3047	0,6953	0,6162	0,3979	0,0987	0,2992	0,0202	0,1364	0,0444	0,0920	0,1426	0,0957	0,0468	0,0360	0,0108
Brittany	145	0,1310	0,8690	0,8345	0,1931	0,0966	0,0966	0,0069	0,0483	0,0276	0,0207	0,0414	0,0414	0,0000	0,0000	0,0000
Cantabria	96	0,2813	0,7188	0,6250	0,4479	0,2292	0,2188	0,0208	0,0313	0,0104	0,0208	0,1667	0,0417	0,1250	0,1250	0,0000
Castelló	49	0,3265	0,6735	0,6283	0,4717	0,0660	0,4058	0,0292	0,1839	0,0558	0,1282	0,1926	0,0674	0,1252	0,1042	0,0210
Galicia	70	0,3857	0,6143	0,5571	0,4000	0,2429	0,1571	0,0714	0,0000	0,0000	0,0000	0,0857	0,0429	0,0429	0,0429	0,0000
GBR (1000 genomes)	46	0,2609	0,7391	0,5217	0,1522	0,0652	0,0870	0,0217	0,0652	0,0217	0,0435	0,0000	0,0000	0,0000	0,0000	0,0000
Girona	131	0,3969	0,6031	0,5022	0,2874	0,0593	0,2281	0,0081	0,1098	0,0161	0,0937	0,1102	0,0939	0,0163	0,0084	0,0079
Île-de-France	91	0,4396	0,5604	0,4693	0,1026	0,0000	0,1025	0,0900	0,0124	0,0000	0,0121	0,0000	0,0000	0,0000	0,0000	0,0000
Ireland	146	0,1849	0,8151	0,7466	0,0068	0,0000	0,0068	0,0000	0,0068	0,0068	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Lleida	104	0,2788	0,7212	0,6518	0,4146	0,1080	0,3066	0,0102	0,1287	0,0395	0,0892	0,1677	0,1279	0,0398	0,0299	0,0099
Madrid	99	0,3131	0,6869	0,6162	0,4949	0,2121	0,2828	0,0303	0,1010	0,0707	0,0303	0,1515	0,0404	0,1111	0,0808	0,0303
Mallorca	48	0,3125	0,6875	0,6875	0,4587	0,1671	0,2917	0,0000	0,1250	0,0208	0,1042	0,1667	0,1667	0,0000	0,0000	0,0000
Midi-Pyrénées	67	0,4030	0,5970	0,5351	0,1070	0,0000	0,1069	0,0313	0,0756	0,0000	0,0754	0,0000	0,0000	0,0000	0,0000	0,0000
Native Basques	229	0,0786	0,9214	0,9214	0,7380	0,3450	0,3930	0,0393	0,0699	0,0393	0,0306	0,2838	0,0786	0,2052	0,1397	0,0655
Nord-Pas-de-Calais	68	0,3824	0,6176	0,4935	0,1251	0,0000	0,1249	0,0484	0,0616	0,0300	0,0316	0,0149	0,0149	0,0000	0,0000	0,0000
Portugal	109	0,3761	0,6239	0,5046	0,4037	0,3211	0,0826	0,0275	0,0459	0,0275	0,0183	0,0092	0,0000	0,0092	0,0092	0,0000
Provence–Alpes-Côte d'Azur	45	0,4444	0,5556	0,5235	0,1223	0,0000	0,1220	0,0731	0,0486	0,0000	0,0478	0,0000	0,0000	0,0000	0,0000	0,0000
Pyrenees	46	0,3043	0,6957	0,6957	0,4241	0,0763	0,3478	0,0217	0,1522	0,0435	0,1087	0,1739	0,1087	0,0652	0,0652	0,0000
Resident Basques	111	0,3784	0,6216	0,5766	0,4775	0,2432	0,2342	0,0450	0,0721	0,0090	0,0631	0,1171	0,0270	0,0901	0,0811	0,0090
Tarragona	120	0,3583	0,6417	0,5750	0,3459	0,0292	0,3167	0,0364	0,1052	0,0136	0,0917	0,1750	0,0917	0,0833	0,0667	0,0167
TSI (1000 genomes)	53	0,5472	0,4528	0,3774	0,0755	0,0189	0,0566	0,0377	0,0000	0,0000	0,0000	0,0189	0,0000	0,0000	0,0000	0,0000
València	79	0,2911	0,7089	0,6669	0,4076	0,1251	0,2826	0,0438	0,0785	0,0479	0,0306	0,1602	0,1070	0,0532	0,0400	0,0133

Supplementary Table 2. STR variances by subhaplogroup and region.

	DF27			Z195			L176.2			M167			Z220			Z278			M153		
	N	Var	sd	N	Var	sd	N	Var	sd	N	Var	sd	N	Var	sd	N	Var	sd	N	Var	sd
All	758	0,330	0,215	510	0,326	0,198	189	0,287	0,213	137	0,245	0,190	267	0,293	0,211	130	0,225	0,141	34	0,146	0,115
Aragón	29	0,372	0,218	18	0,314	0,140	10	0,304	0,225	4	0,239	0,297	7	0,200	0,161	3	0,222	0,272	1		
Basques	154	0,282	0,174	83	0,263	0,179	14	0,300	0,315	7	0,254	0,421	62	0,216	0,209	46	0,177	0,132	15	0,107	0,106
Catalonia	311	0,346	0,218	246	0,343	0,208	108	0,285	0,209	81	0,241	0,172	123	0,319	0,245	41	0,231	0,182	9	0,122	0,151
France	35	0,299	0,207	34	0,302	0,212	15	0,363	0,345	12	0,336	0,330	1								
Mallorca	21	0,341	0,280	14	0,354	0,321	7	0,292	0,325	6	0,311	0,360	7	0,295	0,408						
North Central Spain	105	0,319	0,249	44	0,306	0,262	8	0,248	0,271	6	0,220	0,254	27	0,312	0,325	20	0,279	0,245	1		
València	103	0,349	0,280	71	0,324	0,252	27	0,246	0,233	21	0,192	0,192	40	0,318	0,246	20	0,271	0,180	8	0,188	0,172

Supplementary Note

Let *a* be the absolute frequency of haplogroup M269 (xS116) in a sample of *n* Y chromosomes; similarly, let *b*: S116 (xDF27), *c*: DF27 (xZ195), *d*: Z195 (xL176.2, xZ220), *e*: L176.2 (xM167), *f*: M167, *g*: Z220 (xZ278), *h*: Z278 (xM153), and *i*: M153. Let $s=a+b+c+...+i$. We have three types of samples with partial information: R1b-M269 without further subtyping (let its frequency be *j*), S116 (xU152, xM529, xZ195), but not typed for DF27 (call it *k*), and Z195 (xZ220), not typed for L176.2 (*l*). *j* individuals may belong to any of the *a*,...,*i* subhaplogroups with probability *a/s*,...,*i/s*; *k* can be DF27 (xZ195) with probability $c/(b+c)$, and Z195 (xZ220, xM167) can be either Z195 (xL176.2, xZ220) with probability $d/(d+e)$ or L176.2 (xM167) with probability $e/(d+e)$. Combining these probabilities and turning them into estimated relative frequencies (which we denote with a circumflex over each letter), we have

$$\hat{c} = \frac{c\left(1 + \frac{j}{s} + \frac{k}{b+c}\right)}{n}$$

$$\hat{d} = \frac{d\left(1 + \frac{j}{s} + \frac{l}{d+e}\right)}{n}$$

$$\hat{e} = \frac{e\left(1 + \frac{j}{s} + \frac{l}{d+e}\right)}{n}$$

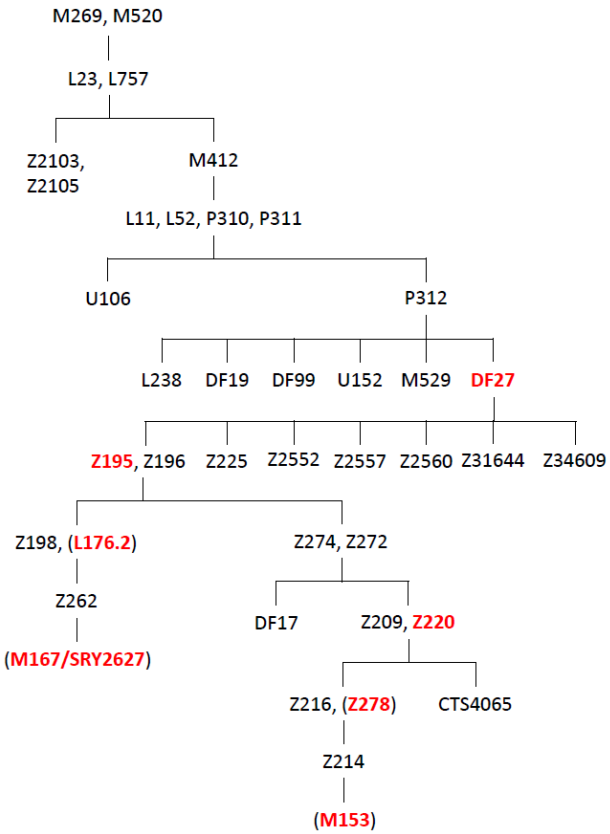
$$\hat{f} = \frac{f\left(1 + \frac{j}{s}\right)}{n}$$

$$\hat{g} = \frac{g\left(1 + \frac{j}{s}\right)}{n}$$

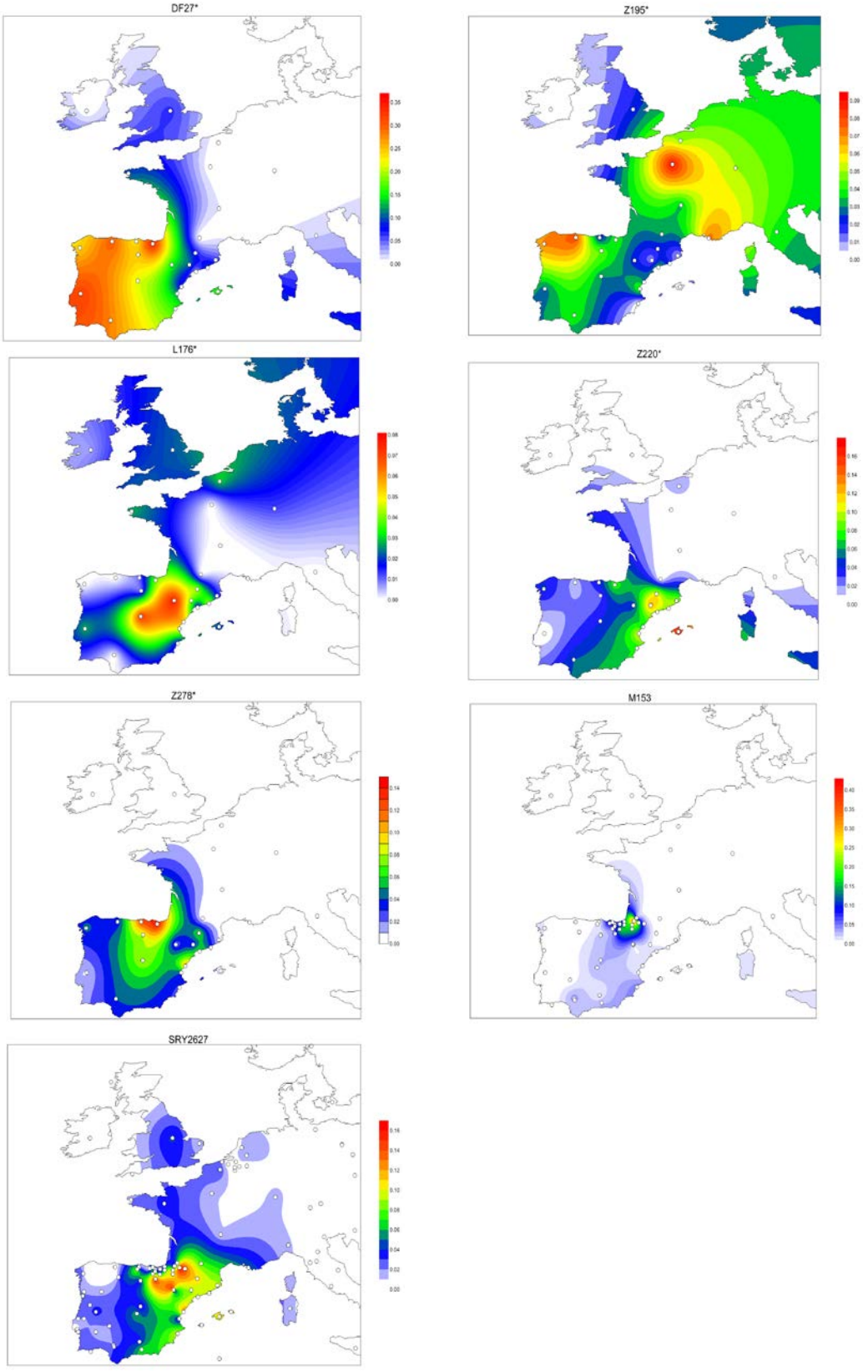
$$\hat{h} = \frac{h\left(1 + \frac{j}{s}\right)}{n}$$

$$\hat{i} = \frac{i\left(1 + \frac{j}{s}\right)}{n}$$

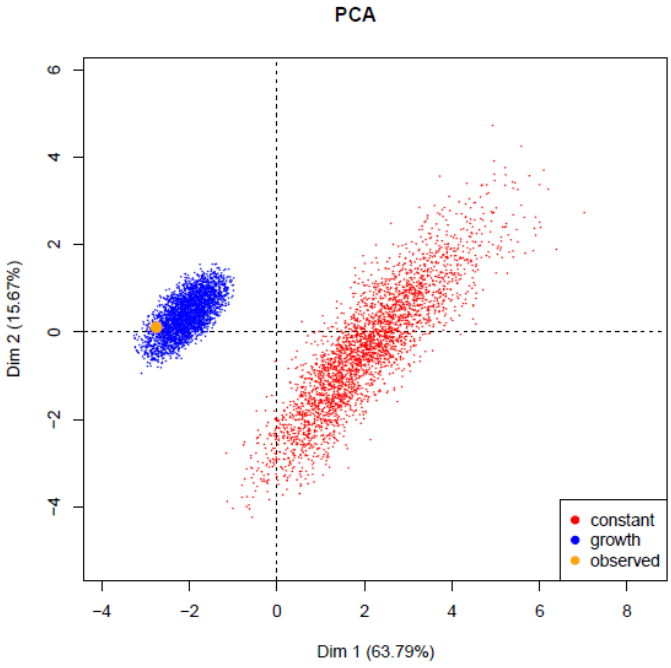
Supplementary Figures



Supplementary Figure 1. R1b-DF27 in the context of the Y-SNP tree compiled in ref. [32] and available from http://www.phylotree.org/Y/tree/index.htm. In red, SNPs typed in this work. In parentheses, SNPs absent from phylotree-Y. Comas separate different SNPs falling the same phylogenetic branch, while slashes indicate alternate names for the same SNP.



Supplementary Figure 2. Additional frequency contour maps of paragroups, and of SRY2627 and of M153 with additional data from the literature. Maps were drawn with SURFER v. 12 (Golden Software, Golden CO, USA).



Supplementary Figure 3. Principal component analysis of summary statistics in stationary and growth ABC simulations; the observed value falls clearly within the cloud of growth simulations.

4.4 Study Number 4

'Effective resolution of the Y chromosome sublineages of the Iberian haplogroup R1b-DF27 with forensic purposes'

The Study Number 4 of the present work corresponds to the attainment of the objective 3: *To design and optimize a new minisequencing method that allows the simultaneous analysis of 15 Y-SNPs for the fine subtyping of the Iberian paternal lineage R1b-DF27, with applicability in both forensic and population analysis.*

In this study a new 15 Y-SNP multiplex was designed and optimized for the fine-resolution subtyping of the haplogroup R1b-DF27 in a single minisequencing reaction. DF27 displays high frequencies in Iberia and Iberian influenced populations, and some of its subhaplogroups show moderate geographical differentiation, which is of interest in forensic genetics in order to link a sample with the bio-geographical origin.

The 15-plex minisequencing panel includes 15 Y-SNPs (U106, P312, U152, M529, L238, DF19, DF27, Z196, L617, L881, DF17, Z220, M153, M167, and S68) strategically chosen based on their ability to resolve the major branches of R1b-DF27, as well as other common Southwest European lineages above DF27. Additionally, we used site-directed mutagenesis with the purpose of producing the derived variants of L881 and DF19, the least common lineages included in the panel. The reproducibility of the assay was assessed by analyzing DNA samples and negative controls several times, by different researchers and using different thermal cyclers.

The obtained results reveal that the 15-plex minisequencing panel is a robust method for subtyping DF27 lineage in a single multiplex reaction. The obtained site-directed mutagenesis products are compatible with minisequencing, making this technique suitable to ascertain the genotyping of rare variants when samples harboring these variants are not available. Furthermore, the short length of the amplicons makes this panel suitable to use with degraded DNA, critical in forensic samples. Finally, the resolution accomplished with this tool enables to improve male lineage discrimination in Iberia, Southwest Europe, and other large areas of the world, as well as making further detailed biogeographical and evolutionary inferences. The geographical differentiation of the sublineages Z220 and M167, included in the panel, could allow to link a vestige with a more specific location in the Iberian Peninsula or with Iberian ancestry.

To conclude, the developed panel is an effective and reproducible method for subtyping DF27 lineage from a minimal quantity of DNA, suitable for the inference of bio-geographical origin and of easy implementation in most forensic and population genetics laboratories.

This study has resulted in an international publication in the journal *International Journal of Legal Medicine* under the heading '*Effective resolution of the Y chromosome sublineages of the Iberian haplogroup R1b-DF27 with forensic purposes*' in September 2018. Q1, IP:2.382. The publication is shown below.

International Journal of Legal Medicine
https://doi.org/10.1007/s00414-018-1936-z

ORIGINAL ARTICLE

Article

Effective resolution of the Y chromosome sublineages of the Iberian haplogroup R1b-DF27 with forensic purposes

Patricia Villaescusa[1], Leire Palencia-Madrid[1], Magdalena Antònia Campaner[1], Jaione Jauregui-Rada[1], Miguel Guerra-Rodríguez[1], Ana María Rocandio[1], Marian M. de Pancorbo[1]

[1]BIOMICs Research Group, Lascaray Research Center, University of the Basque Country UPV/EHU, Avda. Miguel de Unamuno, 3, 01006 Vitoria-Gasteiz, Spain.

Patricia Villaescusa and Leire Palencia-Madrid contributed equally to this work.

Corresponding author: Marian M. de Pancorbo.

Received 6 July 2018, Accepted 7 September 2018, Published online 18 September 2018.

Abstract

Single-nucleotide polymorphisms (SNPs) found within the non-recombining region of the Y chromosome (NRY) represent a powerful tool in forensic genetics for inferring the paternal ancestry of a vestige and complement the determination of biogeographical origin in combination with other markers like AIMs. In the present study, we introduce a panel of 15 Y-SNPs for a fine-resolution subtyping of the haplogroup R1b-DF27, in a single minisequencing reaction. This is the first minisequencing panel that allows a fine subtyping of R1b-DF27, which displays high frequencies in Iberian and Iberian-influenced populations. This panel includes subhaplogroups of DF27 that display moderate geographical differentiation, of interest to link a sample with a specific location of the Iberian Peninsula or with Iberian ancestry. Conversely, part of the intricacy of a new minisequencing panel is to have all the included variants available to test the effectiveness of the analysis method. We have overcome the absence of the least common variants through site-directed mutagenesis. Overall, the results show that our panel is a robust and effective method for subtyping R1b-DF27 lineages from a minimal amount of DNA, and its high resolution enables to improve male lineage discrimination in Iberian and Southwest European descent individuals. The small length of the amplicons and its reproducibility makes this assay suitable for forensic and population genetics purposes.

Keywords

Y chromosome; Y-SNPs; SNaPshot; Site-directed mutagenesis; R1b-DF27.

Introduction

The usefulness of Y chromosome polymorphisms is well established in many different scientific areas. The nonrecombining region of the Y chromosome (NRY) allows to reconstruct patrilineal lineages as it is transmitted unchanged from fathers to their sons, except for occasional mutations that accumulate over time and remain reflected in the evolutionary record [1]. Y chromosome SNPs (Y-SNPs) display low mutation rates and possess a hierarchical structure [2], which makes them informative for evolutionary as well as for forensic studies. Since their geographical distribution is nonrandom [3], they can be employed for inferring the paternal biogeographic ancestry of an unknown contributor to a crime scene [4], an analysis of particular interest in cases where other markers (like Y-STRs) have failed or when the DNA is degraded.

In West Europe, the most common haplogroup is R1b-M269, with frequencies ranging from 40 to 80% [5]. The most important sub-branches of R1b-M269 are U106, more frequent in NC Europe [5, 6], and P312, which is more common in SW Europe [5]. P312, likewise, is divided in the following sublineages: U152, in North Italy and Switzerland [6]; M529 in the British Isles [7]; L238, in Scandinavia [8]; DF27, in the Iberian Peninsula [8, 9]; and DF19, of unknown distribution. DF27, firstly discovered by citizen scientists [8], remained relatively unnoticed in the academic bibliography until not long ago. However, during the last years, the scientific community has grown largely interested in this paternal lineage [10, 11], revealing its near-specificity in the Iberian Peninsula and a potential application in forensic genetics for the determination of biogeographical origin. The recent study of DF27 published by Villaescusa et al. and Solé-Morata et al. [10, 11] has described the different distributions of the sublineages of this paternal lineage in the Iberian Peninsula. Given the geographical differentiation of some of these subhaplogroups (i.e., Z220, in North-Central Spain and M167, in East Iberia), the subtyping of DF27 (L617, L881, Z196, DF17, Z220, M153, M167, and S68; Supplementary Fig. S1) could be considered a powerful tool in forensics for inferring paternal biogeographical ancestry in the Iberian Peninsula.

Numerous technologies are available for SNP genotyping [2, 12], but the most commonly applied methodology to forensic and population studies is the minisequencing or single-base extension (SBE) genotyping due to its sensitivity and multiplexing capability. Many Y-SNP minisequencing panels or assays are available, which include major haplogroups or some European lineages [12],

but most of them require more than one multiplex reaction and do not achieve high phylogenetic resolution.

In order to examine the behavior of the SBE primers on the minisequencing panel, it is significant to test its ability to detect all the alleles of the SNPs included in the design. Even though finding samples from the most common allelic variants is easy, obtaining the least frequent ones can be challenging. In this particular issue, site-directed mutagenesis may be a helpful tool to overcome this inconvenience [13], which we used for subtyping DF19 and L881 subhaplogroups.

In the present study, we design and optimize a minisequencing method for a fine subtyping of the Iberian near-specific paternal lineage R1b-DF27 to the highest phylogenetic resolution to date in a single multiplex reaction. The selected 15 Y-SNPs (U106, P312, U152, M529, L238, DF19, DF27, Z196, L617, L881, DF17, Z220, M153, M167, and S68) were strategically chosen based on their ability to resolve the major branches of R1b-DF27 and provide a good approximation of the biogeographical origin. Additionally, we also used site-directed mutagenesis to produce the least common variants.

Materials and methods

For the development and optimization of the 15-plex minisequencing panel, DNA samples from male individuals with European background were used. Human DNAs were extracted from saliva or peripheral blood samples from healthy male donors who gave their informed consent. Ethical approvals were obtained for this study from the Faculty of Pharmacy UPV/EHU (September 26, 2008, CEISH/119/2012 UPV/EHU), and Spanish DNA National Bank (Ref. 12/0031).

The Y-SNPs selected for the 15-plex minisequencing assay correspond to the diagnostic positions that determine the main subhaplogroups of the paternal lineage R1b-DF27 and some branches above DF27 (Supplementary Fig. S1). The positions were chosen from the updated version of the minimal reference phylogeny for the human Y chromosome PhyloTree Y (9 March 2016) [14] and the more detailed tree maintained by the International Society of Genetic Genealogy (v 12.53; 28 February 2017) [15].

The primers used for amplification (Supplementary Table S1) were designed with Perl Primer v1.1.21 [16]. The specificity of the primers and their non-homology with the X chromosome and other genome regions were confirmed with Primer-BLAST. Potential unfavorable interactions between primers were checked with the web-based version of AutoDimer [17].

Minisequencing primers (Supplementary Table S1) were designed manually. To assure the separation of the extension primers during capillary electrophoresis, their lengths were adjusted

by adding tails of neutral sequence on the 5'-end [18]. Amplification fragments differed in 5 bp in order to allow a clear electrophoretic separation. Unfavorable interactions between minisequencing primers and their specificity to the Y chromosome were checked as described above. Final optimal concentrations for each primer mix were readjusted in line with the different electropherogram intensities.

DF19 and L881 site-directed mutagenesis primers (Supplementary Table S1) were designed manually, by inserting in the primer sequence the necessary nucleotide to produce the derived variant [13]. Each mutagenesis primer was paired with the other end of the amplification primer of its respective Y-SNP, and the suitability of both pairs of primers was checked as described above [17].

PCR, minisequencing, and site-directed mutagenesis primers were synthesized by Integrated DNA Technologies (IDT) and Eurofins. Amplification and mutagenesis primers were purified by standard desalting and minisequencing primers by HPLC. More details on primers are shown in Supplementary Table S1.

PCR multiplex amplification was carried out as follows: 5 µL reaction mix (2×) (Qiagen Multiplex PCR Kit, Qiagen), 1 µL of 10× primer mix, 3 µL of sterile mQ water (Millipore Corporation), and 1 ng of DNA (final volume 10 µL). Thermal cycling was performed in a C1000™ Thermal Cycler (Bio-Rad) in the following conditions: 95 °C for 15 min; 3 cycles at 95 °C for 30 s, 63 °C for 45 s, and 72 °C for 30 s; 15 cycles at 95 °C for 30 s, 63 °C for 45 s (with decrements of 0.2 °C per cycle) and 72 °C for 30 s, 20 cycles at 95 °C for 30 s, 60 °C for 45 s, and 72 °C for 30 s; and a final extension of 7 min at 72 °C. Site-directed mutagenesis was used to produce the derived variants for DF19 and L881. Each mutagenesis reaction was carried out in the following conditions: 8.9 µL of mQ water (Millipore Corporation), 0.6 µL dNTPs (10 mM) (Bioline), 0.6 µL MgCl2 (50 mM) (Bioline), 1.5 µL buffer 10× (Bioline), 0.3 µL bovine serum albumin (10×) (Roche), 0.45 µL of each primer at 10 µM, 0.2 µL Taq polymerase (5 U/µl) (BIOTAQ™ DNA polymerase), and 2 ng of DNA. Amplification success of the mutagenesis derived variants was assessed as described.

PCR products were migrated in 1.5% agarose gels with GelRed (Biotinum) at 100 V for 30 min and visualized with UV light in an UVIdoc gel documentation system (Uvitec Cambridge). Next, PCR products were purified using 0.28 U of exonuclease I (Exo) (Takara) and 0.72 U of shrimp alkaline phosphatase (SAP) (Takara) to 2 µL of PCR product and incubated for 45 min at 37 °C followed by 15 min at 80 °C.

The multiplex minisequencing reaction contained the following: 2 µL of SNaPshot™ Multiplex Kit reaction mix (Applied Biosystems), 0.7 µL of 10× minisequencing primer mix, 3.3 µL of mQ water (Millipore Corporation), and 1 µL purified multiplex PCR product, in a total volume of 7 µL. Thermocycling conditions in a C1000™ Thermal Cycler (Bio-Rad) were the following: 25 cycles at 96 °C for 10s; 50 °C for 5 s; and 60 °C for 30s. Minisequencing products were purified adding 0.75 U of SAP (Takara) to 2 µL of product and incubated for 60 min at 37 °C followed by 15 min at 80 °C.

Finally, a mixture of 1 µL of cleaned minisequencing product, 9.75 µL Hi-DI formamide (Applied Biosystems), and 0.25 µL of Gene-Scan 120LIZ (Applied Biosystems) was prepared and denatured at 96 °C for 6 min. The samples were analyzed using ABI PRISM® 3130 Genetic Analyzer (Applied Biosystems) with a capillary of 36 cm. For optimization, a polymer POP-7® was used. In addition, the final design was also tested with POP-4® polymer. Results were analyzed using GeneMapper® Software v4.0 (Applied Biosystems).

To assess the reproducibility of the 15-plex minisequencing assay, DNA samples and negative controls were analyzed several times by different researchers, on different days and in different thermal cyclers. Afterwards, the mobility of the peaks for the different allelic variants in the electropherograms was compared.

Results

The 15-plex minisequencing panel developed herein includes 15 Y-SNPs (U106, P312, U152, M529, L238, DF19, DF27, Z196, L617, L881, DF17, Z220, M153, M167, and S68) that allow to simultaneously genotype the diagnostic positions of the paternal lineage DF27 and its subhaplogroups, along with other common Southwest European R1b branches above DF27 (Supplementary Fig. S1). More details on frequencies for each Y-SNP in Europe are included in Supplementary Table S2 [5, 8, 10, 11, 19, 20].

The design of the minisequencing assay was optimized to analyze jointly these 15 Y-SNPs in a unique multiplex reaction with up to 1 ng of template DNA (Fig. 1). Moreover, the size of the fragments amplified in the first PCR is short, between 62 and 230 bp, which allows applying this assay to forensic samples. Primer concentrations for both PCRs, the first multiplex amplification and the subsequent minisequencing reaction, were adjusted based on the lowest fluorescent allele signal, in order to obtain balanced intensities for every Y-SNP in the electropherogram. While optimizing this panel, some spurious peaks and high background noise were observed mainly in the green and blue channels of the electropherogram (Supplementary Fig. S2). In order

to solve it, we adjusted the amount of primer mix in the amplification and minisequencing PCRs. Information on how to interpret the electropherograms is included in Supplementary Fig. S3 and Supplementary Table S3.

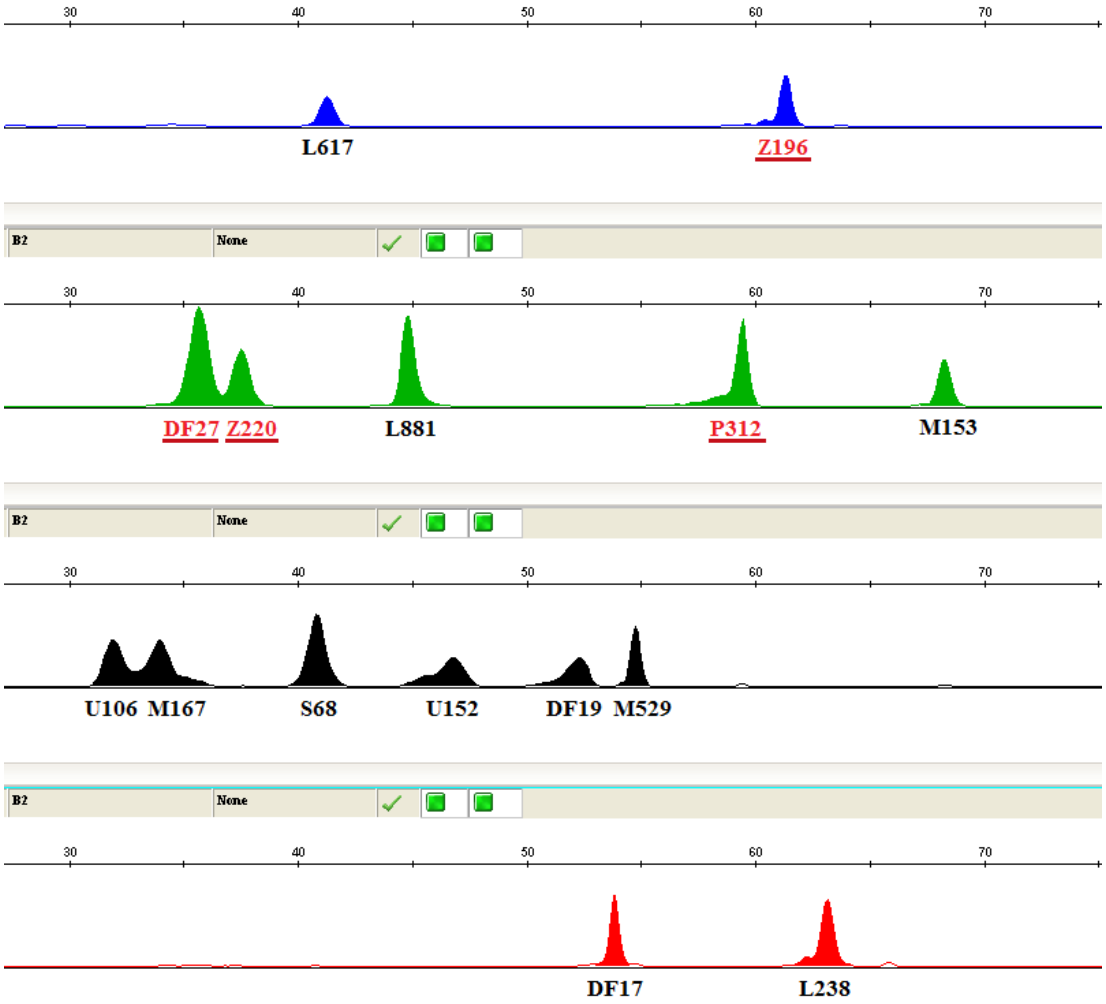


Fig. 1 Electropherogram obtained with the 15-plex minisequencing panel from a male sample belonging to Z220 haplogroup, characterized by the Y-SNPs P312, DF27, Z196, and Z220. Derived allelic variants appear underlined

The effectiveness of this 15-plex for inferring patrilineal biogeographical origin was assessed by analyzing at least two samples displaying haplogroups detectable by our design, except for L238 lineage, of which only one sample was analyzed. For that purpose, we selected individuals from different Southwestern European populations previously analyzed that displayed the haplogroups included in the minisequencing assay (more details about the origin and geographical region of these samples are included in Supplementary Table S4). We analyzed individuals from the branches U106 and P312 and verified that those variants were detected and that no positive results were obtained for the remaining Y-SNPs included in the panel. Similarly, within DF27, we followed the same procedure and observed the same positive results we got by previous

genotyping technologies like high-resolution melting (HRM) and sanger sequencing. Additionally, samples not displaying any of the derived variants included in the design were analyzed. Thus, we ascertained that the 15-plex is able to correctly assign the haplogroup of all the previously genotyped samples [8, 10, 11].

Forasmuch as we were unable to find individuals of the rare lineages DF19 and L881, we generated their derived variants by site-directed mutagenesis in order to verify whether this panel is able to detect them. The resulting electropherograms showed that both mutagenesis products carried the mutated variants (Fig. 2). Thus, we confirmed that the assay would be able to analyze correctly samples belonging to haplogroups DF19 or L881.

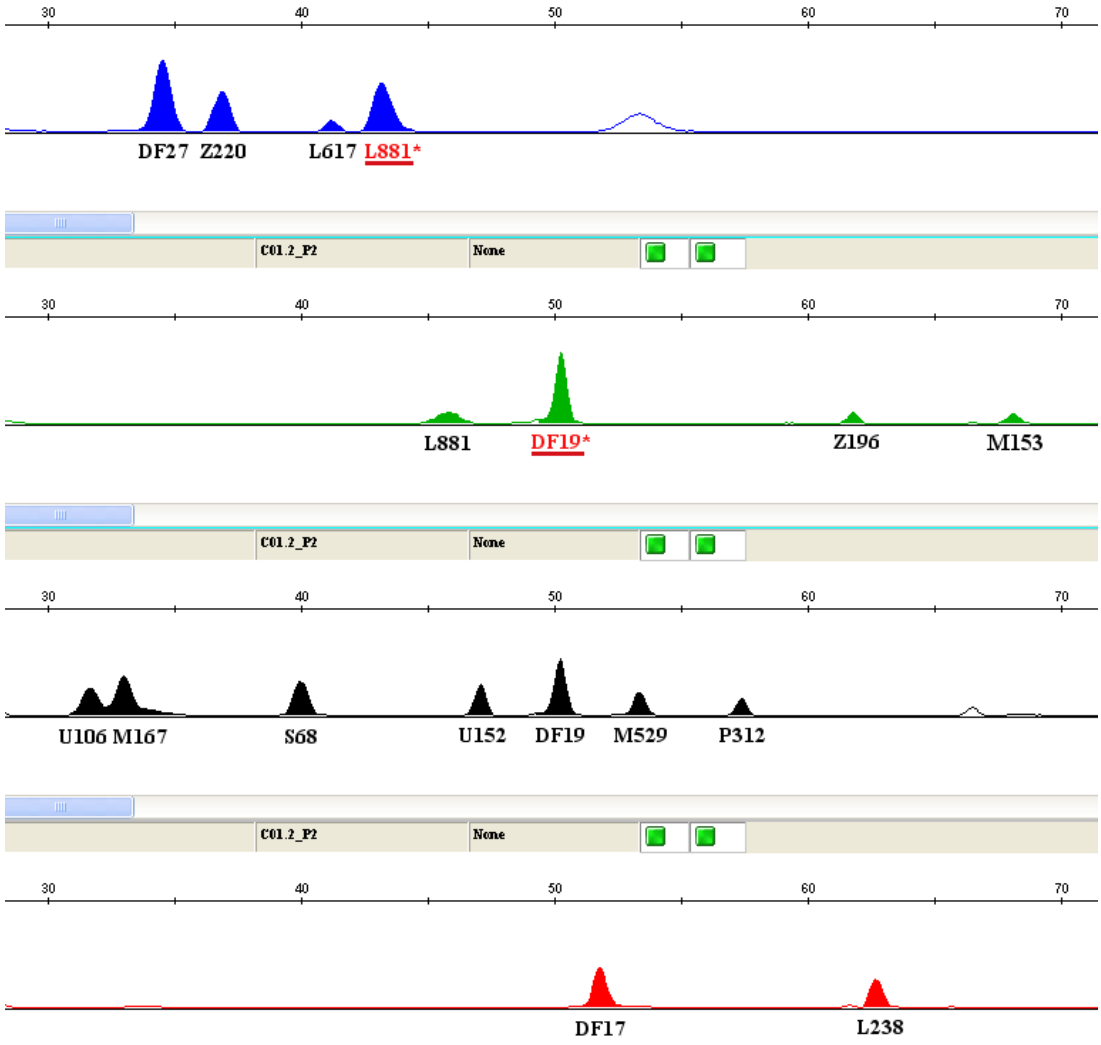


Fig. 2 Electropherogram that shows the position of the mutated variants of L881 and DF19 obtained by site-directed mutagenesis. We obtained this electropherogram by mixing the mutagenesis products of both Y-SNPs. The variants generated by mutagenesis appear underlined and with an asterisk. The ancestral alleles of both Y-SNPs can also be observed in the electropherogram

The reproducibility of the assay was confirmed by analyzing some of the samples twice by two different researchers and using different thermal cyclers, C1000™ Thermal Cycler (Bio-Rad) and 9800 Fast Thermal Cycler (Applied Biosystems). The electropherograms obtained from the same DNA samples showed identical results.

Overall, these results confirm the guiding quality of the method for subtyping haplogroup R1b-DF27 and its potential to be used with forensic samples.

Discussion

In the present study, we introduce a set of 15 Y-SNPs (U106, P312, U152, M529, L238, DF19, DF27, Z196, L617, L881, DF17, Z220, M153, M167, and S68) for fine-resolution subtyping of the Iberian near-specific haplogroup R1b-DF27 in a single minisequencing reaction. Furthermore, this is the first multiplex assay that offers a deep dissection of the paternal lineage DF27, which displays frequencies over 40% in the Iberian Peninsula [8, 10, 11]. The inclusion of its subhaplogroups is of great interest as some of them, like M167 and Z220, show moderate geographical differentiation, being more frequent in Eastern Iberia or North Central Spain respectively. Thus, the typing of DF27 and/or its derived sublineages in forensic samples, in combination with other markers like Y-STRs or ancestry informative markers (AIM), could be of interest for inferring the paternal biogeographical origin of an unknown contributor in a crime scene.

Additionally, the present minisequencing panel can be of special use for the study of Southwest European population introgression in the Latin American populations, as they have been a destination of the historically known Spanish and Portuguese migration [21–23] and other world areas historically influenced by the Spanish presence, like Flanders, Sardinia, or Sicily, or overseas regions like the Philippines.

Conversely, the inclusion in the 15-plex of other R1b- M269 sublineages above DF27 (i.e., U106, P312, U152, M529, L238, and DF19) is also significant since it also allows the analysis of other common Southwest European lineages that are geographically localized [6, 7]. Moreover, given the dispersion of some Southwest European populations (such as Great Britain, Spain, France, or Portugal) over large areas of America, Asia, and Africa throughout history, our panel could allow the study of the European paternal contribution to the genetic substrate of different populations.

Previous Y-SNP panels include haplogroup R1b but not many of its derived subhaplogroups [24–26]; therefore, our design can complete the above mentioned panels in order to provide an increased power of population discrimination with minimal DNA sample consumption. Likewise, the supplementation of this 15-plex minisequencing panel for the Y chromosome with the analysis

of the mitochondrial DNA lineage would complete the information on biogeographical ancestry, which could be of particular interest in the study of admixed individuals.

On the other hand, in the development of any multiplex minisequencing reaction, it is essential to test all the variants. However, this can be difficult when some of the variants display scarce allele frequencies or no samples are available. Therefore, a method should be applied to allow such variants to be included during the optimization of the assay. This is the case with the less frequent haplogroups DF19 and L881. For that reason, we applied site-directed mutagenesis in order to obtain these Y-SNPs. The obtained results confirm that site-directed mutagenesis is a highly appropriate tool to generate rare Y-SNP variants for minisequencing detection, as previously suggested on a mitochondrial DNA minisequencing design [13].

Finally, an advantage of the assay here presented is the reduced number of coamplified fragments, which facilitates the optimization of the method in any forensic laboratory and minimizes the competition effects during the amplification of samples with small quantities of DNA, critical in forensic samples. The size of the PCR amplicons must also be considered. Since it is usual to deal with degraded samples in the forensic routine, short length amplicons are preferred. However, designing amplification primers for the Y chromosome involves additional challenge due to the complex structure of this chromosome. For that reason, we tried to make the amplicon sizes of this assay as short as possible, not exceeding 230 bp. Besides, this design also tried to make the length of the minisequencing primers as short as possible, not exceeding more than 70 bp. This makes this assay a cost-effective approach for genotyping R1b-DF27 and its subhaplogroups, as well as other common Southwest European lineages. Furthermore, although the polymer POP-4® is the most appropriate separation matrix for this type of analysis, we used POP-7®, as it allows both the analysis of fragments and sequences. In any case, the assay was also tested in POP-4®, ensuring that no information was lost and that reliable results are obtained using either polymer.

Conclusion

The 15-plex minisequencing panel provides a robust method for subtyping R1b-DF27 lineage in a single multiplex reaction. We verified that the site-directed mutagenesis products are compatible with minisequencing and, thus, can be used to ascertain the genotyping of the rare variants when control individuals harboring these variants are not available. The high resolution accomplished with this tool enables to improve male lineage discrimination in Iberia, Southwest Europe, and other large areas of the world, as well as further detailed biogeographical and evolutionary inferences. Thus, it can be of relevance for forensic and human population genetics, as well as for

genealogical studies. The short length of the amplicons, its simplicity, and reproducibility allows an easy implementation of the minisequencing panel here presented in most genetic laboratories.

Acknowledgements

The authors are grateful to PhD Maite Álvarez for her human and technical assistance provided on the DNA Bank Service (SGIker) of the University of the Basque Country UPV/EHU (European funding: ERDF and ESF), as well as to the Basque Foundation of Science (BIOEF), and to all the people who voluntarily participated in this study.

Funding information

Funds were provided by the Basque Government (IT-424-07 and IT-833-13). During the execution of this study, Patricia Villaescusa was granted with a predoctoral fellowship by the University of the Basque Country UPV/EHU and Leire Palencia-Madrid with a postdoctoral fellowship by the University of the Basque Country UPV/EHU.

Compliance with ethical standards

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. Jobling MA, Pandya A, Tyler-Smith C (1997) The Y chromosome in forensic analysis and paternity testing. Int J Legal Med 110:118–124

2. Sobrino B, Brión M, CarracedoA(2005) SNPs in forensic genetics: a review on SNP typing methodologies. Forensic Sci Int 154:181–194

3. Underhill PA, Shen P, Lin AA, Jin L, Passarino G, Yang WH, Kauffman E, Bonné-Tamir B, Bertranpetit J, Francalacci P, Ibrahim M, Jenkins T, Kidd JR, Mehdi SQ, Seielstad MT, Wells RS, Piazza A, Davis RW, FeldmanMW, Cavalli-Sforza LL, Oefner PJ (2000) Y chromosome sequence variation and the history of human populations. Nat Genet 26:358–361. https://doi.org/10.1038/81685

4. Kayser M (2017) Forensic use of Y-chromosome DNA: a general overview. Hum Genet 136:621–635. https://doi.org/10.1007/s00439-017-1776-9

5. Myres NM, Rootsi S, Lin AA, JärveM, King RJ, Kutuev I, Cabrera VM, Khusnutdinova EK, Pshenichnov A, Yunusbayev B, Balanovsky O, Balanovska E, Rudan P, Baldovic M, Herrera RJ,

Chiaroni J, di Cristofaro J, Villems R, Kivisild T, Underhill PA (2011) A major Y-chromosome haplogroup R1b Holocene era founder effect in Central and Western Europe. Eur J Hum Genet 19:95–101. https://doi.org/10.1038/ejhg.2010.146

6. Cruciani F, Trombetta B, Antonelli C, Pascone R, Valesini G, Scalzi V, Vona G, Melegh B, Zagradisnik B, Assum G, Efremov GD, Sellitto D, Scozzari R (2011) Strong intra- and inter-continental differentiation revealed by Y chromosome SNPs M269, U106 and U152. Forensic Sci Int Genet 5:e49–e52. https://doi.org/10.1016/j.fsigen.2010.07.006

7. Busby GBJ, Brisighelli F, Sánchez-Diz P et al (2012) The peopling of Europe and the cautionary tale of Y chromosome lineage RM269. Proc Biol Sci 279:884–892. https://doi.org/10.1098/rspb.2011.1044

8. Valverde L, Illescas MJ, Villaescusa P, Gotor AM, García A, Cardoso S, Algorta J, Catarino S, Rouault K, Férec C, Hardiman O, Zarrabeitia M, Jiménez S, Pinheiro MF, Jarreta BM, Olofsson J, Morling N, de PancorboMM(2016) New clues to the evolutionary history of the main European paternal lineage M269: dissection of the Y-SNP S116 in Atlantic Europe and Iberia. Eur J Hum Genet 24:437–441. https://doi.org/10.1038/ejhg.2015.114

9. Rocca RA, Magoon G, Reynolds DF, Krahn T, Tilroe VO, Op den Velde Boots PM, Grierson AJ (2012) Discovery of Western European R1b1a2Y chromosome variants in 1000 genomes project data: an online community approach. PLoS One 7:e41634. https://doi.org/10.1371/journal.pone.0041634

10. Villaescusa P, Illescas MJ, Valverde L, Baeta M, Nuñez C, Martínez-Jarreta B, Zarrabeitia MT, Calafell F, de Pancorbo MM (2017) Characterization of the Iberian Y chromosome haplogroup R-DF27 in Northern Spain. Forensic Sci Int Genet 27:142–148. https://doi.org/10.1016/j.fsigen.2016.12.013

11. Solé-Morata N, Villaescusa P, García-Fernández C, Font-Porterias N, IllescasMJ, Valverde L, Tassi F, Ghirotto S, Férec C, Rouault K, Jiménez-Moreno S, Martínez-Jarreta B, Pinheiro MF, Zarrabeitia MT, Carracedo Á, de Pancorbo MM, Calafell F (2017) Analysis of the R1b-DF27 haplogroup shows that a large fraction of Iberian Y-chromosome lineages originated recently in situ. Sci Rep 7:7341. https://doi.org/10.1038/s41598-017-07710-x

12. Mehta B, Daniel R, Phillips C, McNevin D (2017) Forensically relevant SNaPshot® assays for human DNA SNP analysis: a review. Int J Legal Med 131:21–37. https://doi.org/10.1007/s00414-016-1490-5

13. Palencia-Madrid L, Cardoso S, Castro-Maestre F, Baroja-Careaga I, Rocandio AM, de Pancorbo MM (2018) Development of a new screening method to determine the main 52 mitochondrial haplogroups through a single minisequencing reaction. Mitochondrion S1567-7249:30312–30314. https://doi.org/10.1016/j.mito.2018.02.004

14. Van Oven M, Van Geystelen A, Kayser M et al (2014) Seeing the wood for the trees: a minimal reference phylogeny for the human Y chromosome. Hum Mutat 35:187–191. https://doi.org/10.1002/humu.22468

15. International Society of Genetic Genealogy (2016). Y-DNA haplogroup tree 2016, Version 10.01. http://isogg.org/tree/. Accessed 1 Sep 2017

16. Marshall OJ (2004) PerlPrimer: cross-platform, graphical primer design for standard, bisulphite and real-time PCR. Bioinformatics 20:2471–2472. https://doi.org/10.1093/bioinformatics/bth254

17. Vallone PM, Butler JM (2004) AutoDimer: a screening tool for primer-dimer and hairpin structures. Biotechniques 37:226–231

18. Sanchez JJ, Børsting C, Morling N (2005) Typing of Y chromosome SNPs with multiplex PCR methods. Methods Mol Biol 297:209–228

19. 1000 Genomes Project Consortium, Auton A, Brooks LD et al (2015) A global reference for human genetic variation. Nature 526:68–74. https://doi.org/10.1038/nature15393

20. Larmuseau MHD, Calafell F, Princen SA, Decorte R, Soen V (2018) The black legend on the Spanish presence in the low countries: verifying shared beliefs on genetic ancestry. Am J Phys Anthropol 166:219–227. https://doi.org/10.1002/ajpa.23409

21. Sauer CO (2008) The early Spanish main. Cambridge University Press

22. Noguera MC, Schwegler A, Gomes V, Briceño I, Alvarez L, Uricoechea D, Amorim A, Benavides E, Silvera C, Charris M, Bernal JE, Gusmão L (2013) Colombia's racial crucible: Y chromosome evidence from six admixed communities in the Department of Bolivar. Ann Hum Biol 41:453–459. https://doi.org/10.3109/03014460.2013.852244

23. Morales Padron F (1990) Historia del Decubrimiento y Conquista de América. Gredos

24. Onofri V, Alessandrini F, Turchi C, Pesaresi M, Buscemi L, Tagliabracci A (2006) Development of multiplex PCRs for evolutionary and forensic applications of 37 human Y chromosome SNPs. Forensic Sci Int 157:23–35. https://doi.org/10.1016/j.forsciint.2005.03.014

25. Valverde L, Köhnemann S, Cardoso S, Pfeiffer H, de PancorboMM (2013) Improving the analysis of Y-SNP haplogroups by a single highly informative 16 SNP multiplex PCR-minisequencing assay. Electrophoresis 34:605–612. https://doi.org/10.1002/elps.201200433

26. Bouakaze C, Keyser C, Amory S, Crubézy E, Ludes B (2007) First successful assay of Y-SNP typing by SNaPshot minisequencing on ancient DNA. Int J Legal Med 121:493–499. https://doi.org/10.1007/s00414-007-0177-3

Electronic supplementary material

Supplementary Tables

Supplementary Table S1. Characteristics of the primers and miniprimers used in the 15-plex minisequencing panel.

Y-SNP	db SNP ID	PCR reaction	Sequence (5'-3')	Primer Sense	Final concentration (µM)	Amplicon size (bp)	P	P+T
U106	rs16981293	Amplif.	TTCCTGAATAGCAAATCCCA	FW	0.2	96	20	25
			GCTGTATGTGTCTTCCTGTG	RV				
		Minisec.	GACAATAGCAAATCCCAAAGCTCCA	FW	0.1			
P312	rs34276300	Amplif.	TCCTGCTAATGTATCTGCTG	FW	0.2	115	21	55
			CTCATTTATCACCTCAGTGC	RV				
		Minisec.	CTAAACTAGGTGCCACGTCGTGAAAGTCTGACAAGCTAATGTATCTGCTGCACTG	FW	0.3			
DF27	rs577478344	Amplif.	TTGGCTGGATATGAAATTCTGG	FW	0.1	122	20	25
			GAAGCCCATCAGATTAACAGAG	RV				
		Minisec.	GACAATGGCTTGTAGAGTTTCTGCC	FW	0.2			
U152	rs1236440	Amplif.	AGAAACATTCCACGCTTGAG	FW	0.2	103	26	40
			ATGGTAGTTTAATGGGAGTAGC	RV				
		Minisec.	TGAAAGTCTGACAATCTATACATTACTTTGAGAAGTATGG	FW	0.3			
M529	rs11799226	Amplif.	TAAACCCTCCTCAGCAACAG	FW	0.2	150	20	50
			GGAAGCATTCAGAAGCAGGT	RV				
		Minisec.	ACTAGGTGCCACGTCGTGAAAGTCTGACAAAACAACCGCTCTCTCAGACA	FW	0.1			
L238	rs35199432	Amplif.	AAGAAATGTCAACGGTACAGAG	FW	0.2	125	21	60
			CATACACATTCACAGCAGGT	RV				
		Minisec.	ACTGACTAAACTAGGTGCCACGTCGTGAAAGTCTGACAACACATTCACAGCAGGTAACTG	RV	0.2			
DF19	rs753249165	Mutagen.	GTGAGGGGCCAATAACGG	FW	0.2	95	18	45
		Amplif.	AAAGGGCACTCTGATAGGAC	FW				
			TCCCTATTCAGCCATCTTAGC	RV				
		Minisec.	AGGTGCCACGTCGTGAAAGTCTGACAAGTCCGTTATTGGCCCCTC	RV	0.1			
Z196	–	Amplif.	AACTGTAAGTCTATGCTGCT	FW	0.2	106/104	20	60
			ACAGACTGGTTCTGCTTATGT	RV				
		Minisec.	AACTGACTAAACTAGGTGCCACGTCGTGAAAGTCTGACAACCCAATGGGACACATCACAC	RV	0.1			
L881	–	Amplif.	TGGCTGTGGCTTTACTTTCTG	FW	0.2	211	17	40
			GCAGGACAACCCTTCTTTGA	RV				
		Mutagen.	TACCCGGGGTGCTTCTG	RV				
		Minisec.	GCCACGTCGTGAAAGTCTGACAATCCTCAGAAGCACCCCG	FW	0.1			
L617	–	Amplif.	ACAACAGCACTACTGGACAC	FW	0.3	210	20	35
			TCCCTTCATCCTGAGCTTCA	RV				
		Minisec.	GTGAAAGTCTGACAAGAAGCCAGTCCAAGGTGTGA	FW	0.3			
Z220	rs538725564	Amplif.	TCTCTAACTTCTGGCTTCAAGTG	FW	0.2	102	23	30
			TGGAATGATATCAGCTTCCATGTC	RV				
		Minisec.	CTGACAACACCTCGGCCTCTGTTTTATAA	FW	0.1			
M153	rs375151448	Amplif.	ATTGTCTCCTTTAAGTGGGT	FW	0.3	113	25	65
			TTAATCTGACTTGGAAAGGG	RV				
		Minisec.	AACTGACTAAACTAGGTGCCACGTCGTGAAAGTCTGACAAACACCAATGGTCCTATCTTAATGAA	FW	0.3			
M167	rs1800865	Amplif.	GGAGTGACAACCAAGAAGAG	FW	0.3	229	20	30
			TTTCAAGCTCTGGTTCTGTG	RV				
		Minisec.	AGTCTGACAAAAGGAAGCCCCACAGGGTGC	RV	0.2			
S68	rs775040950	Amplif.	TGTCAGATGCTTAATTGTGTTTC	FW	0.1	62	24	35
			CAGGAGTTATGTGAGGACCC	RV				
		Minisec.	AAGTCTGACAATGTCAGATGCTTAATTGTGTTTCC	FW	0.1			
DF17	rs754186919	Amplif.	ATTAGCCAACTGTAATCTTGGTTAC	FW	0.1	189	17	40
			AGACAGAATCTTATTCCATCACCC	RV				
		Minisec.	TAGGTGCCACGTCGTGAAAGTCTGACAAGGATTGTGTCACTGCGC	FW	0.2			

*These primers are used only for the directed mutagenesis of this Y-SNP.
Amplif.: Amplification primer
Minisec.: Minisequencing primer
Mutagen.: Mutagenesis primer
P: lenght of the primer (bp).
P+ T: minisequencing product lenght, including primer and tag (bp).

Supplementary Table S2. Frequencies (%) of the Y haplogroups included in the 15-plex minisequencing panel in some populations from South, West and Central Europe extracted from the literature.

Population	N	U106	P312[1]	U152	M529	L238	DF19	DF27[2]	Z196[3]	L617	L881	DF17	Z220[4]	M153	M167	S68	Reference
Spain																	
Alicante	115	4,3	7,8	6,0	0,0	0,0	0,0	43,1	–	–	–	–	–	–	–	–	8
Alicante 2	142	–	16,2	–	–	–	–	16,2	1,4	–	–	–	9,9	4,9	9,9	–	11
Andalucía	100	3,0	13,0	4,0	0,0	0,0	0,0	28,0	4,0	–	–	–	9,0	4,0	2,0	–	8, 11
Aragon	92	–	26,1	–	–	–	–	15,2	5,4	0,0	0,0	1,1	6,5	1,1	4,4	1,1	10
Asturias	63	0,0	0,0	7,9	6,3	0,0	0,0	30,2	11,1	0,0	0,0	0,0	1,6	0,0	0,0	0,0	8, 10
Barcelona 1	100	2,0	10,0	6,0	1,0	0,0	0,0	48,0	–	–	–	–	–	–	–	–	8
Barcelona 2	571	–	21,8	–	–	–	–	9,9	6,5	–	–	–	13,2	1,1	9,2	–	11
Basque Country Natives	229	1,3	15,8	2,2	2,2	0,0	0,0	31,0	5,2	1,8	0,0	0,4	21,8	6,6	3,1	0,9	8, 10
Basque Country Residents	111	1,8	6,3	1,8	1,8	0,0	0,0	24,3	5,4	0,0	0,0	0,0	10,8	0,9	6,3	0,0	8, 10
Cantabria	96	2,1	7,3	4,2	6,3	0,0	0,0	22,9	3,1	0,0	0,0	0,0	16,7	0,0	2,1	0,0	8, 10
Castelló	49	–	15,7	–	–	–	–	6,6	8,5	–	–	–	17,2	2,1	12,8	–	11
Galicia	70	4,3	0,0	8,6	7,1	0,0	0,0	24,3	7,1	–	–	–	8,6	0,0	0,0	–	8, 11
Girona	131	–	21,5	–	–	–	–	5,9	2,4	–	–	–	10,2	0,8	9,4	–	11
Lleida	104	–	23,7	–	–	–	–	10,8	5,0	–	–	–	15,8	1,0	8,9	–	11
Madrid	99	2,0	7,1	4,0	1,0	0,0	0,0	21,2	10,1	–	–	–	12,1	3,0	3,0	–	8, 11
Mallorca	48	–	22,9	–	–	–	–	16,7	2,1	–	–	–	16,7	0,0	10,4	–	11
Pyrenees	46	–	27,2	–	–	–	–	7,6	6,5	–	–	–	17,4	0,0	10,9	–	11
València	79	–	25,9	–	–	–	–	12,5	9,2	–	–	–	14,7	1,3	3,1	–	11
Portugal																	
Porto	109	2,7	10,1	3,6	2,7	0,0	0,0	32,1	5,5	–	–	–	0,9	0,0	1,8	–	8, 11
Lisbon	100	7,0	38,0	3,0	3,0	–	–	–	–	–	–	–	–	–	–	–	5
France																	
Alsace	80	–	31,3	–	–	–	–	0,0	6,3	–	–	–	0,0	0,0	1,3	–	11
Auvergne	89	–	43,8	–	–	–	–	1,1	3,4	–	–	–	0,0	0,0	1,1	–	11
Brittany	145	–	6,2	4,1	52,4	0,7	0,0	9,7	3,5	–	–	–	4,1	0,0	2,1	–	8, 11
Île-de-France	91	–	36,7	–	–	–	–	0,0	9,0	–	–	–	0,0	0,0	1,2	–	11
Midi-Pyrénées	67	–	42,8	–	–	–	–	0,0	3,2	–	–	–	0,0	0,0	7,5	–	11
Nord-Pas-de-Calais	68	–	36,8	–	–	–	–	0,0	7,8	–	–	–	1,5	0,0	3,2	–	11
Provence–Alpes-Côte d'Azur	45	–	40,1	–	–	–	–	0,0	7,4	–	–	–	0,0	0,0	4,8	–	11
Var (coastal, E of Rhone)	68	5,9	35,3	19,1	2,9	–	–	–	–	–	–	–	–	–	–	–	5
Vaucluse (upstream Rhone)	61	6,6	29,5	14,8	8,2	–	–	–	–	–	–	–	–	–	–	–	5
Bouches du Rhone (at mouth)	207	8,2	32,4	16,9	6,3	–	–	–	–	–	–	–	–	–	–	–	5
Alpes de Haute Provence	31	12,9	29,0	12,9	19,4	–	–	–	–	–	–	–	–	–	–	–	5
Switzerland (Lower Rhone Valley	51	11,8	7,8	15,7	2,0	–	–	–	–	–	–	–	–	–	–	–	5
Netherlands	87	36,8	6,9	3,4	5,7	–	–	–	–	–	–	–	–	–	–	–	5
Belgium (Flanders)	1087	–	–	–	–	–	–	–	–	–	–	–	4,1	–	1,0	–	20
Austria	18	22,2	0,0	0,0	5,6	–	–	–	–	–	–	–	–	–	–	–	5
Germany (West)	100	24,0	10,0	14,0	1,0	–	–	–	–	–	–	–	–	–	–	–	5
Italy (TSI)	53	3,8	3,8	26,4	0,0	–	–	1,9	3,8	–	–	–	1,9	0,0	0,0	–	19
England (GBR)	46	19,6	8,7	8,7	19,6	–	–	6,5	4,4	–	–	–	0,0	0,0	4,4	–	19
Ireland	146	6,2	17,8	2,1	54,1	0,0	0,0	0,0	0,7	–	–	–	0,0	0,0	0,0	–	8, 11
Denmark	174	17,8	16,7	–	–	–	–	–	–	–	–	–	–	–	–	–	8

[1]: P312 (xU152xM529xDF27xL238xDF19)

[2]: DF27 (xZ196xL617xL881)

[3]: Z195 (xDF17xZ220xM167xS68)

[4]: Z220 (xM153)

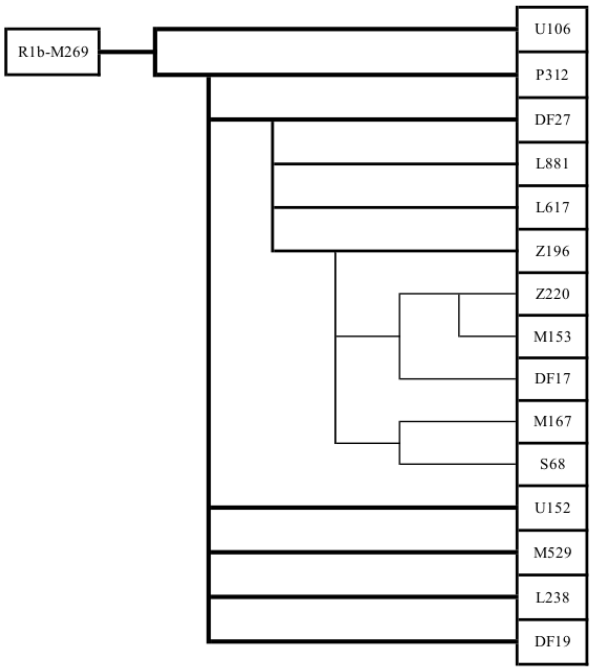
Supplementary Table S3. Ancestral and derived alleles of all the 15 Y-SNPs included in the minisequencing assay. A colour key to facilitate the interpretation of the results in the electropherograms is also provided.

Y-SNP	MUT	Sense	Key to electropherogram	
			Ancestral	Derived
U106	C/T	FW	C	T
P312	C/A	FW	C	A
DF27	G/A	FW	G	A
U152	C/T	FW	C	T
M529	C/G	FW	C	G
L238	A/G	RV	T	C
DF19	G/T	RV	C	A
Z196	T/C (2bp deletion)	RV	A	G
L881	A/G	FW	A	G
L617	G/A	FW	G	A
Z220	G/A	FW	G	A
M153	A/T	FW	A	T
M167	G/A	RV	C	T
S68	C/T	FW	C	T
DF17	T/G	FW	T	G

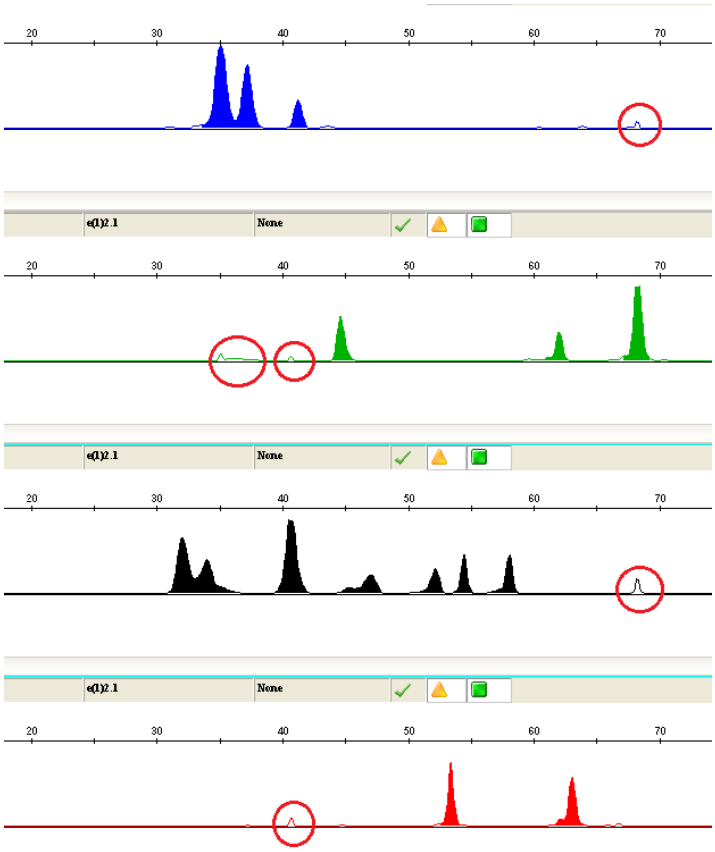
Supplementary Table S4. Origin and genotyping information of the samples used to test the 15-plex minisequencing panel.

Analyzed population sample	YHRD accesion number of the population sample	Genotyping technique	Y-SNP	db SNP ID	Geographic distribution
Basque Country, Iberian Peninsula [8]	-	HRM & Sequencing	U106	rs16981293	North and Central Europe [5, 6]
Basque Country, Iberian Peninsula [8]	YA004063	HRM & Sequencing	P312	rs34276300	South-Western Europe [5, 8]
Basque Country, Iberian Peninsula [8, 10]	YA004063	HRM & Sequencing	DF27	rs577478344	Iberian Peninsula and South-West France, Latin American populations [8, 10, 11]
Generated by site-directed mutagenesis	-	-	L881	-	*
Basque Country, Iberian Peninsula [10]	YA004063	Sequencing	L617	-	*
Basque Country, Iberian Peninsula [10]	YA004063	HRM & Sequencing	Z196	-	*
Basque Country, Iberian Peninsula [10]	YA004063	HRM & Sequencing	Z220	rs538725564	Iberian Peninsula, Basque Country [10, 11]
Basque Country, Iberian Peninsula [10]	YA004063	HRM & Sequencing	M153	rs375151448	Basques, Gascons, Iberian Peninsula [10, 11]
Basque Country, Iberian Peninsula [10]	YA004063	Sequencing	DF17	rs754186919	*
Barcelona, Iberian Peninsula [11]	YA004063	HRM & Sequencing	M167	rs1800865	Basque country, Catalonia, Pyrenees [10, 11]
Basque Country, Iberian Peninsula [10]	YA004063	HRM & Sequencing	S68	rs775040950	*
Basque Country, Iberian Peninsula [8]	YA004063	HRM & Sequencing	U152	rs1236440	Central Europe, North and Central Italy and the Alps [5-8]
Basque Country, Iberian Peninsula [8]	-	HRM & Sequencing	M529	rs11799226	British islands and Britanny [7, 8]
Britanny, France [8]	-	HRM & Sequencing	L238	rs35199432	Scandinavia [8, 9]
Generated by site-directed mutagenesis	-	-	DF19	rs753249165	*

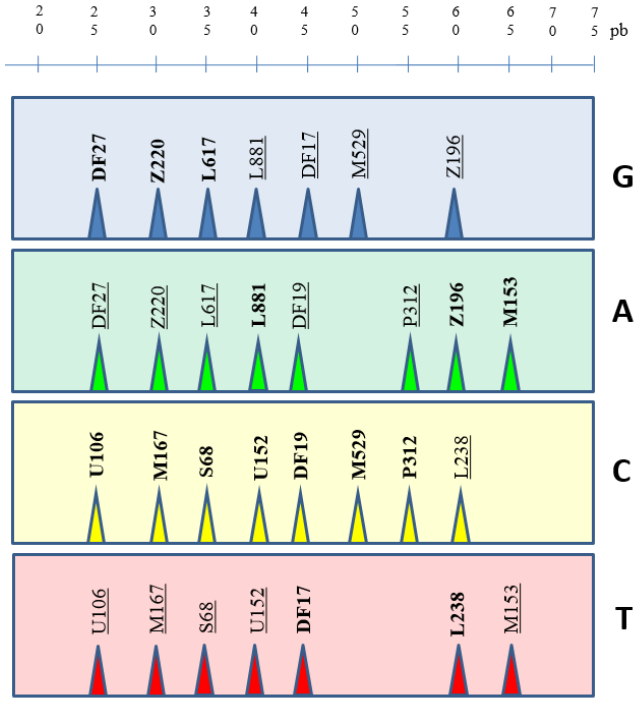
Supplementary Figures



Supplementary Fig. S1 Phylogenetic tree of the Y-SNPs included in the 15-plex minisequencing panel.



Supplementary Fig. S2 Electropherogram obtained during the optimization of the 15-plex minisequencing assay, where spurious peaks and pull-ups (circled) can be observed.



Supplementary Fig. S3 Diagram of the theoretical positions of the Y-SNPs included in the 15-plex minisequencing panel. Ancestral alleles appear in bold letters; Derived alleles appear underlined.

4.5 Study Number 5

'Assessment of a subset of Slowly Mutating Y-STRs for forensic and evolutionary studies'

The present study corresponds to the attainment of the objective 4*: To design, optimize and validate a novel panel of six Slowly Mutating Y-STRs, which can be used in conjunction with the existing multiplex commercial kits for forensic casework, particularly in complex kinship cases and in optimizing the prediction of paternal ancestry based on currently Y-STR panels with medium-high mutation rate.*

The Y-STRs commonly included in commercial kits are highly polymorphic in most populations. In the last years the need to distinguish among close male relatives stimulated the search for Y-STRs with higher mutations rates, known as Rapidly Mutating (RM) Y-STRs. However, these type of Y-STRs are not the best candidates to be included in phylogenetic studies, since they could accelerate the molecular clock estimations due to their higher mutations rates. For that reason, the need to use more stable Y-STRs with lower mutation rates, called Slowly Mutating (SM) Y-STRs, arose in the forensic community, which could complement the routine Y-STR panels especially in exclusion cases where minimal discrepancies are critical.

In the present work a subset of six SM Y-STR loci (DYS388, DYS426, DYS461, DYS485, DYS525 and DYS561) were selected and evaluated in 628 individuals from Asia (Thailand), Central and South America (Amerindians from Guatemala; Hispanics from Colombia and Nicaragua), Africa (Malawi) and Europe (Spain). The sensitivity of the panel was assessed by analyzing serial dilutions of human control DNA. Stability was evaluated by adding two common inhibitors, humic acid and hematin, in the reactions. Additionally, we also analyzed the genetic variability of the selected SM Y-STRs as well as Y-SNPs to assess the correspondence between SM Y-STRs haplotypes and haplogroups through the populations.

The obtained results demonstrate that the novel set of SM Y-STR is a reproducible, sensitive and robust multiplex system for forensic applications in combination with the common commercial panels, particularly for confirming exclusions in biological kinship cases with minimal discrepancies in one or a few loci, since mutation events are rarer to occur in these markers. The SM Y-STR multiplex provided a moderate discrimination power between haplotypes in most of the studied populations, despite the low mutation rate. In addition to that, although the use of the SM Y-STRs for the prediction of Y chromosome haplogroups is not able to reach the same resolution as the Y-STRs included in current panels, the use of our multiplex in combination with them may help

optimize the resolution of the phylogenetic relationships as these markers are more stable than other common Y-STR markers.

Overall, the SM Y-STRs panel has demonstrated to be a robust tool for forensic applications and can be useful in conjunction with current common Y-STR panels. Furthermore, our study also provided an extensive Y-STR haplotype and allele reference dataset for future use in forensics.

This study has resulted in an international publication in the journal *Forensic Science International: Genetics* under the heading '*Assessment of a subset of Slowly Mutating Y-STRs for forensic and evolutionary studies*' in May 2018. Q1, IP: 5.637. The publication is shown below.

Forensic Science International: Genetics 34 (2018) e7–e12

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsigen

Short communication

Assessment of a subset of Slowly Mutating Y-STRs for forensic and evolutionary studies

Miriam Baeta[a], Carolina Núñez[a], Patricia Villaescusa[a], Urko Ortueta[a], Nerea Ibarbia[a], Rene J. Herrera[b], José Luis Blazquez-Caeiro[c], Juan José Builes[d,e], Susana Jiménez-Moreno[f], Begoña Martínez-Jarreta[g], Marian M. de Pancorbo[a,*]

[a] BIOMICs Research Group, Lascaray Research Center, University of the Basque Country UPV/EHU, Avda. Miguel de Unamuno, 3, 01006, Vitoria-Gasteiz, Spain.

[b] Department of Molecular Biology, Colorado College, Colorado Springs, CO, 80903, USA.

[c] Departament of Zoology and Physical Anthropology, University of Santiago de Compostela, Spain.

[d] Genes SAS Laboratory, Medellín, Colombia.

[e] Institute of Biology, University of Antioquia, Medellín, Colombia.

[f] Área Medicina Legal y Forense, Departments of Patología y Cirugía, Universidad Miguel Hernández de Elche, Elche, Alicante, Spain.

[g] Department of Forensic Medicine, University of Zaragoza, Zaragoza, Spain.

[*] Corresponding author.

Received 6 April 2017, Received in revised form 22 November 2017, Accepted 19 March 2018, Available online 20 March 2018.

Abstract

Y-specific short tandem repeat (Y-STR) loci display different mutation rates and consequently are suitable for forensic, genealogical, and evolutionary studies that require different levels of timelines and resolution. Recent efforts have focused on implementing Rapidly Mutating (RM) Y-STRs to assess male specific profiles. However, due to their high mutation rate their use in kinship testing or in phylogenetic studies may be less reliable. In the present study, a novel Slowly Mutating Y-STR (SM) panel, including DYS388, DYS426, DYS461 (Y-GATA-A7.2), DYS485, DYS525, and DYS561, has been developed and evaluated in a sample set of 628 unrelated males from different worldwide populations. This panel is reproducible, sensitive, and robust for forensic

applications and may be useful in conjunction with the common multiplexes, particularly in exclusion of kinship cases where minimal discrimination is reported employing the rapidly mutating Y-STR systems. Furthermore, SM Y-STR data may be of value in evolutionary studies to optimize the resolution of phylogenetic relationships generated with current Y-STR panel sets. In this study, we provide an extensive Y-STR allele and haplotype reference dataset for future applications.

Keywords

Slowly mutating Y-STRs; Allele and haplotype reference dataset; Kinship; Phylogenetic trees analyses.

1. Introduction

Y chromosome markers are suited for forensic and genealogical applications, as well as for ascertaining human evolution and migration events through paternal lineages [1–3]. The non-recombining nature of the male-specific region of the Y chromosome (MSY), as well as the reliability of the molecular clock, enables the reconstruction of haplotype genealogies thorough history [4–6]. The so-called molecular clock is based on the fact that average mutation rates in haplotypes are nearly constant over time. Mutations seem to occur randomly, not depending on particular haplogroups, populations, or time periods [7].

The commonly used Y-STR (Short Tandem Repeat) loci are highly polymorphic in most populations, largely due to their hypermutability, and display mutation rates with values between 10^{-4} and 10^{-2} per locus per generation (Y-Chromosome STR Haplotype Reference Database, YHRD). Until recently, most of the Y-STRs selected for evolutionary, forensic, and genealogical studies exhibit low to midrange mutation rates ($\sim 10^{-3}$), allowing to identify closely related male lineages. However, the need to distinguish among close male relatives has stimulated the search for Y-STR markers with higher mutation rates [8,9]. These markers, known as Rapidly Mutating (RM) Y-STRs, display mutation rates of $\sim 10^{-2}$ per locus per generation. Despite the fact that some of these new loci are already included in widely expanded commercial kits, their application in paternity testing or in missing person identification (when comparison with potential relatives is performed) may not be reliable, due to false exclusions resulting from their high mutation rate [10].

Similarly, RM Y-STRs do not constitute the best candidates to be included in phylogenetic studies, as they could accelerate the molecular clock estimations. In addition, the development of new panels which include more stable or Slowly Mutating (SM) Y-STRs may be useful as a complementary tool to the current Y-STR panels in forensic casework [10], particularly in exclusion

cases where minimal discrepancies are considered critical and reported as exclusions. Likewise, the low mutation rate of SM Y-STRs may provide a higher refinement in the construction of phylogenetic trees linking Y chromosome lineages, since the chance of a random convergence of SM haplotypes is lower compared to other Y-STR markers. Thus, stronger phylogenetic signals may be detected as the number of reticulations and complexities in the networks are reduced [11].

In the present study, we selected a subset of six SM Y-STR loci (DYS388, DYS426, DYS461, DYS485, DYS525, and DYS561) and evaluated its performance in a large number of individuals of Caucasian, Native American, Hispanic, Asian, and African ancestry.

2. Materials and methods

2.1 Selection of Y-STR markers and primer design

Six Y-STR markers with suitable characteristics were selected from the 186 Y-STRs examined in the extensive study of mutability of Ballantyne and cols. [9]: DYS388, DYS426, DYS461 (Y-GATA A7.2), DYS485, DYS525, and DYS561. The main criteria for marker selection were a low mutation rate ($\sim 10^{-4}$ mutations/generation) [12], as well as a gene diversity generally > 0.4 according to the data reported in the literature [13–19]. The primers for the six Y-STRs were designed in order to obtain amplicons under 250 pb using PerlPrimer software v.1.1.21 [20] (Supplementary Table S1). The lack of interactions between primers and specificity for the Y chromosome was checked with Autodimer v.1.0 software [21] and BLASTN alignment tool (http://blast.ncbi.nlm.nih.gov/Blast.cgi), respectively. Forward primers for each marker were modified by the addition of a fluorescent dye at their 5' end: 5-FAM (Abs. = 495 nm; Em. = 520 nm), YAKIMA YELLOW (Abs. = 530 nm; Em. = 549 nm), ATTO 550 (Abs. = 554 nm; Em. = 576 nm), and ATTO 565 (Abs. = 563 nm; Em. = 592 nm) (Eurofins Genomics, Ebersberg, Germany) (Supplementary Table S1). Selected Y-STRs were distributed in the multiplex by expected amplicon length, using a four-dye chemistry. The purpose of this design was to create an open system, where other markers of interest could be easily fitted along the four-dye layout, if needed, in order to complement the multiplex.

2.2 Singleplex reaction

Each primer pair was initially tested in a singleplex PCR reaction using the 2800M control DNA (Promega Corporation, Madison, WI). The reaction consisted of 5 µl of QIAGEN Multiplex PCR kit (Qiagen, Valencia, CA), 0.5 µl of primer mix (final concentration of 0.2 µM), 1 ng of genomic DNA, and Milli-Q water for a final reaction volume of 10 µl. PCR was performed in a GeneAmp 9800 (AB/LT/TFS: Applied Biosystems™, Life Technologies, ThermoFisher Scientific, Waltham, MA, USA)

under the following cycle conditions: an initial denaturation at 95 °C for 15 min was followed by 30 cycles of 94 °C for 30 s, 65 °C for 90 s, 72 °C for 90 s, and a final extension at 72 °C for 10 min. DNA amplification success was evaluated by gel electrophoresis on 1.5% agarose gels, visualized with GelRed (3% µL/ml) (Biotium Inc., Hayward, USA) and UV light (UVItec Cambridge). PCR products were purified using 0.5 µl EXO (Exonuclease I) and 2.5 µl SAP (Shrimp Alkaline Phosphatase) (Takara Bio Inc., Japan) in 10 µl of PCR product. Sequencing was performed using the BigDye Terminator v3.1 Cycle Sequencing Kit (AB/LT/TFS) to confirm the specific amplification of each Y-STR loci. Capillary electrophoresis was conducted on an ABI3130 Genetic Analyzer using the Sequencing Analysis 5.2 software (AB/LT/TFS).

2.3 Multiplex PCR amplification, electrophoresis, and data analysis

The multiplex PCR amplification was carried out following the same conditions described for the singleplex reaction, using 0.2 µM of each PCR primer (Supplementary Table S1). Fluorescently labeled PCR products (1 µl) were mixed with 0.5 µl of Genescan 500 LIZ size standard (AB/LT/TFS) and 9 µl Hi-Di formamide and separated by capillary electrophoresis on an ABI3130 Genetic Analyzer (AB/LT/TFS). Fragment size determination and allele designation was performed with GeneMapper ID v.4.0 software (AB/LT/TFS) and Gene Scan 500 LIZ (AB/LT/TFS) as internal size standard. Panel for GeneMapper ID software were constructed through the electrophoresis analysis of reference samples with alleles of known size, since they were previously sequenced. Bins were updated when new alleles were found in the population study. Y-STR typing quality control was assured with the simultaneous electrophoresis analysis of samples with known SM Y-STR profile.

2.4 Sensitivity and stability studies

For the sensitivity study serial dilutions of the 2800M human control DNA were analyzed in triplicate: 10 ng/µl, 1.6 ng/µl, 1 ng/µl, 400 pg/µl, 200 pg/µl, 100 pg/µl, 50 pg/µl, and 25 pg/µl. To examine the stability and robustness of the SM Y-STR multiplex, two common PCR inhibitors, humic acid and haematin, which may be found in forensic casework samples, were added to the amplification reactions. The study was performed using duplicate samples with 1 ng of 2800M control DNA and the following concentrations of inhibitors: 5000 µM, 3000 µM, 1500 µM, 1000 µM, 750 µM, 500 µM, 300 µM, 150 µM, and 100 µM of haematin (Sigma-Aldrich Corporation, St. Louis, MO, USA); and 3000 ng/µl, 2000 ng/µl, 1000 ng/µl, 500 ng/µl, 300 ng/µl, 250 ng/µl, 200 ng/µl, 100 ng/µl, 50 ng/µl, and 25 ng/µl of humic acid (Sigma-Aldrich Corporation, St. Louis, MO, USA).

2.5 Population study

In order to determine the genetic variability of the six Y-STRs analyzed herein, relevant population groups were studied for these markers. A total of 628 samples of unrelated males were included from populations of Asia (Thailand, Bangkok, N = 102), Central and South America (native Americans from Guatemala, Mayans, N = 50; Hispanics from Colombia, N = 60; and from Nicaragua, N = 66), Africa (Chewas from Malawi, Lilongwe, N = 31), and Europe (Caucasians from Spain: Alicante, N = 50; Barcelona, N = 54; Madrid, N = 62; resident and autochthonous individuals from the Basque Country, N = 53 and N = 100, respectively). Two different groups (resident and autochthonous) were considered for Basque Country population according to previous reports [22,23], which described significant differences among them. The inclusion criteria used in order to define autochthonous Basques were the Basque origin of the surnames of the individuals and the geographical origins of their ancestors (at least until the third generation back) within the Basque area. The resident group corresponds to those individuals that live in the Basque Country but whose paternal ancestors are not native Basques.

All samples were collected from healthy volunteer donors after informed consent according to the ethical guidelines of the Helsinki Declaration. Samples from Alicante were provided by the University Miguel Hernández (Spain), and samples from Aragon and Nicaragua were obtained from the University of Zaragoza (Spain). Samples from Madrid and Barcelona were provided by the Spanish National DNA Bank Carlos III (BNADN Ref. 12/0031) (Spain), and samples from Thailand, Colombia, and Africa were provided by Colorado College (US), University of Antioquia (Colombia), and University of Santiago de Compostela (Spain), respectively. Samples from the Basque Country and Guatemala were part of the collection from BIOMICs Research Group of the University of Basque Country (Spain). The study was approved with the favorable ethical reports from the Faculty of Pharmacy of the University of the Basque Country, signed on 26th September 2008, the Independent Ethics Committee Zugueme No PROZU315-12 (Guatemala C.A.) in 2012 and the Colorado College IRB on August 29, 2014.

Extracted genomic DNA was quantified by using the Scientific NanoDrop™ 1000 Spectrophotometer (ThermoFisher Scientific Inc., Wilmington, DE) and then diluted to a 1.5 ng/µl concentration.

2.6 Forensic parameters and statistical analysis

Allele frequencies and genetic diversity (GD) for each locus were calculated using Arlequin software v.3.5.2.2 [24]. Pairwise allelic comparisons were calculated as the number of alleles

which differ between all possible pairs of samples in each population group, using an in-house developed macro in Microsoft Excel 2016. The discrimination capacity (DC) (number of different haplotypes observed in a given population) was calculated by dividing the number of different haplotypes by the total number of individuals in the population. Pairwise R_{ST} genetic distances and the corresponding *p* values between the populations were determined employing Arlequin software. Significance *p* values were adjusted with the sequential Bonferroni correction (α = 0.05/{[(1+n)/2]*n}; n=number of populations) [25] in order to account for potential Type I errors due to the multiple comparisons performed. Pairwise R_{ST} genetic distances were visualized by a heatmap plot obtained with the R statistical package included in Arlequin. A Non-Metric Multi-Dimensional-Scaling plot (NMDS), based on pairwise R_{ST} genetic distances, was obtained using PAST software v.3.04 [26] and the x-y-z coordinates were represented using the rgl package (http://cran.r-project.org/package=rgl) for R software [27].

2.7 Y-SNP analysis

In order to evaluate the correspondence of the SM Y-STR haplotypes and Y chromosome single nucleotide polymorphism (Y-SNP) haplogroups across populations, 319 samples, which were not previously YSNP typed, were analyzed for the following Y-SNPs: CDEF-M168, DE-M145, C-M130, E-P170, H1-M69, G-M201, IJ-P126, I-M258, KLT-M9, T-M272, L-M11, N-M231, O-M175, P1-M45, Q-M242, Q1a2-M3, R-M207, and R1b-M269. The nomenclature of the genotyped mutations follows the minimal reference phylogeny for the human Y chromosome [28], supplemented with the ISOGG v12.166 haplogroup tree (http://www.isogg.org/tree). The analysis was performed using the 16 Y-SNP multiplex PCR-minisequencing assay described in Valverde and cols. [29] or High Resolution Melting (HRM). The primers used for the amplification of each Y-SNP in HRM with the corresponding annealing temperature are shown in Table S2. The conditions for the HRM analysis of Y-SNPs are described in Villaescusa and cols. [23]. Y-SNP data from the remaining samples (N=309) can be found in [23,30,31]. The Factorial Correspondence Analysis was computed using the software Genetix v.4.05.2 [32], showing relationship among the multilocus genotypes and the sample haplogroup.

3. Results and discussion

In the present study, the development and evaluation of a novel panel including six Slowly Mutating (SM) Y-STRs are reported. Parameters of forensic and phylogenetic interest were obtained for this subset of markers in populations from four different continents.

3.1 Multiplex development and validation studies

The novel multiplex developed herein includes six Y-STR loci (DYS388, DYS426, DYS461, DYS485, DYS525, and DYS561), which were selected among the 186 Y-STRs included in the comprehensive study of mutability of Ballantyne and cols. [9].

The main criterion for the marker selection was to choose relatively stable or SM Y-STRs with a low mutation rate of $\sim 10^{-4}$ mutations/generation [9]. Furthermore, as a second condition, the Y-STRs selected should present intra-population gene diversity in groups from different origin. The six best candidate loci selected mostly displayed a gene diversity > 0.4 in different worldwide population groups, according to the scarce data reported in the literature for these markers [13–19]. Given the priority to the stability criterion, candidates with higher gene diversity were not found. Detailed information for the six Y-STRs selected is outlined in Supplementary Table S1 and the final multiplex design is displayed in Fig. S1. The distribution of the 6 SM Y-STR allows the addition of other potential Y-STR markers of interest.

Sensitivity and stability studies were performed to evaluate the performance of the panel for forensic casework (Fig. S2). Sensitivity studies of the multiplex system allowed setting up the minimum quantity of DNA recommendable to obtain complete genetic profiles. Complete profiles, with peak heights above 50 RFU, were obtained with DNA input starting from 200 pg. Lower sample inputs resulted in allelic drop-out events. The stability of the new panel in the presence of two common inhibitors in forensic casework, such as humic acid and haematin, was also evaluated. Full genetic profiles were obtained with ≤ 500 ng/µl of humic acid or ≤ 500 µM of haematin using replica samples. These results demonstrated the sensitivity and robustness of this new multiplex panel.

3.2 Y-STR population study

A total sample set of 628 males representing populations from four different continents was studied to examine the allele diversity of the new set of SM Y-STRs (DYS388, DYS426, DYS461, DYS485, DYS525, and DYS561). The haplotypes obtained with the novel panel are provided in Supplementary Table S3. Populations from Spain were pooled together since non- significant differences were observed among them ($p > 0.0033$, after Bonferroni correction for multiple comparisons), with the exception of autochthonous Basques which shown significant differences and, therefore, it was treated as a single group.

Allele frequency and gene diversity for each locus in the analyzed populations are given in Supplementary Table S4. The number of alleles for each locus in the whole dataset and in each

population is compared in Fig. S3. In the whole population sample set, the six loci exhibited a range of alleles up to 9. The gene diversity per locus ranged from 0.42 to 0.54, except for the marker DYS388, which displayed the lowest average gene diversity value (~ 0.29). The most discriminative marker differed among the analyzed populations: DYS461 was the most variable locus for the Thailand, Spain, and the autochthonous Basque groups; DYS485 for the three American populations; and DYS561 for Malawi.

Haplotype diversity (HD) and the haplotype discrimination capacity (DC) were evaluated for each population, in addition to the number of different haplotypes (Supplementary Table S5). The populations from Africa and Asia displayed the highest haplotype diversity values for the SM Y-STR multiplex (0.9527 ± 0.0213 and 0.9493 ± 0.0132, respectively). Among the Latin American populations, the group from Nicaragua displayed higher diversity (0.9235 ± 0.0281) than the Hispanic group from Colombia (0.8949 ± 0.0353) and Native Americans from Guatemala (0.8392 ± 0.0463). Finally, for the Caucasian groups, the general population from Spain showed a diversity markedly higher (0.8437 ± 0.0246) than the Basques autochthonous, which displayed the lowest values (0.6990 ± 0.0505) of all the studied populations. The differences observed in haplotype diversity are expected given the different evolutionary and demographic histories of the populations here studied. High diversity values were displayed by populations which historically have experienced important complex demographic events, such as the Hispanic groups [33,34] or the Asian and African populations analyzed [35,36]. On the other hand, lowest values were obtained for those populations, such as the Native American or Basque groups, that have been characterized as genetic isolates due to cultural or/and geographic barriers [37,38].

The capacity of the 6-plex SM Y-STR panel to discriminate among male haplotypes differed among the analyzed populations. In the Latin American groups, i.e. the admixed groups from Nicaragua and Colombia, the novel panel allowed the differentiation of more than half of the haplotypes, obtaining DC values of 0.6212 and 0.5167, respectively. Most of the haplotypes were distinguished by at least three locus differences (65.22% in Nicaragua and 57.46% in Colombia) (Fig. 1). Similar results were obtained for the African (DC = 0.6129) and Asian (DC = 0.4902) groups, where around half of the haplotypes could be discriminated using this multiplex and most of haplotypes exhibited two or three differences among them. In the Native American population from Guatemala, the DC was more limited given that only 19 haplotypes were observed in the 50 individuals analyzed (DC = 0.3800), 12 of these haplotypes were unique. In this population, around half of the individuals possessed haplotypes differing in two or three loci (51.84%). The Caucasian populations showed also low DC values, in particular the autochthonous Basque group (DC = 0.220) compared to the general Spanish population (DC = 0.3653). In this last group 80 out of 219

individuals could be differentiated using this 6-loci multiplex, with 59 unique haplotypes. In this case, the distribution of haplotypes, with 62.34% of them differing in two or more loci, is indicative of an influx of different male lineages in the gene pool. On the other hand, in the autochthonous Basque group, only 22 different haplotypes (13 unique ones) were identified among the 100 individuals. In fact, most of the haplotypes in this sample set displayed only one (34.55%) or two (21.62%) locus differences among them. This result is in agreement with previous reports that also indicated a limited influx of male lineages from other regions in the autochthonous group [38].

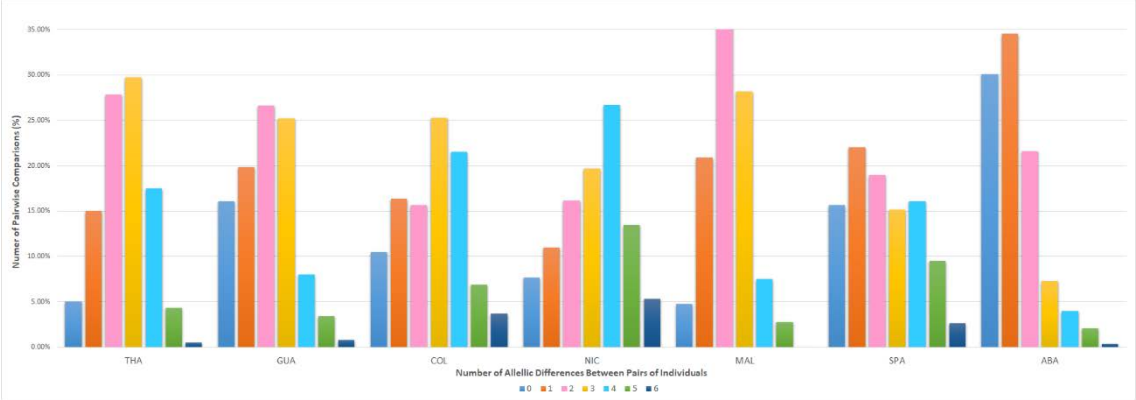


Fig. 1. Distribution of the number of locus differences between pairs of individuals in the analyzed population sets. Populations are: THA: Asians from Thailand; COL: Hispanics from Colombia; GUA: Native Americans from Guatemala; NIC: Hispanics from Nicaragua; MW: Africans from Malawi; SPA: European Caucasians from Spain; and ABA: European Caucasians Autochthonous Basques.

The 6-plex SM Y-STR panel provided a moderate power of discrimination between male haplotypes in most populations, despite of the low mutation rate. Yet, the inclusion of SM Y-STR markers in casework may be a valuable tool in exclusion of kinship cases where minimal discrepancies have been found using the routine panels, and when de novo mutations may account for the allele inconsistencies. The presence of one or more discrepancies in the SM Y-STRs may offer further evidence for the genuine exclusion of the biological parenthood, since mutation events are rarer to occur in these markers among close relatives.

3.3 Population-specific and shared Y-STR haplotypes across populations

Population-specific SM Y-STR haplotypes (only found in a single population) were mostly identified in the Asian, African, and Native American groups, with values of 80.00%, 78.95%, and 63.16%, respectively. On the other hand, in the groups from Spain and Latin America a higher number of Y-STR haplotypes in common is observed. This is expected given the relatively common European ancestral genetic pool of these populations.

Overall, among the shared SM Y-STR haplotypes, it is noticeable that the haplotype 12-12-11-15-10-10 (DYS388-DYS426-DYS461-DYS485-DYS525-DYS561) was the most abundant in the general population from Spain (39.27%) and autochthonous Basques (54%), as well as in the Latin American groups from Colombia (31.67%) and Nicaragua (27.27%). In contrast, no individuals with this haplotype were found in the Asian population from Thailand and neither in the African population from Malawi, and only one male carried this haplotype in the Native American population from Guatemala. The Y-SNP typing revealed that individuals carrying this haplotype belonged to the R1b haplogroup. In the Native American population from Guatemala the most common haplotype, assigned to the Q haplogroup according to Y-SNPs, was shared by 19 individuals (12-12-11-14-9-10), and it was also present in one Nicaraguan individual. In the case of the African population, the most frequent haplotype was detected in 5 males (12-11-12-14-10-11), and it was also seen in one Colombian male. This haplotype was associated to the E haplogroup according to Y-SNP typing. Finally, in the Asian group the most recurrent haplotype, which corresponded to the O haplogroup, was observed in 19 individuals (12-11-10-15-9-10) and it was specific from this population.

3.4 Y-STR population comparison

The pairwise R_{ST} genetic distances indicate that the Africans from Malawi, Asians from Thailand and Native Americans from Guatemala exhibit significant differences from all groups ($p \leq 0.0018$) (Supplementary Table S6 and Fig. 2A). On the other hand, the two Hispanic groups from Nicaragua and Colombia did not show significant differences between them, neither from general Spanish population. Therefore, the data from this novel multiplex system demonstrates the permanence of male Spanish genetic ancestry in the Hispanic groups sampled. Significant differences were observed from all the other groups in the case of the autochthonous Basque group, despite low R_{ST} values were observed from general Spanish population.

The Non-Metric Multi-Dimensional-Scaling plot (NMDS), based on pairwise R_{ST} genetic distances, shows the segregation among the groups from different continents (Fig. 2B). The populations from Africa, Asia and Native Americans are found apart from each other and the rest of the populations. On the other hand, the Spanish population plots close to the admixed Latin-American groups, and further apart from the autochthonous Basque group.

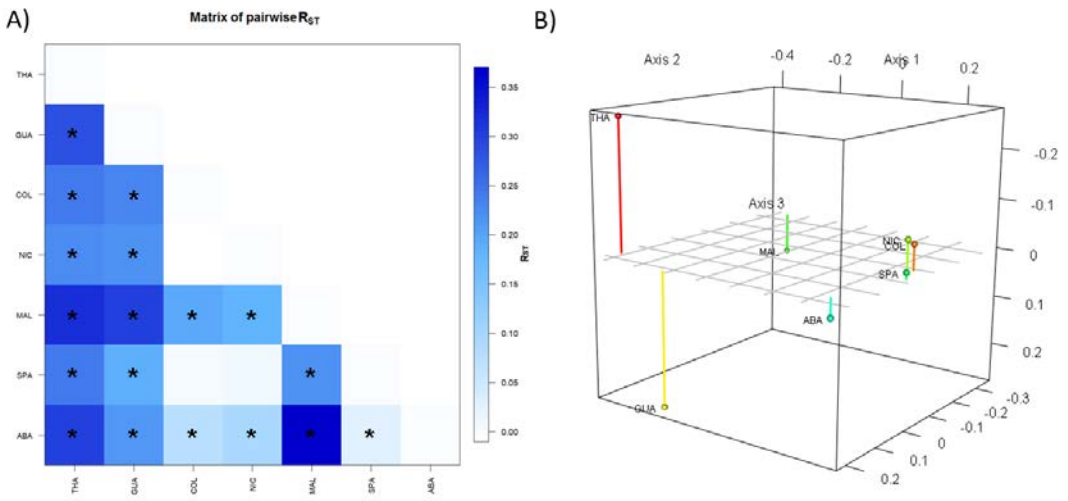


Fig. 2. A) Heat plot of pairwise RST values between all populations (blue=high values; white=low values). *Indicates significant difference after Bonferroni correction (P = 0.0009). B) Non-metric Multi-Dimensional-Scaling (NMDS) (3D projection with minimum stress of 0.0723) representation of genetic distances based on RST estimates. Populations are: THA: Asians from Thailand; COL: Hispanics from Colombia; GUA: Native Americans from Guatemala; NIC: Hispanics from Nicaragua; MW: Africans from Malawi; SPA: European Caucasians from Spain; and ABA: European Caucasians Autochthonous Basques. gr2ce Analysis of SM Y-STRs haplotypes in a three-Dimensional plot colored by haplogroups.

3.5 Congruency between SM Y-STR haplotypes and Y-SNP haplogroups across populations

A Factorial Correspondence Analysis 3D plot was constructed including the most abundant haplogroups identified in the study (in decreasing frequency: R, O, E, Q, J, I, G, and T) to visualize the correspondence of the 6 SM Y-STR haplotypes and Y-SNP haplogroups across populations (Fig. S4). Moderate clustering was observed for the different haplogroups along the axes, being 74.49% of the variance explained by the first three dimensions.

A more detailed analysis indicated that most individuals which displayed an identical SM Y-STR haplotype belonged to the same Y-SNP haplogroup (82.67%) (Supplementary Table S3). Different level of Y-STR haplotype diversification was detected within the haplogroups. High SM Y-STR haplotype resemblance was observed within haplogroups such as R, particularly in R-M269, observing mostly haplotypes identical (zero differences) to the most abundant haplotype or near identical (one mutation-step difference). These results are in accordance to previous studies that reported haplotype similarity within R-M269 [11,39]. Other haplogroups, such as E, J, and I displayed higher diversification, which could be due to a stronger differentiation at the subhaplogroup level [40,41].

Therefore, these results do not necessarily point to the capacity of unambiguous assignation of individuals in haplogroups using SM Y-STRs, as it has previously been stated for other Y-STRs

markers [30,42]. However, it shows that inclusion of these SM Y-STR loci in the analyses, together with conventional Y-STRs, may help to optimize the phylogenetic signals over current Y-STR panels.

4. Conclusions

In light of the findings in this study the novel set of Slowly Mutating Y-STRs (DYS388, DYS426, DYS461, DYS485, DYS525, and DY561) has demonstrated to be a reproducible, sensitive and robust multiplex system. This panel may be used in conjunction with the existing commercial multiplexes for forensic casework, particularly for confirming the exclusions in kinship cases where minimal discrepancies in one or few loci are reported using the regularly employed panels. The assessment of additional disparities in the SM Y-STRs may provide further evidence for the genuine exclusion of the biological kinship, since mutation events are rarer to occur in these markers. Furthermore, SM Y-STR data can be used to optimize and to increase the resolution of the phylogenetic trees based only on the current Y-STR panel sets. In addition, the results obtained in this study highlight the potential of combining mixed systems (slowly and rapidly mutating Y-STRs) to address different evolutionary time windows. In this study, we have provided an extensive Y-STR allele and haplotype reference dataset for future applications.

Conflict of interest statement

The authors have declared no conflict of interest.

Acknowledgements

Funds were provided by the Basque Government (Grupo Consolidado IT833-13). The authors are grateful to the Spanish National DNA Bank for Barcelona and Madrid sample collection and the Basque Foundation for Health Research and Innovation (BIOEF). We gratefully acknowledge PhD Maite Alvarez for her technical and human support provided by the DNA Bank Service (SGIker) of the University of the Basque Country (UPV/EHU) and European funding (ERDF and ESF).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.fsigen.2018.03.008.

References

[1] M.F. Hammer, A.J. Redd, Forensic applications of Y chromosome STRs and SNPs, Forensics Law Enforc. 133 (2006).

[2] T.E. King, M.A. Jobling, What's in a name? Y chromosomes, surnames and the genetic genealogy revolution, Trends Genet. 25 (2009) 351–360.

[3] G.D. Poznik, Y. Xue, F.L. Mendez, T.F. Willems, A. Massaia, M.A. Wilson Sayres, Q. Ayub, S.A. McCarthy, A. Narechania, S. Kashin, Y. Chen, R. Banerjee, J.L. Rodriguez-Flores, M. Cerezo, H. Shao, M. Gymrek, A. Malhotra, S. Louzada, R. Desalle, G.R.S. Ritchie, E. Cerveira, T.W. Fitzgerald, E. Garrison, A. Marcketta, D. Mittelman, M. Romanovitch, C. Zhang, X. Zheng-Bradley, G.R. Abecasis, S.A. McCarroll, P. Flicek, P.A. Underhill, L. Coin, D.R. Zerbino, F. Yang, C. Lee, L. Clarke, A. Auton, Y. Erlich, R.E. Handsaker, C.D. Bustamante, C. Tyler-Smith, Punctuated bursts in human male demography inferred from 1,244 worldwide Y chromosome sequences, Nat. Genet. 12 (2016) 809.

[4] P.A. Underhill, P. Shen, A. Lin, L. Jin, G. Passarino, W.H. Yang, E. Kauffman, B. Bonné-Tamir, J. Bertranpetit, P. Francalacci, M. Ibrahim, T. Jenkins, J.R. Kidd, S.Q. Mehdi, M.T. Seielstad, R.S. Wells, A. Piazza, R.W. Davis, M.W. Feldman, L.L. Cavalli-Sforza, P.J. Oefner, Y chromosome sequence variation and the history of human populations, Nat. Genet. 26 (2000) 358–361.

[5] O. Semino, G. Passarino, P.J. Oefner, A.A. Lin, S. Arbuzova, L.E. Beckman, G. De Benedictis, P. Francalacci, A. Kouvatsi, S. Limborska, M. Marcikiae, A. Mika, B. Mika, D. Primorac, A.S. Santachiara-Benerecetti, L.L. Cavalli-Sforza, P.A. Underhill, The genetic legacy of paleolithic Homo sapiens in extant europeans: a Y chromosome perspective, Science (2000) 1155–1159.

[6] D. Contu, L. Morelli, F. Santoni, J.W. Foster, P. Francalacci, F. Cucca, Y-chromosome based evidence for pre-neolithic origin of the genetically homogeneous but diverse Sardinian population: inference for association scans, PLoS One. 3 (2008) e1430.

[7] I.L. Rozhanskii, Mutation rate constants in DNA genealogy (Y chromosome), Adv. Anthropol. 1 (2011) 26–34.

[8] K.N. Ballantyne, V. Keerl, A. Wollstein, Y. Choi, S.B. Zuniga, A. Ralf, M. Vermeulen, P. De Knijff, M. Kayser, A new future of forensic Y-chromosome analysis: rapidly mutating Y-STRs for differentiating male relatives and paternal lineages, Forensic Sci. Int. Genet. 6 (2012) 208–218.

[9] K.N. Ballantyne, A. Ralf, R. Aboukhalid, N.M. Achakzai, M.J. Anjos, Q. Ayub, J. Balažic, J. Ballantyne, D.J. Ballard, B. Berger, C. Bobillo, M. Bouabdellah, H. Burri, T. Capal, S. Caratti, J. Cárdenas, F. Cartault, E.F. Carvalho, M. Carvalho, B. Cheng, M.D. Coble, D. Comas, D. Corach, M.E. D'Amato, S. Davison, P. de Knijff, M.C.A. De Ungria, R. Decorte, T. Dobosz, B.M. Dupuy, S. Elmrghni, M. Gliwiński, S.C. Gomes, L. Grol, C. Haas, E. Hanson, J. Henke, L. Henke, F. Herrera-Rodríguez, C.R. Hill, G. Holmlund, K. Honda, U.D. Immel, S. Inokuchi, M.A. Jobling, M. Kaddura, J.S. Kim, S.H. Kim, W. Kim, T.E. King, E. Klausriegler, D. Kling, L. Kovačević, L. Kovatsi, P. Krajewski, S. Kravchenko,

M.H.D. Larmuseau, E.Y. Lee, R. Lessig, L.A. Livshits, D. Marjanović, M. Minarik, N. Mizuno, H. Moreira, N. Morling, M. Mukherjee, P. Munier, J. Nagaraju, F. Neuhuber, S. Nie, P. Nilasitsataporn, T. Nishi, H.H. Oh, J. Olofsson, V. Onofri, J.U. Palo, H. Pamjav, W. Parson, M. Petlach, C. Phillips, R. Ploski, S.P.R. Prasad, D. Primorac, G.A. Purnomo, J. Purps, H. Rangel-Villalobos, K. Reogonekbała, B. Rerkamnuaychoke, D.R. Gonzalez, C. Robino, L. Roewer, A. Rosa, A. Sajantila, A. Sala, J.M. Salvador, P. Sanz, C. Schmitt, A.K. Sharma, D.A. Silva, K.J. Shin, T. Sijen, M. Sirker, D. Sváková, V. Škaro, C. Solano-Matamoros, L. Souto, V. Stenzl, H. Sudoyo, D. Syndercombe-Court, A. Tagliabracci, D. Taylor, A. Tillmar, I.S. Tsybovsky, C. Tyler-Smith, K.J. van der Gaag, D. Vanek, A. Völgyi, D. Ward, P. Willemse, E.P.H. Yap, R.Y.Y. Yong, I.Z. Pajnič, M. Kayser, Toward male individualization with rapidly mutating Y-Chromosomal short tandem repeats, Hum. Mutat. 35 (2014) 1021–1032.

[10] U. Rogalla, M. Woźniak, J. Swobodziński, M. Derenko, B.A. Malyarchuk, I. Dambueva, M. Koziński, J. Kubica, T. Grzybowski, A novel multiplex assay amplifying 13 Y-STRs characterized by rapid and moderate mutation rate, Forensic Sci. Int. Genet. 15 (2015) 49–55.

[11] M.H.D. Larmuseau, N. Vanderheyden, A. Van Geystelen, M. van Oven, P. de Knijff, R. Decorte, Recent radiation within Y-chromosomal haplogroup R-M269 resulted in high Y-STR haplotype resemblance, Ann. Hum. Genet. 78 (2014) 92–103.

[12] K.N. Ballantyne, M. Goedbloed, R. Fang, O. Schaap, O. Lao, A. Wollstein, Y. Choi, K. Van Duijn, M. Vermeulen, S. Brauer, R. Decorte, M. Poetsch, N. Von Wurmb-Schwark, P. De Knijff, D. Labuda, H. Vézina, H. Knoblauch, R. Lessig, L. Roewer, R. Ploski, T. Dobosz, L. Henke, J. Henke, M.R. Furtado, M. Kayser, Mutability of Y chromosomal microsatellites: rates, characteristics, molecular bases, and forensic implications, Am. J. Hum. Genet. 87 (2010) 341–353.

[13] J.M. Butler, R. Schoske, P.M. Vallone, M.C. Kline, A.J. Redd, M.F. Hammer, A novel multiplex for simultaneous amplification of 20 Y chromosome STR markers, Forensic Sci. Int. 129 (2002) 10–24.

[14] J.M. Butler, A.E. Decker, P.M. Vallone, M.C. Kline, Allele frequencies for 27 Y-STR loci with U.S. Caucasian African American, and Hispanic samples, Forensic Sci. Int. 156 (2006) 250–260.

[15] E.K. Hanson, J. Ballantyne, An ultra-high discrimination Y chromosome short tandem repeat multiplex DNA typing system, PLoS One 2 (2007) e688.

[16] P. Sánchez-Diz, L. Gusmão, S. Beleza, A. Benítez-Pérez, A. Castro, O. García, L.P. Solla, H. Geada, P. Martín, B. Martínez-Jarreta, M.D.F. Pinheiro, E. Raimondi, S.M. Silva De La Fuente, M.C. Vide, M.R. Whittle, M.T. Zarrabeitia, A. Carracedo, A. Amorim, Results of the GEP-ISFG

collaborative study on two Y-STRs tetraplexes: GEPY I (DYS461 GATA C4, DYS437 and DYS438) and GEPY II (DYS460, GATA A10, GATA H4 and DYS439), Forensic Sci. Int. 135 (2003) 158–162.

[17] S.-K. Lim, Y. Xue, E.J. Parkin, C. Tyler-Smith, Variation of 52 new Y-STR loci in the Y Chromosome Consortium worldwide panel of 76 diverse individuals, Int. J. Legal Med. 121 (2007) 124–127.

[18] M. Jacobs, L. Janssen, N. Vanderheyden, B. Bekaert, W. Van de Voorde, R. Decorte, Development and evaluation of multiplex Y-STR assays for application in molecular genealogy, Forensic Sci. Int. Genet. Suppl. Ser. 2 (2009) 57–59.

[19] A. Nebel, D. Filon, C. Hohoff, M. Faerman, B. Brinkmann, A. Oppenheim, Haplogroup-specific deviation from the stepwise mutation model at the microsatellite loci DYS388 and DYS392, Eur. J. Hum. Genet. 9 (2001) 22–26.

[20] O.J. Marshall, PerlPrimer: cross-platform, graphical primer design for standard, bisulphite and real-time PCR, Bioinformatics 20 (2004) 2471–2472.

[21] P.M. Vallone, J.M. Butler, AutoDimer: a screening tool for primer-dimer and hairpin structures, Biotechniques 37 (2004) 226–231.

[22] L. Valverde, M. Rosique, S. Köhnemann, S. Cardoso, A. García, A. Odriozola, J.M. Aznar, D. Celorrio, M. Schuerenkamp, J. Zubizarreta, M.C. Davis, G. Hampikian, H. Pfeiffer, M.M. De Pancorbo, Y-STR variation in the Basque diaspora in the Western USA: evolutionary and forensic perspectives, Int. J. Legal Med. 126 (2012) 293–298.

[23] P. Villaescusa, M. Illescas, L. Valverde, M. Baeta, C. Nuñez, B. Martínez-Jarreta, M. Zarrabeitia, F. Calafell, M.M. de Pancorbo, Characterization of the Iberian Y chromosome haplogroup R-DF27 in Northern Spain, Forensic Sci. Int. Genet. 27 (2016) 142–148.

[24] L. Excoffier, G. Laval, S. Schneider, Arlequin (version 3.0): an integrated software package for population genetics data analysis, Evol. Bioinform. 1 (2005) 47–50 (Online).

[25] W.R. Rice, Analyzing tables of statistical tests, Evolution 43 (1989) 223–225.

[26] Ø. Hammer, D.A.T. Harper, P.D. Ryan, PAST: paleontological statistics software package for education and data analysis, Palaeontol. Electron. 4 (2001) 1–9.

[27] D. Adler, D. Murdoch, rgl: 3D Visualization Device System (OpenGL) R Package, (2016).

[28] M. Van Oven, A. Van Geystelen, M. Kayser, R. Decorte, M.H. Larmuseau, Seeing the wood for the trees: a minimal reference phylogeny for the human Y chromosome, Hum. Mutat. 35 (2014) 187–191.

[29] L. Valverde, S. Köhnemann, S. Cardoso, H. Pfeiffer, M.M. De Pancorbo, Improving the analysis of Y-SNP haplogroups by a single highly informative 16 SNP multiplex PCR-minisequencing assay, Electrophoresis 34 (2013) 605–612.

[30] C. Nuñez, M. Geppert, M. Baeta, L. Roewer, B. Martínez-Jarreta, Y chromosome haplogroup diversity in a Mestizo population of Nicaragua, Forensic Sci. Int. Genet. 6 (2012) e192–5.

[31] L. Valverde, M.J. Illescas, P. Villaescusa, A.M. Gotor, A. García, S. Cardoso, J. Algorta, S. Catarino, K. Rouault, C. Férec, O. Hardiman, M. Zarrabeitia, S. Jiménez, M.F. Pinheiro, B.M. Jarreta, J. Olofsson, N. Morling, M.M. de Pancorbo, New clues to the evolutionary history of the main European paternal lineage M269: dissection of the Y-SNP S116 in Atlantic Europe and Iberia, Eur. J. Hum. Genet. 24 (2016) 437–441.

[32] K. Belkhir, P. Borsa, L. Chikhi, N. Raufaste, F. Bonhomme, GENETIX 4.05 logiciel sous Windows TM pour la génétique des populations, (2004).

[33] J.J. Builes, M.L. Bravo, C. Gómez, C. Espinal, D. Aguirre, A. Gómez, J. Rodríguez, P. Castañeda, A. Montoya, M. Moreno, A. Amorim, L. Gusmão, Y-chromosome STRs in an antioquian (Colombia) population sample, Forensic Sci. Int. 164 (2006) 79–86.

[34] C. Nuñez, M. Baeta, C. Sosa, Y. Casalod, J. Ge, B. Budowle, B. Martínez-Jarreta, Reconstructing the population history of Nicaragua by means of mtDNA, Y-chromosome STRs, and autosomal STR markers, Am. J. Phys. Anthropol. 143 (2010) 560–591.

[35] A.M. Oliveira, L. Gusmão, P.M. Schneider, I. Gomes, Detecting the paternal genetic diversity in west Africa using Y-STRs and Y-SNPs, Forensic Sci. Int. Genet. Suppl. Ser. 5 (2015) e213–e215.

[36] W. Kutanan, J. Kampuansai, S. Fuselli, S. Nakbunlung, M. Seielstad, G. Bertorelle, D. Kangwanpong, Genetic structure of the Mon-Khmer speaking groups and their affinity to the neighbouring Tai populations in Northern Thailand, BMC Genet. 12 (2011) 56.

[37] L. Roewer, M. Nothnagel, L. Gusmão, V. Gomes, M. González, D. Corach, A. Sala, E. Alechine, T. Palha, N. Santos, A. Ribeiro-dos-Santos, M. Geppert, S. Willuweit, M. Nagy, S. Zweynert, M. Baeta, C. Núñez, B. Martínez-Jarreta, F. González-Andrade, E. Fagundes de Carvalho, D.A. da Silva, J.J. Builes, D. Turbón, A.M. Lopez Parra, E. Arroyo-Pardo, U. Toscanini, L. Borjas, C. Barletta, E.

Ewart, S. Santos, M. Krawczak, Continent-wide decoupling of Y-Chromosomal genetic variation from language and geography in native South Americans, PLoS Genet. 9 (2013) e100346.

[38] L. Valverde, S. Köhnemann, M. Rosique, S. Cardoso, M. Zarrabeitia, H. Pfeiffer, M.M. De Pancorbo, 17 Y-STR haplotype data for a population sample of Residents in the Basque Country, Forensic Sci. Int. Genet. 6 (2012) e109–111.

[39] N. Sole-Morata, J. Bertranpetit, D. Comas, F. Calafell, Recent radiation of R-M269 and high Y-STR haplotype resemblance confirmed, Ann. Hum. Genet. 78 (2014) 253–254.

[40] S. Rootsi, C. Magri, T. Kivisild, G. Benuzzi, H. Help, M. Bermisheva, I. Kutuev, L. Barać, M. Pericić, O. Balanovsky, A. Pshenichnov, D. Dion, M. Grobei, L.A. Zhivotovsky, V. Battaglia, A. Achilli, N. Al-Zahery, J. Parik, R. King, C. Cinnioğlu, E. Khusnutdinova, P. Rudan, E. Balanovska, W. Scheffrahn, M. Simonescu, A. Brehm, R. Goncalves, A. Rosa, J.-P. Moisan, A. Chaventre, V. Ferak, S. Füredi, P.J. Oefner, P. Shen, L. Beckman, I. Mikerezi, R. Terzić, D. Primorac, A. Cambon-Thomsen, A. Krumina, A. Torroni, P.A. Underhill, A.S. Santachiara-Benerecetti, R. Villems, O. Semino, Phylogeography of Y-chromosome haplogroup I reveals distinct domains of prehistoric gene flow in europe, Am. J. Hum. Genet. 75 (2004) 128–137.

[41] O. Semino, C. Magri, G. Benuzzi, A.A. Lin, N. Al-Zahery, V. Battaglia, L. Maccioni, C. Triantaphyllidis, P. Shen, P.J. Oefner, L.A. Zhivotovsky, R. King, A. Torroni, L.L. Cavalli-Sforza, P.A. Underhill, A.S. Santachiara-Benerecetti, Origin, diffusion, and differentiation of Y-Chromosome haplogroups E and J: inferences on the neolithization of Europe and later migratory events in the mediterranean area, Am. J. Hum. Genet. 74 (2004) 1023–1034.

[42] M. Muzzio, V. Ramallo, J.M.B. Motti, M.R. Santos, J.S. López Camelo, G. Bailliet, Software for Y-haplogroup predictions: a word of caution, Int. J. Legal Med. 125 (2011) 143–147.

Electronic supplementary material

Supplementary Tables

Supplementary Table S1. Loci information, mutation rate, number of alleles observed, allele range size (in bp), PCR primers and final concentrations, and fluorescent dyes.

Locus	Repeat motif	Mutation rate[a] (per locus per generation)	No. of alleles observed	Allele range size (in bp)[b]	Primer Sequences (5'-3')[c]	Primer concentration (µM)	Dye (in F primer)
DYS388	$(\mathbf{ATT})_n$	4.25×10^{-4}	10-18	191-215	F: GTGAGTTAGCCGTTTAGCGA	0.2	5'-FAM
					R: TAGTCCCAGCTACTCAGCAG		
DYS426	$(\mathbf{GTT})_n$	3.98×10^{-4}	10-14	181-193	F: GAAGCTCAACTGTTTGTAATCTGG	0.2	5'-ATTO 550
					R: CTGGGTGACAAGACGAGAC		
DYS461 (Y-GATA-A7.2)	$(\mathbf{TAGA})_n$	9.89×10^{-4}	9-13	174-190	F: GCAGAGGATAGATGATATGGATA	0.2	5'-ATTO 565
					R: CAGGTAAATCTGTCCAGTAGTGA		
DYS485	$(\mathbf{TTT})_{0\text{-}1}(\mathbf{TTA})_n$	4.04×10^{-4}	12-18	216-234	F: ACTTCGCCACTACATAATATGTCC	0.2	5'-YAKYE
					R: AAGGCTGAGGCTTAAGAATCAC		
DYS525	$(\mathbf{AGAT})_n$	9.78×10^{-4}	8-13	206-226	F: GATAGGGAGATGATACATAGAAG	0.2	5'-ATTO 550
					R: CATCCATCTGTTTATCTTCCCA		
DYS561	$(\mathbf{GATA})_n(GACA)_4$	9.41×10^{-4}	8-12	207-223	F: TTAATGCTTGCCTGATGCCA	0.2	5'-ATTO 565
					R: AGTGATCTATGATCCCAACAACTC		

Supplementary Table S2. Y-SNP characteristics, primer sequences and analysis conditions.

Y-SNP	Primer Sequences (5'-3')[a]	Method[b]	Annealing Temperature (C°)	Amplicon Size (in bp)	Mutation (anc/der)	db SNP ID	Y chr position hg19	Reference
E-P170	F: TTGTTTCCTTTGGCAAACTG	HRM & Seq.	55	109	G/A	rs9786025	15,021,522	Valverde et al 2013
	R: GGCATTTCCACAAATACACTG							
G-M201	F: TGTCAAATTGTGACACTGCA	HRM & Seq.	57	149	G/T	rs2032636	15,027,529	Valverde et al 2013
	R: CTTCATCCAACACTAAGTACCT							
IJ-P126	F: CTCCAGACAAATCTCGTCTC	HRM & Seq.	56	90	C/G	rs17250163	21,225,770	Valverde et al 2013
	R: CCTTTACCAAGTAGTCACCTG							
I-M258	F: CAGGATTTGTCAAGGATGGG	HRM & Seq.	55	106	T/C	rs9341301	15,023,364	Valverde et al 2013
	R: GCTATGACTAAGAGGGATTCCA							
T-M272	F: ATTACTCTTTGCTCTCCCGA	HRM & Seq.	55	114	A/G	rs9341308	22,738,775	Valverde et al 2013
	R: CCCAGAATACACTTTATCCTCC							
O-M175	F: TTAAACTCTCTGAATCAGGCACAT	HRM & Seq.	56	70	TTCTC/Del	-	15,508,704..15,508,712	**Present study**
	R: TGATACCTTTGTTTCTGTTCATTC							
Q-M242	F: TCTACGGCATAGAAAGTTTGTG	HRM & Seq.	55	138	C/T	rs8179021	15,018,582	Valverde et al 2013
	R: CTAGAACAACTCTGAAGCGG							
Q1a2-M3	F: AGGGCATCTTTCATTTTAGG	HRM & Seq.	59,5	156	G/A	rs3894	19,096,363	**Present study**
	R: GTGGATTTGCTTTGTAGTAGG							
R-M207	F: GGGCAAATGTAAGTCAAGCA	HRM & Seq.	56	81	A/G	rs2032658	15,581,983	Valverde et al 2013
	R: CACTTCAACCTCTTGTTGGA							
R1b-M269	F: ACATGGTATCACAATAGAAGGG	HRM & Seq.	60.5	216	T/C	rs9786153	22,739,367	Valverde et al 2015
	R: TCCAAGGTGCTGGGATTAC							

[a] F: forward ; R: Reverse

[b] Detection method: Seq (Sequencing) and HRM (High Resolution Melting)

Supplementary Table S3. Y-STR haplotypes from the studied population groups. N: number of individuals. [1] Núñez et al. 2012; [2] Valverde et al. 2015; [3] Villaescusa et al. 2017.

Corresponds to Attached Table 3 in Appendix section.

Supplementary Table S4. Allele frequencies and gene diversity (GD) for each of the 6 Y-STR markers in the studied population samples. N: number of individuals.

Population	Asians from Thailand	Native Americans from Guatemala	Hispanics from Colombia	Hispanics from Nicaragua	Africans from Malawi	European Caucasians from Spain	European Caucasians Autochthonous Basques
N	102	50	60	66	31	219	100
DYS388							
10	0.1275 ± 0.0332		0.0167 ± 0.0167		0.0323 ± 0.0323		
11	0.0098 ± 0.0098	0.0400 ± 0.0280					0.0100 ± 0.0100
12	0.7745 ± 0.0416	0.8400 ± 0.0524	0.8167 ± 0.0504	0.7576 ± 0.0532	0.9355 ± 0.0449	0.8402 ± 0.0248	0.9400 ± 0.0239
13	0.0490 ± 0.0215	0.1000 ± 0.0429	0.0667 ± 0.0325	0.0758 ± 0.0328	0.0323 ± 0.0323	0.0639 ± 0.0166	0.0200 ± 0.0141
14	0.0196 ± 0.0138			0.0455 ± 0.0258		0.0365 ± 0.0127	0.0200 ± 0.0141
15			0.0333 ± 0.0234	0.0455 ± 0.0258		0.0365 ± 0.0127	
16	0.0196 ± 0.0138	0.0200 ± 0.0200	0.0167 ± 0.0167	0.0606 ± 0.0296		0.0228 ± 0.0101	0.0100 ± 0.0100
17			0.0333 ± 0.0234	0.0152 ± 0.0152			
18			0.0167 ± 0.0167				
GD	0.3844	0.2882	0.3311	0.4187	0,1269	0,2881	0.1166
DYS426							
10						0.0046 ± 0.0046	
11	0.9412 ± 0.0234	0.0400 ± 0.0280	0.3500 ± 0.0621	0.4242 ± 0.0613	1.0000 ± 0.0000	0.2740 ± 0.0302	0.0500 ± 0.0219
12	0.0588 ± 0.0234	0.9600 ± 0.0280	0.6500 ± 0.0621	0.5758 ± 0.0613		0.7032 ± 0.0309	0.9400 ± 0.0239
13						0.0137 ± 0.0079	
14						0.0046 ± 0.0046	0.0100 ± 0.0100
GD	0.1118	0.0784	0.4627	0.4960	0,0000	0,4322	0.1150
DYS461							
9	0.0588 ± 0.0234	0.2000 ± 0.2000		0.0303 ± 0.0213		0.0320 ± 0.0119	
10	0.4118 ± 0.0490	0.2000 ± 0.2000	0.2000 ± 0.0521	0.1364 ± 0.0426	0.0645 ± 0.0449	0.1096 ± 0.0212	0.1500 ± 0.0359
11	0.4118 ± 0.0490	0.8000 ± 0.0571	0.6667 ± 0.0614	0.6515 ± 0.0591	0.2581 ± 0.0799	0.7260 ± 0.0302	0.7600 ± 0.0429
12	0.0784 ± 0.0268	0.1400 ± 0.0496	0.1333 ± 0.0443	0.1667 ± 0.0462	0.5484 ± 0.0909	0.1005 ± 0.0204	0.0900 ± 0.0288
13	0.0392 ± 0.0193	0.2000 ± 0.2000		0.0152 ± 0.0152	0.1290 ± 0.0612	0.0320 ± 0.0119	
GD	0.6562	0.3461	0.5062	0.5361	0,6323	0,4508	0.3958
DYS485							
12		0.0600 ± 0.0339		0.0303 ± 0.0213		0.0320 ± 0.0119	0.0200 ± 0.0141
13	0.0098 ± 0.0098	0.0400 ± 0.0280		0.0303 ± 0.0213		0.0183 ± 0.0091	0.0500 ± 0.0219
14	0.0294 ± 0.0168	0.6000 ± 0.0700	0.2000 ± 0.0521	0.1667 ± 0.0462	0.8065 ± 0.0721	0.0868 ± 0.0191	0.0800 ± 0.0273
15	0.6765 ± 0.0466	0.2000 ± 0.0571	0.6667 ± 0.0614	0.5303 ± 0.0619	0.0645 ± 0.0449	0.7717 ± 0.0284	0.8200 ± 0.0386
16	0.2451 ± 0.0428	0.1000 ± 0.0429	0.0833 ± 0.0360	0.1212 ± 0.0405	0.0968 ± 0.0540	0.0548 ± 0.0154	0.0200 ± 0.0141
17	0.0392 ± 0.0193		0.0500 ± 0.0284	0.1212 ± 0.0405		0.0320 ± 0.0119	0.0100 ± 0.0100
18					0.0323 ± 0.0323	0.0046 ± 0.0046	
GD	0,4846	0.5967	0.5147	0.6699	0,3462	0,3934	0.3210

DYS525							
8	0.0294 ± 0.0168					0.0091 ± 0.0064	
9	0.7549 ± 0.0428	0.7400 ± 0.0627		0.0606 ± 0.0296	0.0968 ± 0.0540	0.0320 ± 0.0119	0.0200 ± 0.0141
10	0.1863 ± 0.0387	0.2200 ± 0.0592	0.7333 ± 0.0576	0.6818 ± 0.0578	0.7097 ± 0.0829	0.7854 ± 0.0278	0.9500 ± 0.0219
11	0.0098 ± 0.0098	0.0400 ± 0.0280	0.1833 ± 0.0504	0.1818 ± 0.0478	0.1935 ± 0.0721	0.1461 ± 0.0239	0.0300 ± 0.0171
12	0.0098 ± 0.0098		0.0667 ± 0.0325	0.0606 ± 0.0296		0.0228 ± 0.0101	
13	0.0098 ± 0.0098		0.0167 ± 0.0167			0.0046 ± 0.0046	
GD	0.3982	0,4106	0.4311	0.5021	0,4645	0,3618	0.0972
DYS561							
8						0.0046 ± 0.0046	
9	0.0392 ± 0.0193	0.0200 ± 0.0200	0.1167 ± 0.0418	0.1212 ± 0.0405	0.1290 ± 0.0612	0.1096 ± 0.0212	0.0700 ± 0.0256
10	0.6275 ± 0.0481	0.8200 ± 0.0549	0.7500 ± 0.0564	0.7273 ± 0.0552	0.4839 ± 0.0912	0.7534 ± 0.0292	0.8700 ± 0.0338
11	0.3137 ± 0.0462	0.1600 ± 0.0524	0.1000 ± 0.0391	0.1515 ± 0.0445	0.3548 ± 0.0874	0.1324 ± 0.0230	0.0600 ± 0.0239
12	0.0196 ± 0.0138		0.0333 ± 0.0234		0.0323 ± 0.0323		
GD	0.5110	0.3078	0.4198	0.4401	0,6430	0,4046	0.2370

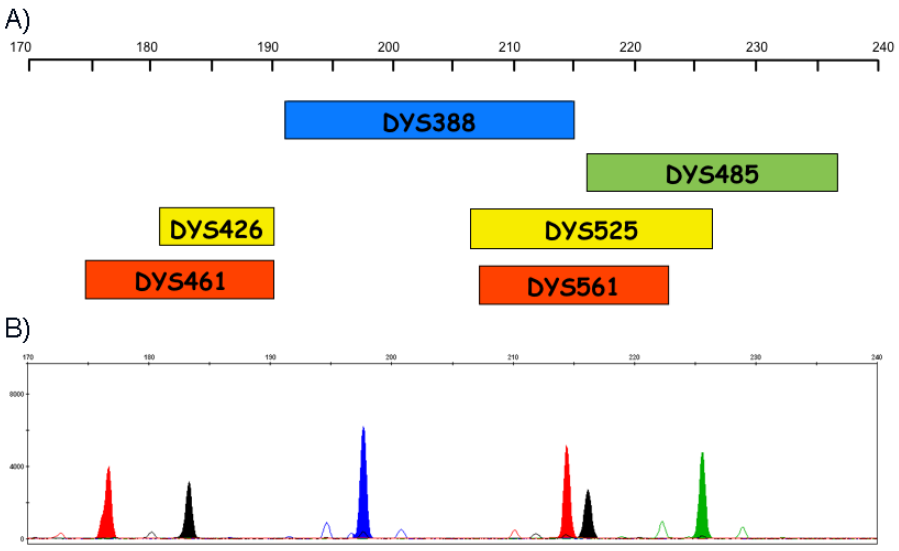
Supplementary Table S5. Diversity parameters obtained for the populations analyzed with the six Y-STR multiplex.

Population	N	Different haplotypes	Unique haploypes	Population specific haplotypes	HD	DC
Asians from Thailand	102	50	34	40	0.9493 ± 0.0132	0,4902
Native Americans from Guatemala	50	19	12	12	0.8392 ± 0.0463	0,3800
Hispanics from Colombia	60	31	22	11	0.8949 ± 0.0353	0,5167
Hispanics from Nicaragua	66	41	34	16	0.9235 ± 0.0281	0,6212
Africans from Malawi	31	19	13	15	0.9527 ± 0.0213	0,6129
European Caucasians from Spain	219	80	59	42	0.8437 ± 0.0246	0,3653
European CaucasiansAutochthonous Basques	100	22	13	7	0.6990 ± 0.0505	0,2200

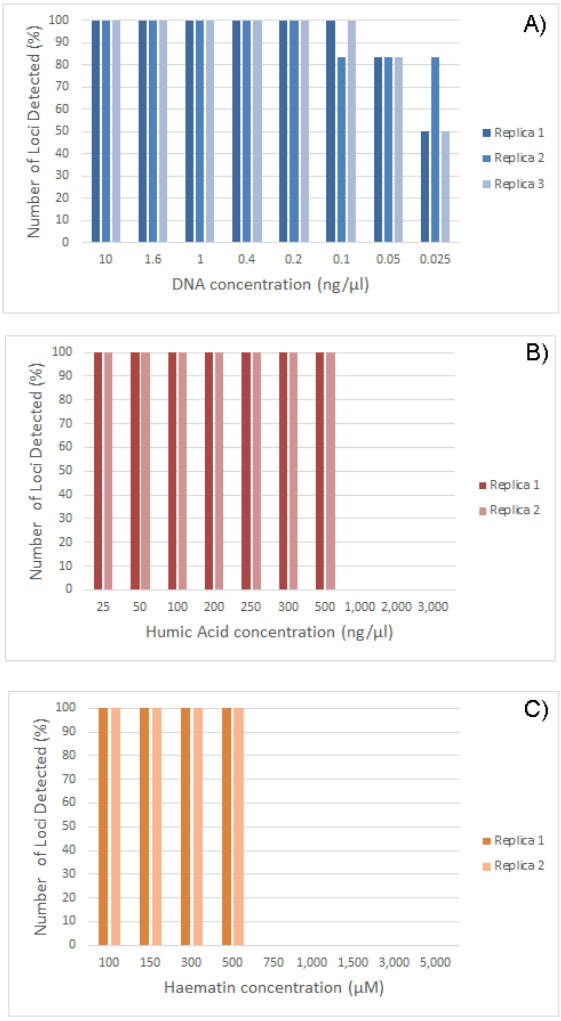
Supplementary Table S6. Pairwise RST value estimates (under the diagonal) and their significance (above the diagonal) between all populations, based on the 6 Y-STR loci studied. Significant values (p<0.0018 after Bonferroni correction) are highlighted in bold italics.

	Asians from Thailand	Native Americans from Guatemala	Hispanics from Colombia	Hispanics from Nicaragua	Africans from Malawi	European Caucasians from Spain	European Caucasians Autochthonous Basques
Asians from Thailand	*	0.0000±0.0000	0.0000±0.0000	0.0000±0.0000	0.0000±0.0000	0.0000±0.0000	0.0000±0.0000
Native Americans from Guatemala	***0,2856***	*	0.0000±0.0000	0.0000±0.0000	0.0000±0.0000	0.0000±0.0000	0.0000±0.0000
Hispanics from Colombia	***0,2418***	***0,2305***	*	0.8677±0.0039	0.0000±0.0000	0.2784±0.0037	0.0000±0.0000
Hispanics from Nicaragua	***0,2262***	***0,2188***	-0,0107	*	0.0000±0.0000	0.0529±0.0020	0.0000±0.0000
Africans from Malawi	***0,3183***	***0,2988***	***0,1966***	***0,1806***	*	0.0000±0.0000	0.0000±0.0000
European Caucasians from Spain	***0,2409***	***0,1897***	0,0025	0,0126	***0,2159***	*	0.00109±0.0003
European Caucasians Autochthonous Basques	***0,3030***	***0,2123***	***0,0730***	***0,0936***	***0,3698***	***0,0250***	*

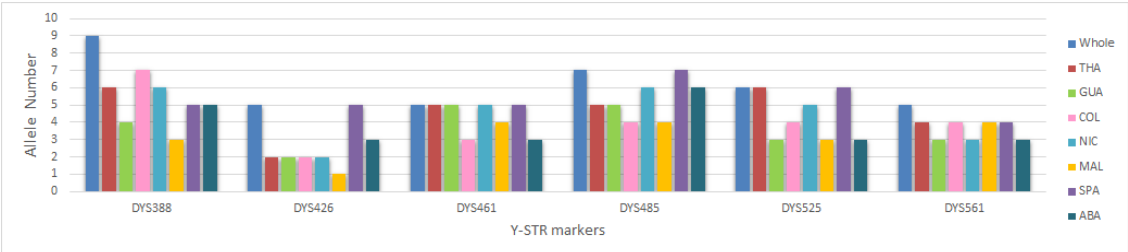
Supplementary Figures



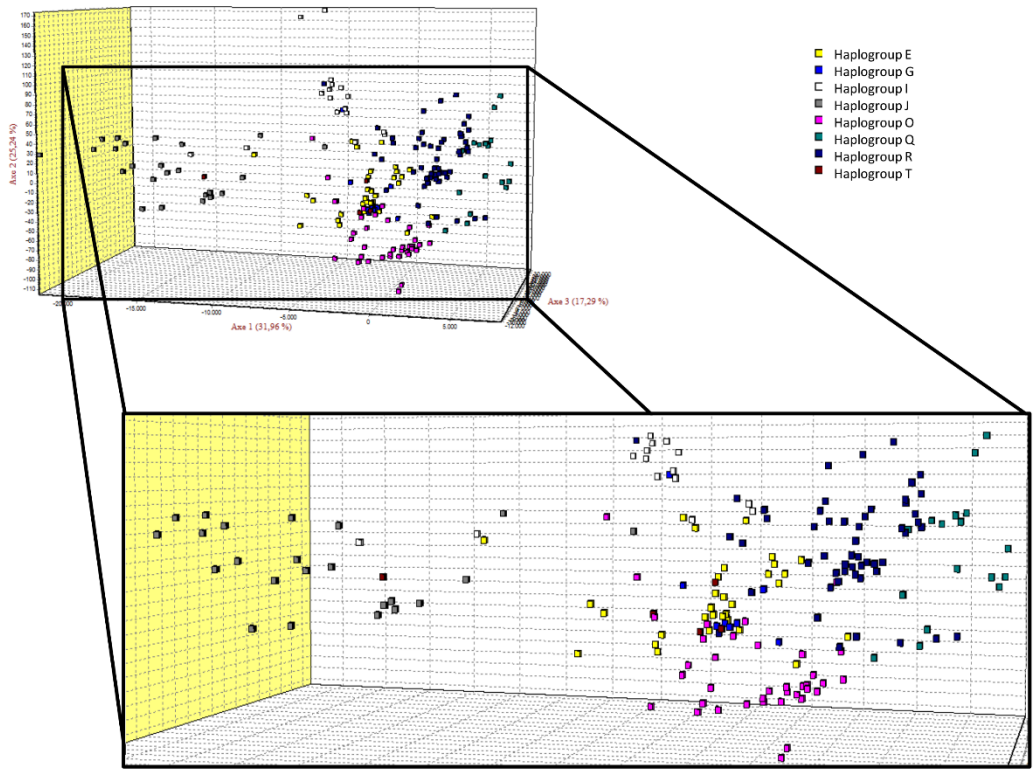
Supplementary Figure S1. A) Diagram of the panel developed in the present study. (B) A representative electrophoregram showing the profile of 1.5 ng control DNA amplified at the optimized PCR conditions. The peaks correspond to: DYS388 (blue), DYS485 (green), DYS426 (black), DYS525 (black), DYS461 (red), and DYS561 (red). The GeneMapper ID-X plots are presented as combined all dyes.



Supplementary Figure S2. Number of alleles determined in the six SM Y-STRs for each population and the whole dataset. Populations are: THA: Asians from Thailand; COL: Hispanics from Colombia; GUA: Native Americans from Guatemala; NIC: Hispanics from Nicaragua; MW: Africans from Malawi; SPA: European Caucasians from Spain; and ABA: European Caucasians Autochthonous Basques.



Supplementary Figure S3. Performance testing of the 6 SM Y-STRs multiplex measured as average (%) of loci detected. A) Sensitivity test of template DNA (2800M DNA) ranging from 10 ng to 25 pg. B) Inhibitory effects of humic acid ranging from 25 ng to 3000 ng. C) Inhibitory effects of haematin ranging from 100 µM to 5000 µM.



Supplementary Figure S4. Factorial Correspondence Analysis of SM Y-STRs haplotypes in a three-Dimensional plot colored by haplogroups.

5. Discussion

5.1 The paternal genetic landscape of Southwestern Europe

The present doctoral thesis work has focused on the study of the paternal lineage R1b-M269 through the dissection in its sublineages by the analysis of Y chromosome SNPs (Y-SNPs) in order to reconstruct the most probable evolutionary scenario of its origin, which has provided valuable results of forensic and population interest. In addition to that, with the purpose of responding to the demand of more multiplex tools of forensic application, two novel panels of Y-SNPs and Y-STRs were developed that have enabled, respectively, to attain higher haplogroup resolution of the branch M269 and to resolve particular cases that conventional Y-STR panels are not able to.

5.1.1 Haplogroup composition of Southwestern Europe

The paternal genetic landscape of Europe is defined by haplogroup R, to which more than 50% of the men belong to. R lineage is mainly subdivided in the sub-branches R1a-M420 and R1b-M343, which are more common in East Europe and West Europe respectively (Figure 17). However, the genetic landscape of Southwestern Europe is also defined by other haplogroups apart from R, like E, G, I, J, and N [258,344].

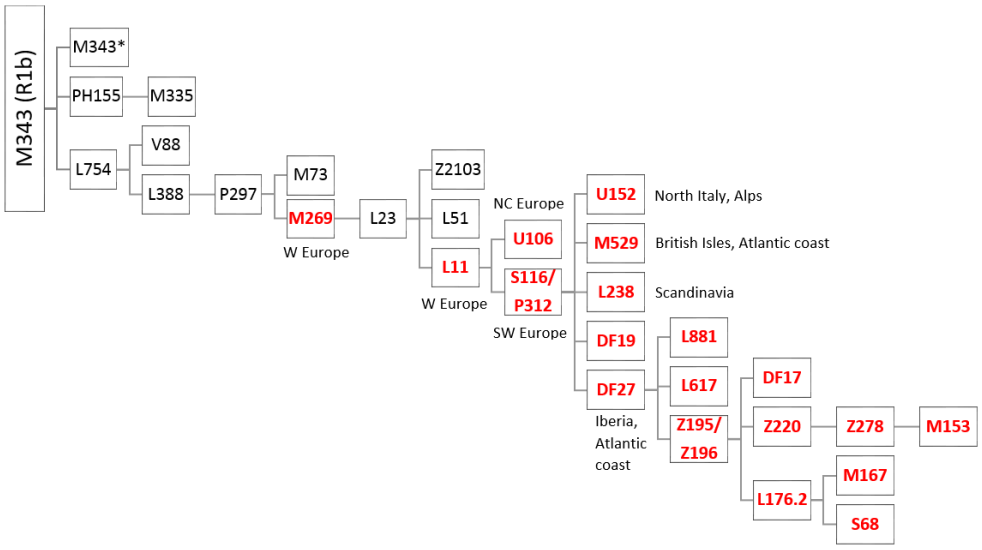


Figure 17. Simplified phylogenetic tree of the R1b-M343 haplogroup and geographic location of the main subhaplogroups, if known. In red bold letters are represented the Y-SNPs analyzed in the present doctoral thesis work.

In the present doctoral thesis work, we studied the most common West European paternal lineage, the R1b derived branch R1b1a1b-M269, and its sublineages. Even though the distribution and structure of the other less frequent European haplogroups were already defined, in the case of M269, its structure and how it expanded was still not completely known [257]. For that reason, the information provided by the present thesis is of high value, as it has allowed us to refine the

distribution of M269 and its subhaplogroups in populations from Southwestern Europe (*Study Number 1*). Understanding the distribution of this lineage and its sublineages in Europe, as well as in other regions of European influence, is of great interest not only in population genetics but also for forensic purposes. M269 can be used as a marker of West European ancestry, as its detection in a vestige would point to European paternal biogeographical origin and, thus, allow to connect a crime contributor to a more concrete geographical area. Furthermore, it could also allow to trace European migrations to other areas of the world.

5.1.2 Dissection and structure of M269

Previous studies by Balaresque and colleagues, and Myres and colleagues [257,258], revealed that most of the male individuals that currently inhabit Central and Western Europe belong to the paternal lineage R1b-M269, in frequencies between 40-90%. This haplogroup displays the highest frequencies in the Franco-Cantabrian area and shows a west-east decreasing frequency cline with distance (Figure 18). Our results in *Study Number 1* reveal frequencies for M269 concordant with the previous studies, and improves the coverage in populations from Southwestern Europe, especially from the Atlantic Coast.

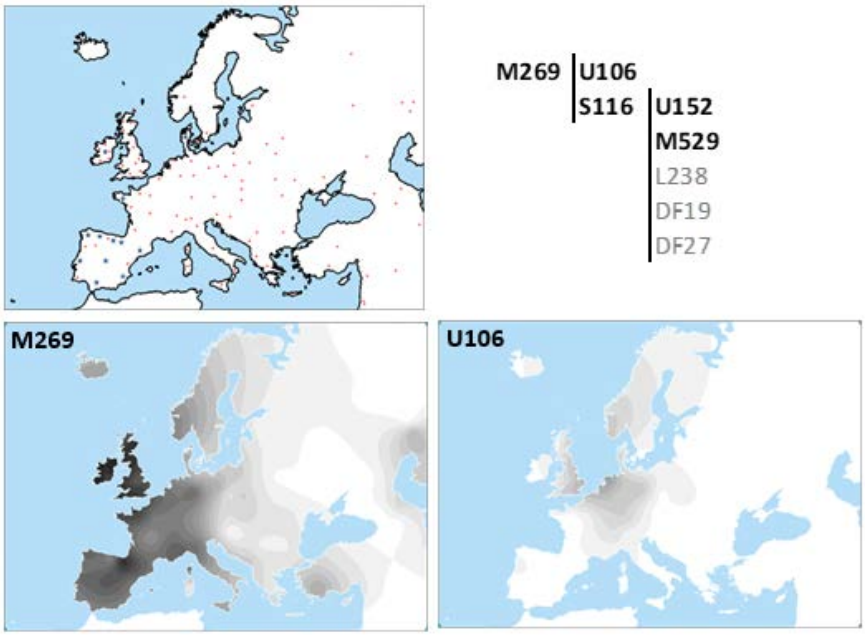


Figure 18. Frequency distribution maps of the data compiled in *Study Number 1* (blue stars) and the data from [258,259,345] (red points). The Y-SNPs used for the construction of these maps are highlighted in bold in the upper right tree.

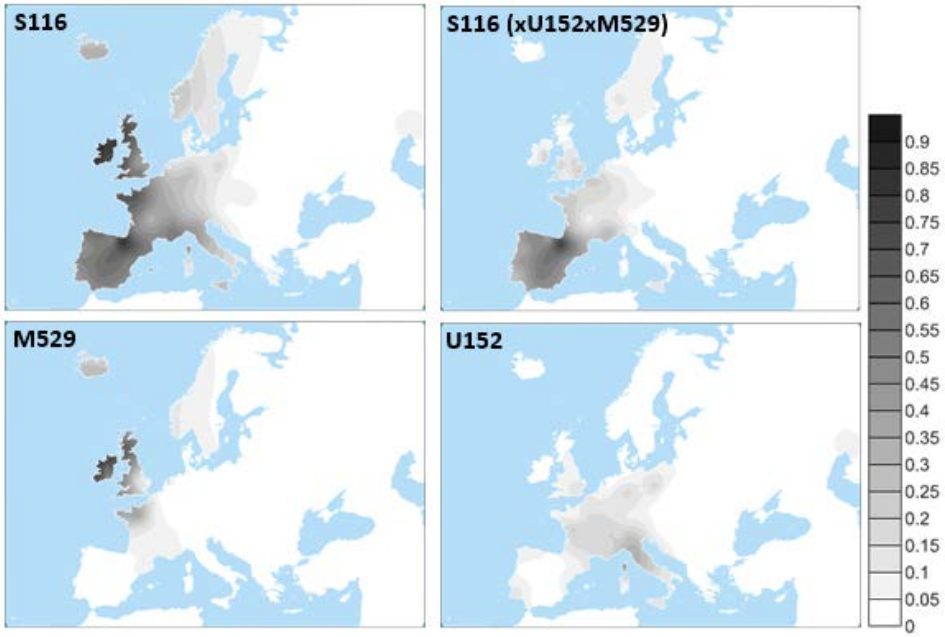


Figure 18. Continuation. Frequency distribution maps of the data compiled in *Study Number 1* (blue stars) and the data from [258,259,345] (red points). The Y-SNPs used for the construction of these maps are highlighted in bold in the upper right tree.

The two main M269 subhaplogroups, namely U106 and S116 (also known as P312), showed distinct areas of distribution in Europe: U106 is more frequent in Central and Northern Europe while S116 is the dominant subtype in Western and Southwestern Europe (Figures 18, 19). Surprisingly, the distribution of S116 found in our study differs from the one proposed by Myres and colleagues [258]. Myres and colleagues [258] detected a frequency peak for S116 in the Upper Danube Basin and Paris, with a declining frequency towards Italy, Southern France, the Iberian Peninsula, and the British Isles. In contrast, our data shows maximum frequencies in Northern Iberia, the French western coast, and the British Isles, which raised questions on the possible expansion of S116 during the early Neolithic LBK culture (Linearbandkeramik or Linear Pottery culture) as Myres and colleagues [258] proposed. These discrepancies in the frequency distributions with Myres and colleagues [258] could be due to the inclusion of new populations form the Atlantic coast and Iberia, with allowed a better coverage of those European areas. Furthermore, a more recent study has suggested that the dissemination of S116 throughout most of its present-day distribution may be linked to individuals of the Bell Beaker culture [346], a more recent culture linked to the early Bronze Age.

Regarding the sublineages of S116, as previously noted [258,259], U152 is more common in Northern Italy and the Alpine region, whereas M529 is more frequent in the British Isles and Brittany. Our results reveal a striking distribution of U152 in Iberia, displaying frequency peaks in the coastal corners in the Southwest (13%), Northwest (8%), and Northeast (6%). This pattern could be

explained by migration from the Alpine region, where this haplogroup peaks, to the Iberian Peninsula following a coastal route. As for M529, our new population data revealed high frequencies (>50%) of this Y-SNP in Brest (Brittany), outside the British Isles, which raises questions whether it originated in the British Isles (where it is most common nowadays) or somewhere else in the European continent (Figure 18).

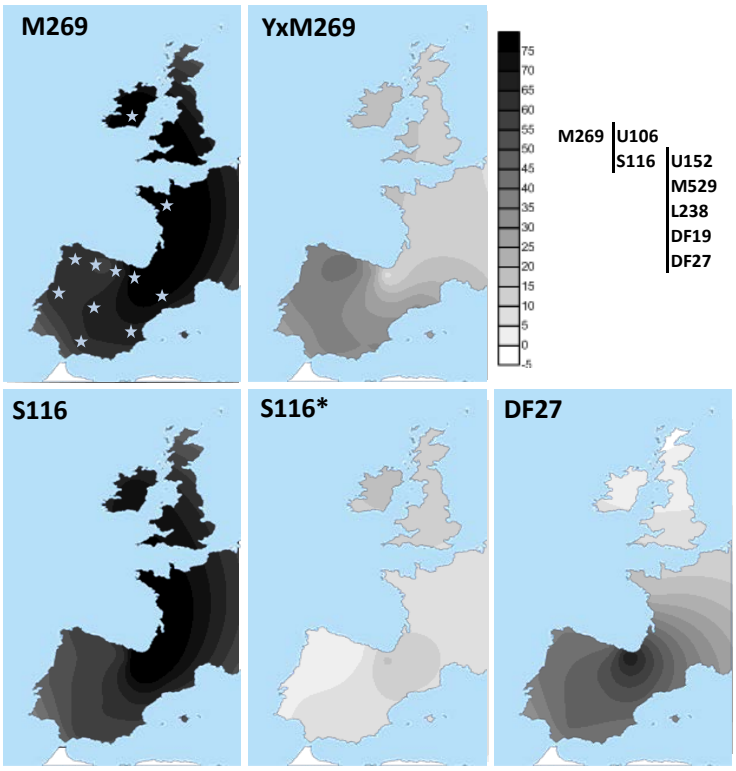


Figure 19. Frequency distribution maps of M269, S116, and DF27 in the Atlantic Coast and Iberian Peninsula obtained in *Study Number 1*. The stars in M269 map indicate the samples of population analysed. The upper right tree includes the Y-SNPs used for constructing the distribution maps.

Considering all of the discussed above, the dissection analysis of the M269 sublineage S116 provided informative results that allowed further completing the history of M269 lineage. Owing to the discovery of the highly frequent sublineage DF27, first described by Rocca and colleagues [224], the paragroup S116* (xU152, xM529) was largely resolved. Previous studies of S116 lineage and sublineages in populations from Southwestern Europe [258,259] found the highest frequencies of the paragroup S116* in Iberia, between 28-52%. *Study Number 1* confirms that most of those frequencies correspond to the lineage DF27. Furthermore, the median joining network analyses constructed from Y-STRs showed a bipartite structure corresponding to the individuals belonging to S116* and DF27 haplogroup (Figure 20). DF27 was found in frequencies between 40-48% in Iberia, reaching peak values in the Basque Country (>60%). Thus, the haplogroup DF27 occupies a different geographic area from that of the other S116 sublineages U152 and M529.

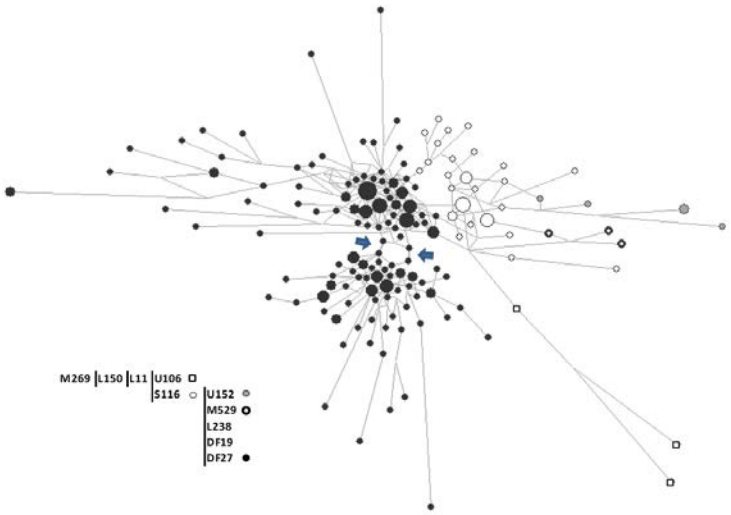


Figure 20. Median joining network of the M269 haplogroup in the Basque native population (bearing Basque surnames) obtained in *Study Number 1*. The blue arrows indicate a phylogenetic split of DF27 haplogroup into two groups bearing the alleles 14/18 and 15/19 in the Y-STR haplotype DYS437/DYS448.

5.1.3 The origin and controversy of M269

The origin and expansion of M269 has been the subject of heated debate due to the differences in the time to the most recent common ancestor (TMRCA) estimated by various authors [257,258,295]. The most widely accepted theories argued that M269 originated in the Franco-Cantabrian refuge, and that the current pattern of frequencies is the result of its postglacial expansion during the Paleolithic [294,295,347]. Other theories proposed its origin in Eastern Europe during the Neolithic, based on the higher diversity of Y-STR haplotypes in that area [257], or during the Mesolithic period considering coalescence times and frequency distribution [258].

The differences on the estimated TMRCA for M269 based on Y-STR markers is mainly due to the fact that estimates are sensitive to the calibration of the mutation rates and to the mathematical model applied to perform the inference [348]. In order to facilitate the discussion in the present thesis work, the mentioned TMRCAs were detailed in the same time scale (years ago, ya), although the authors used different ways to express it. Balaresque and colleagues [257] used a germinal mutation rate (GMR) and obtained TMRCAs for M269 between 5,500 and 8,000 ya, which would agree with a Neolithic expansion. Conversely, Morelli and colleagues [295] used the evolutionary mutation rate (EMR) and obtained much older TMRCAs, between 14,800-32,600 ya, supporting a Paleolithic origin. Two studies [258,349] obtained TMRCAs (8,590-11,950 ya and 8,500-12,500 ya, respectively) more compatible with a Neolithic expansion also employing the EMR. Overall, it is clear that age estimates based on Y-STR variation have proven to be a difficult topic since, apart from the mutation rate and the mathematical model, the set of Y-STR markers used for the

estimation and the individuals selected for the calculation are also critical. In *Study Number 1*, making the calculations with all the precautions reported so far, we obtained coalescent times selecting the EMR that dated the origin of S116 lineage 9,711-13,635 ya, which would place the origin of this lineage, and M269, in the Paleolithic.

Given the controversy in calculating TMRCAs, in *Study Number 1* we considered that time estimates should be taken with caution until more complete Y chromosome sequencing allowed more accurate time scales, and/or genotyped and reliably dated archaeological remains became available. In that context, since our TMRCA results were more compatible with an origin of M269 during the Paleolithic, an arrival of M269 to Europe during the Neolithic was regarded to be unlikely, since it would assume the creation of a complex scenario of expansions of sublineages in a relative short time. Moreover, the advent of M269 during the Neolithic would also assume a rapid expansion of the lineage throughout Europe, replacing most of the previously settled lineages.

Regarding the place of origin and/or expansion of M269, classical theories located it in the Franco-Cantabrian refuge, in the Paleolithic, due to its maximum frequencies and the pattern of decreasing frequency with increasing distance from that area. However, our dissection of M269 in the refuge area in *Study Number 1* raises questions about its origin in that region. The Basque population inhabits the core of the Franco-Cantabrian region and almost all of their M269 sublineages belong to the subhaplogroup S116. If M269 had originated in that area, it would be logical to find more M269 sublineage variability. Considering all of the above, in the present thesis work, we considered more likely a place of origin and/or expansion of M269 in Eastern Europe with a subsequent migration to the west, with the appearance of its sublineages during the advance of the lineage through West Europe over the time. Nonetheless, the homogeneity in the variability of the Y-STRs within M269 made it difficult to pinpoint a more likely origin [259]. The maximum frequencies of S116 and DF27 lineages in the Basque region, with its pattern of decreasing frequency gradient with distance, and the extremely low frequencies of M529 and U152, suggest that this area could be the source for S116 and its subhaplogroup DF27. Thus, we proposed the following scenario (Figure 21):

1. Origin of M269 in Eastern Europe.
2. Origin of L11 on the way of westward expansion of M269 [258].
3. Spread of L11 throughout West Europe, as suggested by the frequencies of L11* in different parts of the Atlantic coast [258,259].
4. Origin of U106 around the southern coast of the North Sea.

5. Origin of S116 in the Franco-Cantabrian area.
6. Origin of DF27 sublineage from individuals inhabiting the Franco-Cantabrian area, while other S116 individuals spread to the rest of Iberia and Europe along the Atlantic and Mediterranean coast giving rise to the subhaplogroups M529 and U152, respectively.
7. U152, DF27 and M529 spread and occupied their current territories, with M529 and U152 re-entering the Iberian Peninsula during subsequent migrations.

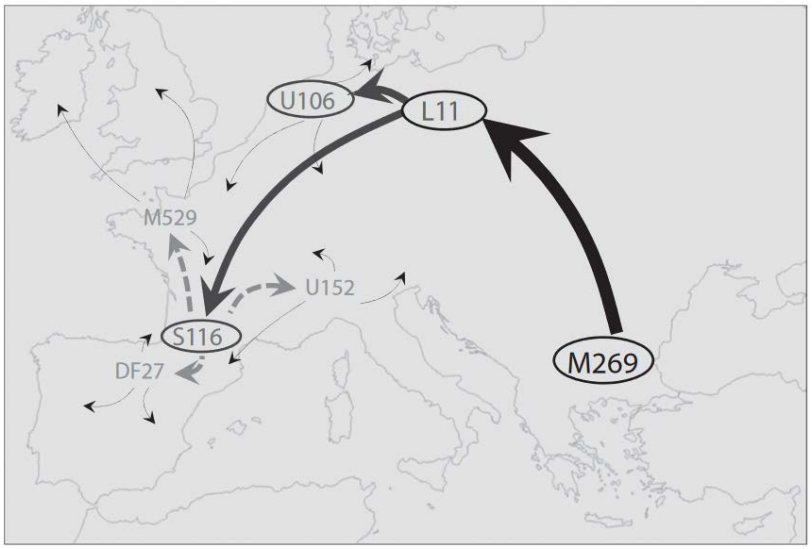


Figure 21. Evolutionary proposal for sublineages of M269 in Europe proposed in *Study Number 1*. The older the movement, the thicker the arrow. The thinner arrows indicate the current distribution of the younger sublineages here studied.

Finally, thanks to the last published studies of Y chromosome resequencing and analysis of ancient DNA, it has been possible to obtain bias-free whole sequences of the male specific region of the Y chromosome (MSY) that have reliably added a temporal scale to the spread of the Y chromosome diversity. One of the most outstanding features of the recent evolutionary history of the Y chromosome is that it appears to have happened in bursts, with lineages rising to high frequencies in the wake of major changes in lifestyle and technological innovations, such as the arrival of the Neolithic or the recently acknowledged demographic upheaval caused by the Bronze Age in Europe [165,350].

The direct dating performed from MSY sequence variation has put the origin of M269 in the early Bronze Age, around 4,500 years ago (ya) [165,168], which is consistent with the information provided by the ancient DNA record where the first R1b-M269 ancient individuals have appeared in archaeological sites from the late Neolithic and early Bronze Age [346,350–352]. The studies of Haak and colleagues[350], and Allentoft and colleagues[351] associated M269 lineage with the arrival of Yamnaya steppe migrants in Central and Northern Europe after 5,000-4,500 ya, as a surge in this lineage is

indeed seen at that time. Furthermore, Poznik and colleagues [165] observed evidence of population bursts in Western Europe around 4,800-5,900 ya associated to lineages within R1b-L11 (a R1b-M269 sublineage). The later time also coincides with the origins of the Corded Ware culture in Eastern Europe and the Bell-Beakers in Western Europe [350], the latter associated with M269 lineage and the spread of its sublineage S116 throughout most of its present-day distribution [346]. Moreover, Martiniano and colleagues [352] detected a discontinuity in the Y chromosome during the Bronze Age in the Iberian Peninsula after analyzing Neolithic and Bronze Age samples from Portugal. These last findings have finally brought light to the M269 controversy, and although our TMRCA estimation in *Study Number 1* was not reliable due to using the EMR, our proposed dispersion scenario of M269 arriving to West Europe from Eastern Europe seems more or less compatible with the last reported findings about the origins and expansion route of M269.

5.2 The Iberian near-specific paternal lineage DF27

5.2.1 Paternal lineages in the Iberian Peninsula

In the Iberian Peninsula, which nowadays hosts the countries of Spain and Portugal, the most common paternal lineage is the West European R1b-M269 in frequencies over 50%, as shown by *Study Number 1* and Myres and colleagues [258]. Among the M269 branches present in Iberia, the one that stands out the most is the haplogroup DF27, as described in *Studies Number 1, 2,* and *3*, which occurs in frequencies over 30% in that area. Apart from that main R1b sublineage, other minor paternal lineages are observed like J2 (8%), E1b1b (7%), I2a (4.5%), and G (3%), among others [240,353–355].

The highest contribution of paternal lineages to the Iberian Peninsula seem to have been from Late Neolithic and/or Bronze Age origin [346,352], apparent by the prevalence of M269 and its derived sublineages, as detailed in *Studies Number 1, 2,* and *3*. Later migrations to the Iberian Peninsula may have also contributed with more lineages due to the influence of different historical groups such as Phoenicians, Carthaginians, Jews, Romans, Vikings, and Levantine Arabs, but in a smaller proportion [353–356]. Considering all of the above, it is clear that the Iberian Peninsula seems to possess a complex genetic structure defined by different human migrations. For that reason, the fine knowledge of the paternal landscape of this region is of high interest, as it would allow making more reliable biogeographic inferences in Forensic Genetics. In this regard, the results provided by *Studies Number 1, 2,* and *3* are of great value for forensic, population and evolutionary genetics.

The paternal lineage DF27 was first reported in the study of Rocca and colleagues [224] conducted by citizen scientists, which discovered new variants within haplogroup R1b-L11 using the data of

the 1,000 Genomes Project [357]. This lineage was known among the burgeoning amateur genetic genealogy community, although no publications were devoted to it until the publication of the *Studies Number 1, 2, and 3*.

5.2.2 Distribution and structure of DF27 haplogroup

The research conducted by Rocca and colleagues [224] first reported that out of 49 samples previously categorized as belonging to S116* (xU152, xM529) 42 belonged to the newly defined Y-SNP DF27. Most of the DF27 derived samples were from Iberian or Latin American populations and linked the newly discovered variant to the previously unclassified S116* (xU152, xM529) reported in Iberia and some regions of France [259].

The *Studies Number 1, 2,* and *3* analyzed the haplogroup DF27 and/or its sublineages in populations from Spain, Portugal, and France, and collected frequencies from the 1,000 Genomes Project [357] (Figure 22A). DF27 was found at frequencies 30-50% in Iberia, displaying maximum values in the Native Basque population (70%). Outside of Iberia, the frequencies drop to a range of 6-20%. Thus, we confirmed that DF27 is located in different geographic area that the one occupied by the other two major S116 sublineages U152 and M529, as discussed in the previous section. The dissection of DF27 in its main sublineages (Figure 22B) performed in *Studies Number 2* and *3*, allowed us to obtain a detailed picture of its distribution and phylogenetic structure in West Europe, observing a pattern of distribution similar to its mother haplogroup. The same studies revealed that the sublineages of DF27 show a moderate geographical differentiation: Western Iberia (especially Asturias, Portugal, and Galicia) is characterized by low values of R1b-Z195 derived chromosomes and relatively high frequencies of the paragroup DF27* (xZ195, xL881, xL617); North-Central Spain (Basque Country and Cantabria) displays relatively high frequencies of the subhaplogroup Z220 compared to the branch L176.2, which is more abundant in Eastern Iberia (Catalonia, Valencia, and Balearic Islands) (Figure 22A).

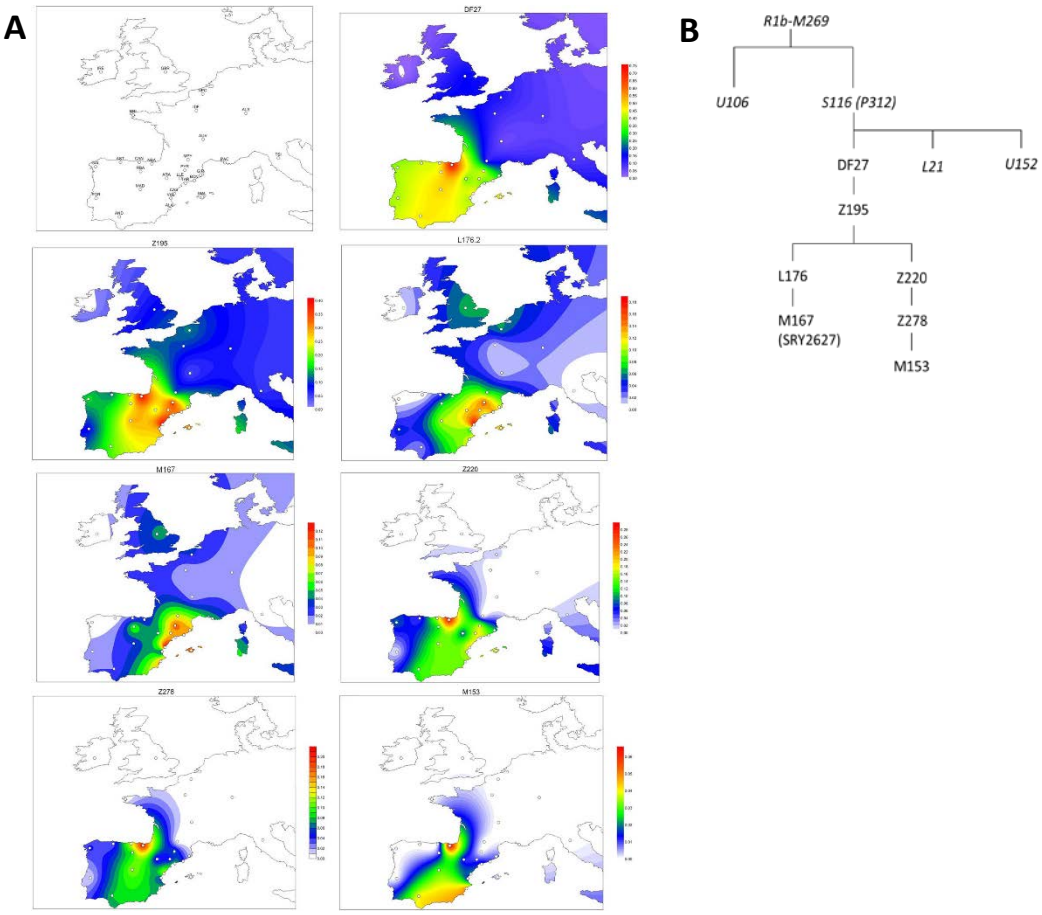


Figure 22. A) Contour maps of the derived allele frequencies of the SNPs analyzed in *Study Number 3*. B) Simplified phylogenetic tree of the R1b-M269 haplogroup.

The Z220 sublineages Z278 and M153 display a similar pattern of frequency distribution and peaks as Z220. In the case of M153, our results and other available studies [256,354,358] show that it is confined in the Iberian Peninsula, with higher frequencies among the Basque population but rarely present at frequencies >1% elsewhere. Conversely, the L176.2 sublineage M167 (also known as SRY2627) peaks in Catalonia and the lands settled from Catalonia in the 13th Century [359], Valencia and the Balearic Islands. In addition to the typing of this Y-SNP performed on *Studies Number 2* and *3*, other authors also analyzed this haplogroup in Iberian and other European populations [247,256,354,358,360–363], confirming its distribution centered in the eastern half of Iberia. Considering all together, the geographic differentiation of DF27 subhaplogroups observed in *Study Number 3* remind of past historical East-West patternings of Iberia:

1) In the pre-Roman era, the Iberian Peninsula was divided in two areas due to the presence of two types of peoples speaking different languages. The Indo-European Celts occupied the center and the West of the Iberian Peninsula while the non-Indo-European eponymous Iberians inhabited the Mediterranean Coast and hinterland (Figure 23) [364,365].

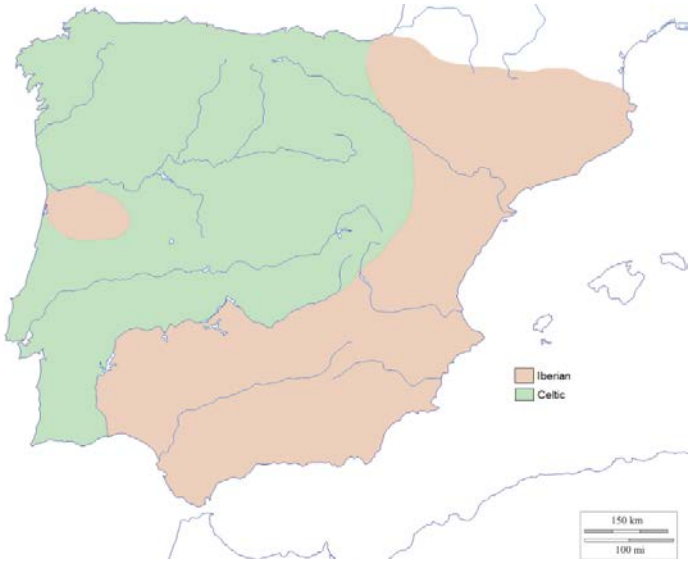


Figure 23. Peoples inhabiting the Iberian Peninsula in the pre-Roman era and their relative position. Adapted from [365].

2) In the Middle Ages, when the Christian kingdoms in the north part of Iberia expanded southwards and regained control of the lands held by the Muslims (Figure 24) [366].

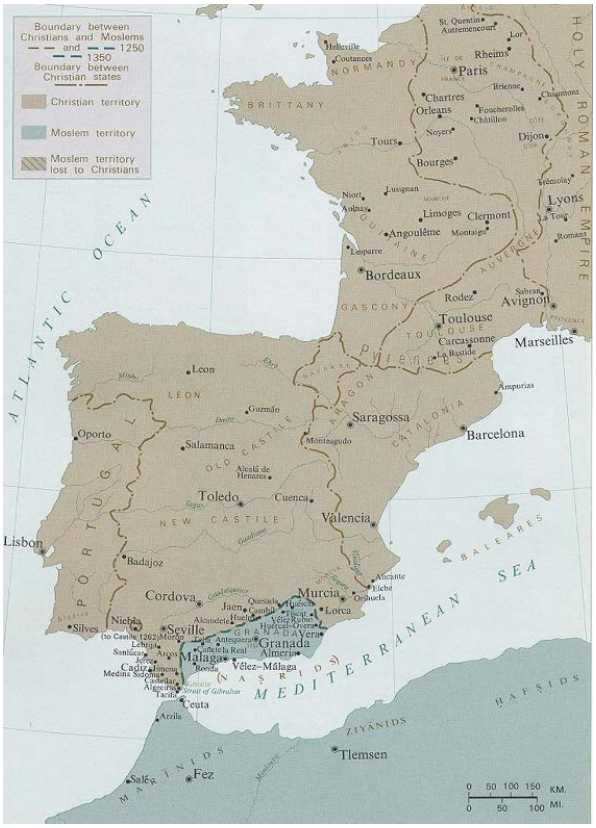


Figure 24. Overview of the territory partition at the end of the Christian Reconquista in the Iberian Peninsula between the years 1,250-1,350. Extracted from [367].

On the other hand, in *Study Number 2* we gathered frequency data of DF27 and its subhaplogroups from the 1,000 Genomes Project [357] in order to get a better picture of the global distribution of

DF27. Until the releases of *Studies Number 1, 2,* and *3* the presence of DF27 in the academic literature was rather obscure, it was not directly genotyped in any study, and only scarce information concerning two of its subhaplogroups (M167 and M153) was available. The bibliographic search included several populations from Europe, America, Asia, and Africa. The frequencies found in other European populations (Britain, Italy, Finland) verified the decreasing frequency pattern with distance detected for DF27 in *Studies Number 1, 2,* and *3*. Furthermore, we also detected the presence of DF27 in Latin America, finding frequencies similar to those observed in the Iberian Peninsula in Mestizo populations from the areas associated with a strong Iberian influence during the Colonial period (especially Colombia and Puerto Rico). The present finding raised the interest of using DF27 as an indicative of the degree of patrilineal Iberian versus Native American admixture. As expected, DF27 was absent in the populations from Asia and Africa, although it was found in very low frequencies in African-Americans and Afro-Caribbeans.

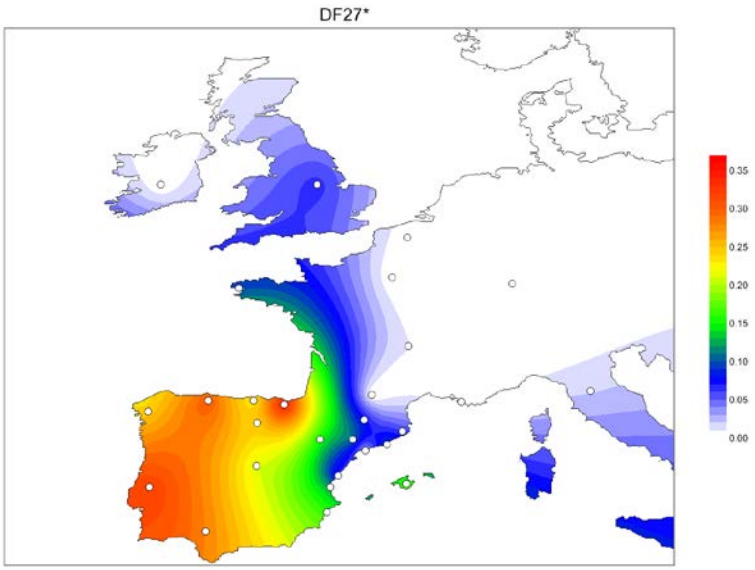


Figure 25. Frequency contour map of DF27* obtained in *Study Number 3*.

Conversely, the frequency of the paragroup of DF27* (xZ195, xL881, xL617) was the highest in the Native Basques (30%) and Western Iberian populations like Asturias, Galicia, and Portugal (Figure 25). The high frequency of the paragroup could mean that these individuals may harbor yet unknown branches of DF27. The median joining networks constructed in *Study Number 2* with individuals from the paragroup did not show any pattern of internal variability, not hinting the presence of new subhaplogroups not yet found. Likewise, it is also feasible that these individuals may just harbor individual variations. Since we genotyped pre-ascertained SNPs, a global characterization of the whole sequence diversity of DF27 haplogroup through next generation sequencing would allow to solve the paragroup and to run more precise statistical analyses.

The networks constructed in *Study Number 1* to evaluate the Y-STR variation within DF27 haplogroup hinted to a bipartite structure owing to the presence of three different haplotypes in the Y-STRs DYS437, DYS448, and YGATAH4. *Studies Number 2* and *3* confirmed the phylogenetic split that separated the haplotypes that carried the derived allele for Z220 (that is, belonging to Z220*, Z278*, and M153) from the rest of DF27 chromosomes (that is, belonging to DF27*, Z196*, L176.2*, M167, and S68). The median haplotype for the Z220 derived chromosomes was 11-14-18 at YGATAH4-DYS437-DYS448 while it was 12-15-19 for the rest of the DF27 chromosomes (Figure 26). These same results were also confirmed by principal component analysis (PCA) among Y-STR haplotypes. In addition to that, Z220 node groups also showed divergent branches that separated Z220* chromosomes from Z278* due to differing haplotypes in DYS390 and DYS456.

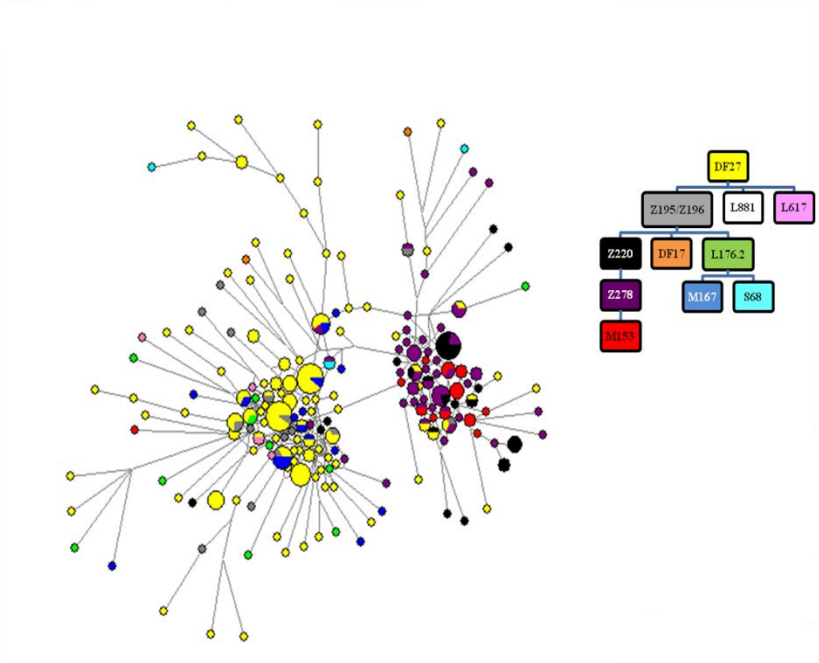


Figure 26. Median joining network of DF27 haplogroup in the populations of Asturias, Cantabria, native Basques, resident Basques, and Aragon obtained in *Study Number 2*. The phylogenetic split for DF27 haplogroup is due to differing haplotypes for YGATAH4-DYS43-DYS448 Y-STRs.

5.2.3 Origin and evolution of DF27

The age of DF27 haplogroup seems clear considering the TMRCA estimations performed in *Studies Number 2* and *3* with independent samples (our samples versus 1,000 Genomes Project [357]) and independent methods (variation in 14-15 Y-STRs versus whole Y chromosome sequences). DF27 appeared around 4,000-4,500 ya, which coincides with the population upheaval in West Europe at the transition between the Neolithic and the Bronze Age [350,351]. As discussed in the previous section, the estimates obtained in *Study Number 1* using the EMR were not considered to be reliable.

Regarding the population dynamics of DF27, the Y-STR diversity in the *Study Number 3* dataset is much more compatible with a population growth model than with stationarity, as shown by the Approximate Bayesian Computing (ABC) results. However, contrary to other lineages such as S116, G2a, I2, or J2a, the start of its growth is closer to the TMRCA of the haplogroup. These results indicate that DF27 started its expansion around 3,000-3,500 ya, shortly after its origin. The median time for the start of the expansion is older in DF27 in comparison with other groups, which could suggest the effect of a different demographic process. Nevertheless, all the highest probability density (HPD) intervals broadly overlap and thus, suggest that a common demographic history may have affected the whole of the Y chromosome diversity in the Iberian Peninsula. The HPD intervals cover a broad timeframe, and could reflect the post-Neolithic population expansions that occurred from the early Bronze Age to the Roman Empire [366].

While when the DF27 lineage appeared seems clear, where it might have originated has been more difficult to determine, as shown in *Studies Number 2* and *3*. If we considered only haplogroup frequencies, DF27 would seem to have appeared in the Basque Country. However, Y-STR internal diversity measures and age estimates for DF27 and most of its sublineages were lower in the Basque Country than in the other populations. The high frequencies found in the Basques could be better explained by genetic drift due to geographical and cultural isolation rather than by an origin in that area, something that could have also decreased the internal diversity of DF27 among the Basque people. We also considered an origin for this lineage outside of the Iberian Peninsula, which would be similar to the external origin of M269, even if DF27 reaches the highest frequency in that region. This search for an external origin would be limited to France or Great Britain, as DF27 is rare or absent elsewhere. The Y-STR data available for France (no data was available of Great Britain) pointed to a lower diversity, and the obtained TMRCAs were younger than in Iberia. We consider unlikely that DF27 originated in France since unlike in the case of the Basques, genetic drift in a traditionally isolated population would seem an improbable explanation for this pattern. For that reason, the most plausible hypothesis would be a local origin of DF27 in the Iberian Peninsula.

Within Iberia Aragon showed the highest diversity and the oldest age estimates for DF27, Z195, and L176.2. Nonetheless, since the sample size was small, any conclusion should be taken with caution. By contrast, the TMRCA estimations for lineages Z220 and Z278 were older in populations from North-Central Spain. Concerning the sublineage M153, since it seems almost restricted to the Basque Country, a local origin in that area seems possible, although the diversity and age values cannot be compared due to the scarcity of M153 chromosomes outside of that region.

5.2.4 Relevance and forensic applicability of DF27

As discussed in detail in the present doctoral thesis work, DF27 possess relatively high frequencies in Iberia and Iberian-derived populations and is rare elsewhere. For that reason, a potential Forensic Genetics application for this Y-SNP would be for inferring the paternal biogeographic origin of an unknown male contributor to a crime scene. This could be helpful in missing person cases, mass disaster identifications, and cases where the person of interest is of Iberian descent, especially when dealing with admixed populations composed of individuals of diverse origin. The finding of a vestige that possess DF27 paternal lineage could point to an individual of Iberian ancestry, particularly if a sublineage within DF27 is also found, since they are uncommon outside of Iberia. Apart from that, *Study Number 3* showed that some of the sublineages of DF27, like Z220 and L176.2, display moderate geographical differentiation inside the Iberian Peninsula, being more frequent in Eastern Iberia or North Central Spain respectively. The presence of these lineages could vaguely point to some concrete areas of the Iberian Peninsula, although it should be considered with caution.

Despite the potential interest to use the Y-SNP DF27 as a marker of Iberian ancestry for forensics, there are some limitations concerning its specificity and sensitivity that should be taken into account. If we compare the frequency of DF27 in Iberia with the CEU sample (Utah residents with Northern and Western European ancestry) of Europeans-Americans from the 1,000 Genomes Project [357], DF27 is just 4.19 times more frequent in Iberians than in CEU, a ratio that increases to 6.82 for the sublineage Z220. Thus, the analysis of DF27 alone could not guarantee significant investigative leads in many cases. Part of this limitation derives from the intrinsic qualities of the Y chromosome and the Y-SNPs, lack of recombination and low mutation rate that ends up in their transmission practically unchanged from fathers to sons, preventing to discriminate between male individuals from the same paternal lineage. In addition to that, we cannot disregard the effect of recent migrations, which has led to populations growing more diverse and the dissemination of paternal lineages previously restricted to more specific geographic areas.

Hence, we consider that DF27 is a potentially useful marker in Forensic Genetics for determining paternal biogeographical ancestry that could be used in conjunction with other markers such as AIMs (Ancestry Informative Markers) and/or mitochondrial DNA in order to ascertain the ancestry of a vestige.

Apart from the evident forensic application for DF27 and its sublineages, these haplogroups may also be used to trace migratory events involving Spanish or Portuguese men, particularly outside of Europe. A clear example of this can be observed in the Latin American populations, where DF27

seems to correlate with the amount of male-mediated Spanish admixture, as it is clearly less frequent in the populations with a stronger Native American component, such as Peru and Mexico [357]. Furthermore, even within Europe DF27 can still be informative to detect short-range migration events such as that from Northern France to Flanders [368], or from Spain to the Low Countries [369]. Thus, using DF27 in particular could serve to trace other migration events even within the Iberian Peninsula, such as the medieval expansion of the Aragon kingdom towards the Mediterranean in the 14th-15th centuries, or the Castilian occupation of Flanders in the 17th century. In this regard, the study published by Larmuseau and colleagues [369] analyzed the impact of the presence of Spaniards during the Dutch Revolt on the genetic variation in the Low Countries. By the analysis of the DF27 subhaplogroups Z195 and M167 it could be verified that there was no higher occurrence of Iberian specific lineages associated with a historical gene flow event in the Low Countries like the one that happened during the 16th century with the Spanish Furies [370,371]. The use of DF27 served to assess the impact of those circumstances on the genetic variation in the current autochthonous populations of the Low Countries.

Finally, DF27 could also be relevant in genealogical studies, particularly in the study of the Y chromosome in connection with surnames, since the latter is often transmitted through the male line [159]. For this purpose, Y-STRs are usually analyzed and, considering the Y-STR mutation rates, the similarity in Y-STR haplotypes between men sharing the same surname is usually taken as indicative of a shared genealogical origin. However, in the case of M269, the most common European paternal lineage, the diversity within the haplogroup is rather small [361,372] and using only Y-STRs may result in homoplasy, rather than shared origin, causing Y-STR haplotype convergence. For that reason, Y-SNPs are much more informative in the case of surnames connected to M269 lineage, as individuals belonging to different M269 subhaplogroups cannot be distinguished from each other based solely on Y-STR haplotype variation. In this context, when trying to ascertain the founding events of surnames within Iberia, that is, to study the history of surnames from this region, the Y-SNP DF27 should be used instead of Y-STRs. Since the majority of the male individuals from this geographic area belong to the lineage M269, the analysis of Y-STRs would not be informative enough for the study of Iberian surnames and, thus, analyzing the Iberian near-specific DF27 would provide much more information.

5.3 Estimating the time to the most recent common ancestor of Y-SNPs

5.3.1 Interest

The estimation of the time to the most recent common ancestor (TMRCA) is a popular method to measure the age of a concrete DNA mutation, such as genetic markers associated to paternal or maternal lineages. The inference of the divergence time between different populations has been of great interest in the study of population evolution. The availability of genome-wide data thanks to the advent of next generation sequencing (NGS) has provided an unprecedented opportunity to study the demographic changes and migration patterns throughout the history of the human populations [373]. Many of the existing population genetics inference methodologies have been built on the basis of the coalescent theory [374,375], although these can be generally classified according to the type of genetic data used as input and the assumptions about population demography. Some methods assume a model for the genealogy while others do not [376].

Several studies have provided refined Y chromosome phylogenies [165,168] where the branch lengths are proportional to time allowing, thus, the direct estimation of the TMRCA of the nodes. Furthermore, the coalescence of their branches can be used to trace the effective population size back in time. These studies have allowed obtaining reliable TMRCAs for several Y-SNPs from whole Y chromosome sequences [168]. On the other hand, Y-STRs are also commonly used in population genetics since they exhibit a higher degree of variation and, owing to that, can be a great tool for discriminating between closely related chromosomes. Though their mutation processes can be rather complex and saturate much faster than SNPs, Y-STR markers can also provide good estimates for relative recent events [140].

The estimation of the TMRCA in *Studies Number 2* and *3* allowed us to estimate reliably the age of the lineage DF27 and its sublineages both from Y-STRs and from whole Y chromosome sequences present in the 1,000 Genomes Project dataset [357]. The comparison of the ages estimated through different genetic data (Y-STRs vs. whole chromosome) enabled us to verify the quality of the method used for the TMRCA calculation. Thanks to the solid estimation obtained with the algorithm Rho by following the advised precautions and using suitable mutation rates, we could understand the historical context of the origin of DF27 and its demography, how such a young haplogroup gave rise to a great diversity of sublineages in a relatively short time. Some of its subhaplogroups, like Z195, originated almost simultaneously with appearance of DF27, and the other two main subbranches appeared around 1,000-1,500 years after the TMRCA of DF27 (*Studies Number 2 and 3*).

5.3.2 Limitations of TMRCA estimation from Y-STRs

As discussed in the previous sections, the estimation of the TMRCA has been a controversial issue in population genetics mainly due to the methods, since it has been difficult to assess their reliability due to the lack of datasets where the true time of interest is known [233]. The main reason of the discrepancies between the obtained TMRCAs is the selected mutation rate, as well as the mathematical model applied to perform the inference [233,348]. Mutation rate assumptions have a large impact on molecular dating. By using the evolutionary mutation rate (EMR) proposed by Zhivotovsky [310,311] considerably older TMRCAs are obtained in comparison with the ones obtained using the germinal mutation rate (GMR) [349]. Another three variables that should be considered are the generation time, whose values in the literature range from 15 to 30 years, the set of Y-STRs used, and the presence of 'outlier' individuals. Concerning this last variable, Boattini and colleagues stated that the presence of 'outlier' haplotypes could inflate significantly the ages of haplogroups and suggested that for dating purposes these haplotypes should be detected and removed [362].

In the present thesis work, the TMRCA for several Y-SNPs was estimated using four approaches:

1) Using Rho (ρ) statistic selecting the EMR with a set of Y-STRs (*Study Number 1*).
2) Using ρ statistic selecting a mean GMR adjusted to the set of Y-STRs used for the estimation (*Study Number 2*).
3) Using a weighted ρ statistic, which takes into account that mutations at slow STRs take longer to accumulate than mutations at faster STRs, selecting a mean GMR adjusted to the set of Y-STRs (*Study Number 3*).
4) Using ρ statistic with a mutation rate adjusted to the ~ 10.36 Mb of the Y chromosome amenable to short-read sequencing and SNP detection [165] (*Study Number 3*).

In this way, we obtained older TMRCAs for S116 and DF27 using the EMR, placing their origin in the Paleolithic (11,673 ± 1,962 and 10,468 ± 1,831 years ago respectively). However, with the approaches II and III, which used a GMR adjusted to our set of Y-STRs, we obtained TMRCAs for DF27 around 4,000-4,500 ya, transferring its origin at the transition between the Neolithic and the Bronze Age. Furthermore, using whole genome sequence variability with approach IV we obtained remarkably similar estimations to those obtained with Y-STRs, verifying the quality of the method and confirming that Y-STRs can also provide good time estimations if we consider a suitable mutation rate, among other variables.

Finally, we selected the ρ statistic for inferring the TMRCA. The accuracy of this method has been highly perceived as independent of demographic parameters, but this assumption is false [377] and has often led to wrong TMRCA estimates. Taken that into account, we consider that using ρ properly adjusted and selecting a suitable set of Y-STRs and mutation rate, this method can provide reliable estimations for relative recent times. Nevertheless, we are aware of the limitations of this statistic, and recommend to use it with caution since demography has proven to be a strong confounding factor in estimating molecular dates accurately, especially for populations in which bottlenecks, founder events, and size changes have played important historical roles [377].

5.4 Evaluation of the new 15 Y-SNP minisequencing multiplex

5.4.1 Assessment of the 15 Y-SNP minisequencing panel

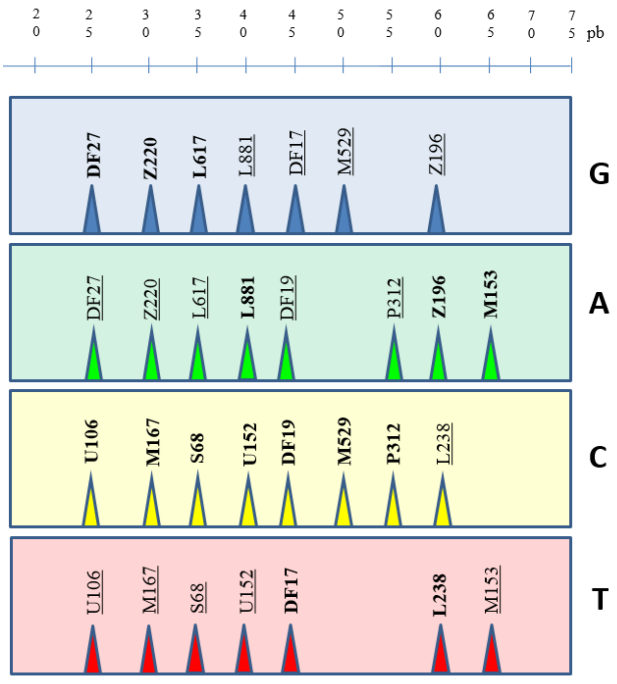


Figure 27. Diagram of the theoretical positions of the Y-SNPs included in the 15-plex minisequencing panel presented in *Study Number 4*. Ancestral alleles appear in bold letters; Derived alleles appear underlined.

In *Study Number 4* a novel Y-SNP minisequencing multiplex was developed for the fine subtyping of the Iberian paternal lineage DF27 in a unique reaction (Figure 27). This panel was designed for its application in forensic and population genetics, specifically for the inference of paternal biogeographical ancestry. In forensic casework, DNA is not always in optimal condition due to degradation, low copy number, or the presence of inhibitors [378–380]. These factors, alone or altogether, may give raise in incomplete or null genetic profiles and, therefore, robust panels with

short length amplicons are preferred in forensic routine. With that in mind, we carefully designed the 15-plex to obtain the shortest length of amplicons, with the additional challenge that involves designing primers for the analysis of the Y chromosome owing to its complex structure [110,118]. Besides, we also made the length of the minisequencing primers as short as possible to make this method a cost-effective approach of easy implementation. Moreover, the reduced number of coamplified fragments enables the optimization of the assay in any forensic laboratory and minimizes competition effects during the amplification of samples with low quantities of DNA.

On the other hand, the development of any multiplex minisequencing assay is a complex process where apart from the reaction conditions and marker selection, it is critical to test all the included variants. This entails that the minisequencing panel must able to detect all the alleles of the markers included in the assay, which is usually verified by analyzing samples that display all the alleles of the selected variants. However, this can prove to be a difficult task when some of the variants display scarce frequencies or no samples of the particular allelic variant are available. Therefore, a method is necessary to allow such variants to be included during the optimization of the minisequencing assay. The application of site-directed mutagenesis in *Study Number 4* allowed us to confirm that the 15-plex Y-SNP panel was able to detect the derived allelic variants of the rare Y-SNPs DF19 and L881 by producing them *in vitro*. We placed the changed nucleotide as close as possible to the 5' extreme of the primer in order to prevent primer-DNA hybridization problems due to the mismatch produced by the changed nucleotide. These results confirm that site-directed mutagenesis is a highly appropriate tool to produce variants *in vitro* for minisequencing reactions, as was also suggested on a mitochondrial DNA minisequencing assay [319].

5.4.2 Applicability of the novel 15-plex minisequencing panel

The new 15-plex is a reproducible method that allows subtyping the Iberian near-specific lineage DF27 in a single reaction to the highest phylogenetic resolution to date. Some of the included DF27 subhaplogroups, such as Z220 and M167, display moderate geographical differentiation, as discussed in previous sections and *Study Number 3*. Thus, the typing of DF27 and its subhaplogroups in forensic samples by the use of the present multiplex, in combination with other markers like Y-STRs or AIMs, could be of interest for inferring the paternal biogeographic origin of an unknown contributor to a crime scene. Moreover, given the lower dispersion of the DF27 sublineages the detection of one of them in an area outside of the Iberian Peninsula could hint a suspect of Iberian ancestry, which would be useful in admixed populations composed of different ethnic groups (such as the United States or some Latin American countries).

Conversely, the inclusion in the 15-plex of other common M269 sublineages above DF27 (i.e., U106, S116, U152, M529, L238, and DF19) that are geographically localized in West Europe (as described in *Study Number 1*), makes this multiplex also applicable to a broader range of samples, as it also detects West European lineages. Therefore, the present 15-plex could be used for forensic purposes as well as for the study of the European paternal contribution to the genetic substrate of different population groups given the dispersion of some West European populations (such as Great Britain, France, Spain, or Portugal) over large areas of America, Asia, and Africa throughout history [366].

Previous Y-SNP minisequencing panels include the European haplogroup R1b but not many of its derived sublineages [177,179,180,183], hence, our 15-plex can be combined with those panels in order to provide an increased power of paternal lineage discrimination. Likewise, this Y-SNP multiplex could also be combined with the analysis of the mitochondrial DNA lineage to complete the information on biogeographical ancestry, both paternal and maternal, which would be of particular interest in the study of admixed populations or individuals.

5.4.3 Application of the 15-plex to real cases

Y-SNPs possess some characteristics that makes them suitable for their use in forensics [74,95] and have certain advantages over other makers. In certain cases where the genotyping of other markers has failed due to the DNA being degraded or in low copy number, or no suspects are known, is where the analysis of Y-SNPs can provide valuable clues by inferring the paternal biogeographic ancestry of a vestige [95]. Therefore, the new 15-plex for the subtyping of DF27, which has demonstrated to be a highly useful and efficient multiplex, has been incorporated to the Diagnostic Service of the Biological Parentage, Genetic Identification, and Ancestry Identification of the DNA Bank of the University of the Basque Country UPV/EHU. Within this service, it has also been applied in paternal biogeographical ancestry inference of skeletal remains exhumed from mass graves of the Spanish Civil War and posterior dictatorship by our group.

5.5 Evaluation of the novel Slowly Mutating Y-STR panel

5.5.1 Efficiency of the novel multiplex

In *Study Number 5* a new Slowly Mutating (SM) Y-STR panel was developed and evaluated for forensic purposes (Figure 28). We selected the best candidate makers with a low mutation rate (~ 10^{-4} mutations/generation) [151] and an intra-population gene diversity >0.4 in different world population groups. The sensitivity studies performed allowed setting a minimum quantity of DNA

of 200 pg where complete genetic profiles are obtained with peak heights above 50 RFU, which makes this multiplex suitable for forensic samples. We also evaluated the stability of the panel in the presence of two common inhibitors in forensic casework, such as hematin and humic acid, obtaining complete profiles with ≤ 500 ng/µl of both inhibitors. The obtained results demonstrated the robustness and sensitivity of this new multiplex panel for forensic use.

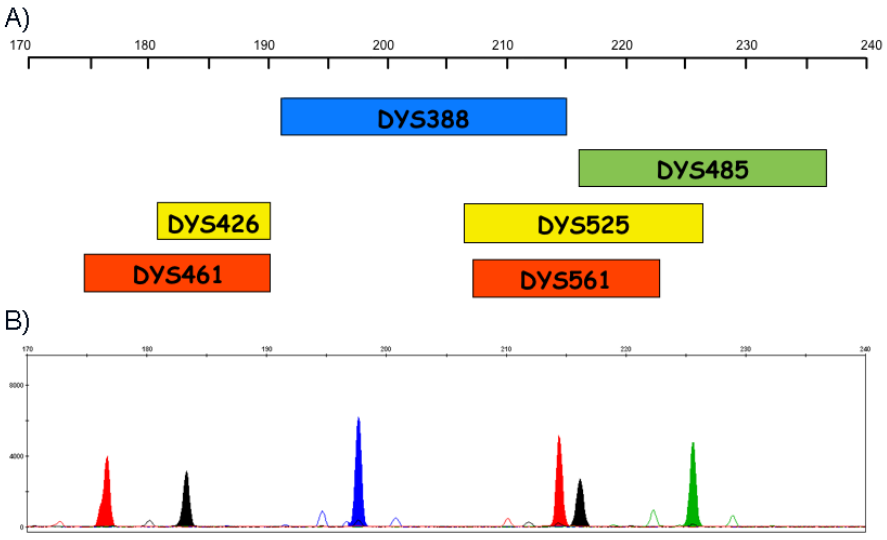


Figure 28. A) Diagram of the panel developed in *Study Number 5*. (B) A representative electropherogram showing the profile of 1.5 ng control DNA amplified at the optimized PCR conditions. The peaks correspond to: DYS388 (blue), DYS485 (green), DYS426 (black), DYS525 (black), DYS461 (red), and DYS561 (red). The GeneMapper ID-X plots are presented as combined all dyes.

The population study revealed that the SM Y-STR panel provides a moderate power of discrimination between male haplotypes in most populations in spite of the low mutation rate. Nevertheless, the inclusion of SM Y-STR markers in forensic casework in combination with routine panels may be a valuable tool in exclusion of kinship cases since mutation events are rare to occur in these markers among close relatives. The panel allowed the differentiation of half of the haplotypes in the Latin American populations, and similar results were obtained for the African and Asian groups. The Native American populations, as well as the Europeans, showed lower discrimination capacity, the first ones due to their genetic isolation attributable to cultural or/and geographic barriers [381], and the seconds owing to the high resemblance of Y-STR haplotypes belonging to M269 lineage [361,372].

In the same way, we observed relative correspondence of SM Y-STR haplotypes to concrete Y-SNPs, particularly in the case of the haplogroups R1b, Q, and O. These results do not point to the capability of an unambiguous prediction of haplogroups using only SM Y-STRs, as it has been

previously sated for other Y-STRs [163,382], but hints their potential to be included in routine analysis with conventional Y-STRs to help optimize Y chromosome haplogroup prediction. It is necessary to consider that haplogroup prediction from Y-STR haplotypes is not always trustworthy and does not work well for the prediction of some lineages due to the scarcity of reference data [163]. In this sense, both conventional Y-STRs and SM Y-STRs are limited in their prediction capability.

5.5.2 Applicability of the SM Y-STR panel

The 6-plex panel was designed considering the potential utility of SM Y-STRs in forensic casework, particularly in exclusion cases where minimal discrepancies are considered to be critical and end up reported as exclusions. In this sense, the inclusion of SM Y-STR markers may be a useful tool in exclusion kinship cases where minimal discrepancies were found using the conventional Y-STR panels, and when *de novo* mutations may account for the allele inconsistencies. The presence of one or more discrepancies in the SM Y-STRs may offer further evidence for the truthful exclusion of the biological parenthood, considering that mutations occur more infrequently in these markers.

Likewise, the low mutation rate of the SM Y-STRs pointed that they could also be used for phylogenetic studies. Although the sole analysis of SM Y-STRs does not allow an unambiguous assignation of individuals in haplogroups, we consider that the inclusion of these loci in the analyses, together with the conventional Y-STRs, may be helpful to optimize the phylogenetic signals upon current Y-STR panels since they are more stable than other common Y-STR markers (although less stable than Y-SNPs), and allow constructing more robust phylogenetic relationships.

Overall, we consider that the present multiplex has demonstrated that it is a reproducible, sensitive and robust method to be used in tandem with the existing commercial multiplexes for forensic casework, and highlights the potential of combining mixed Y-STR systems (slowly and rapidly mutating) to address distinct time windows.

5.5.3 Application if the SM Y-STR panel to real cases

The use of the SM Y-STR multiplex in conjunction with other commercial multiplexes show potential for particular kinship cases as well as for optimizing and increasing the resolution of the phylogenetic relationships based only on the conventional Y-STR panel sets. Considering this, the SM Y-STR panel has been added to the Diagnostic Service of the Biological Parentage and Genetic Identification of the DNA Bank of the University of the Basque Country UPV/EHU. The 6-plex has been applied in samples from skeletal remains exhumed from mass graves of the Spanish Civil

War and posterior dictatorship for concrete kinship cases where minimal discrepancies were found with current Y-STR panels.

6. Conclusions

Conclusions

1) The characterization of the paternal landscape of Southwestern Europe by the analysis of Y-SNPs in population samples from Spain, Portugal, France, Ireland, and Denmark has revealed that the predominant lineage is R1b-M269 with a distribution in concordance with previous studies. As previously described, we confirmed the geographical location of the main M269 sublineages, being U106 more frequent in Northern and Central Europe while S116 is more common in Southwestern Europe.

2) S116 haplogroup presents a different distribution from the one proposed before by other authors, displaying a decreasing gradient with distance from Northern Iberia, the French western coast, and the British Isles. The maximum frequencies of S116, S116*, and DF27 in the Franco-Cantabrian region, the diversity of S116 sublineages, and their spatial distributions in the Iberian Peninsula and Atlantic coast point to the origin of S116 in the Franco-Cantabrian area. Finally, the paragroup S116* was largely resolved by haplogroup DF27, which is located in a different geographic area that the ones occupied by the other two major S116 subhaplogroups, U152 and M529.

3) The paternal lineage DF27 displays an Iberian near-specific distribution, with frequencies over 30-50% in Iberia that quickly drop to a range of 6-20% outside of that region, being absent in populations from Asia and Africa. Likewise, DF27 is also present in Latin America in areas associated with Iberian or European influence during the Colonial period, such as Colombia and Puerto Rico, being less frequent in populations with stronger Native American component like Mexico or Peru.

4) The high frequencies of the paragroup DF27* could point to the existence of yet unknown subhaplogroups of DF27, although no pattern of internal variability was observed in the median joining networks constructed with the individuals from the paragroup. Nevertheless, only a global characterization of the whole sequence diversity of DF27 through next generation sequencing would allow eventually resolving the paragroup.

5) The TMRCA estimations performed from Y-STRs and whole Y chromosome sequences date the origin of DF27 lineage 4,000-4,500 years ago, overlapping with the population upheaval in West Europe at the transition between the Neolithic and the early Bronze

Age. Considering the age of DF27 in the different populations and the internal diversity of the Y-STRs, the most feasible theory is that DF27 originated in the northeast Iberian Peninsula, although it has not been possible to point out a more specific location within Iberia. However, caution should be taken when estimating TMRCAs because the calculation methods, mutation rates, and the demographic history of each population could affect the accuracy of the estimated times to a high degree.

6) DF27 is a potentially useful marker in Forensic Genetics for determining paternal biogeographical ancestry that could be used in conjunction with other markers such as AIMs and/or mitochondrial DNA to ascertain the Iberian or European paternal ancestry of a vestige. In addition to that, the analysis of DF27 and its sublineages could also be relevant for the study of migratory events involving Spanish or Portuguese men, short-range migrations events within Europe, and genealogical studies involving founding events of surnames within Iberia.

7) A new Y-SNP multiplex system for the analysis of 15 Y-SNPs allowing the fine subtyping of the haplogroup DF27 has been developed. The high resolution accomplished by this panel makes it suitable for paternal biogeographic inference, particularly for Iberian and Southwest European paternal ancestry. Additionally, site-directed mutagenesis proved to be a highly appropriate tool to produce genetic variants *in vitro* for minisequencing reactions, allowing the inclusion of all variants during the optimization process. Thus, this new minisequencing panel has proven to be a robust and reproducible method of easy implementation in most forensics or population genetics laboratories.

8) A novel Slowly Mutating (SM) Y-STR panel that includes six markers with a low mutation rate has been developed for its application in forensic casework in conjunction with the existing commercial multiplexes. The SM Y-STRs panel could be helpful for confirming exclusions in kinship cases where minimal discrepancies in one or a few loci are reported using the regular Y-STR panels, as well as for evolutionary studies, optimizing and increasing the resolution of the haplogroup prediction solely based on the conventional Y-STR panel sets. The SM Y-STR panel proved to be a reproducible, sensitive, and robust system for forensic use through its validation studies, and provided an extensive allele and haplotype reference dataset for future applications.

7. References

1. Watson JD, Crick FH. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*. 1953;171(4356):737-738.

2. Wilkins MHF, Stokes AR, Wilson HR. Molecular structure of deoxypentose nucleic acids. *Nature*. 1953;171(4356):738-740.

3. Lander ES, Linton LM, Birren B, et al. Initial sequencing and analysis of the human genome. *Nature*. 2001;409(6822):860-921.

4. Venter JC, Adams MD, Myers EW, et al. The Sequence of the Human Genome. *Science*. 2001;291(5507):1304-1351.

5. Human Genome Sequencing Consortium I. Finishing the euchromatic sequence of the human genome. *Nature*. 2004;431(7011):931-945.

6. Levy S, Sutton G, Ng PC, et al. The Diploid Genome Sequence of an Individual Human. *PLoS Biol*. 2007;5(10):e254.

7. Wheeler DA, Srinivasan M, Egholm M, et al. The complete genome of an individual by massively parallel DNA sequencing. *Nature*. 2008;452(7189):872-876.

8. Feingold EA, Good PJ, Guyer MS, et al. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science*. 2004;306(5696):636-640.

9. McPherson JD, Marra M, Hillier L, et al. A physical map of the human genome. *Nature*. 2001;409(6822):934-941.

10. Birney E, Stamatoyannopoulos JA, Dutta A, et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*. 2007;447(7146):799-816.

11. Dunham I, Kundaje A, Aldred SF, et al. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012;489(7414):57-74.

12. Cantor C, Smith C. *Genomics: The Science and Technology behind the Human Genome Project*. John Wiley & Sons; 1999.

13. Butler JM. *Fundamentals of Forensic DNA Typing*. Elsevier; 2010.

14. Butler JM. *Advanced Topics in Forensic DNA Typing: Methodology*. Elsevier; 2012.

15. Siegel J, Knupfer G, Saukko P. *Encyclopedia of Forensic Sciences*. 2nd ed. Academic Press; 2000.

16. Landsteiner K. Zur Kenntnis der antifermentativen, lytischen und agglutinierenden Wirkungen des Blutserums und der Lymphe. *Zentralblatt für Bakteriol Parasitenkd und Infekt*. 1900;27:357-362.

17. Dungern E, Hirschfeld L. Über Vererbung gruppenspezifischer Strukturen des Blutes. *Z Indukt Abstamm Vererbungsl*. 1911;5(1):196-197.

18. Goodwin W, Linacre A, Hadi S. *An Introduction to Forensic Genetics*. Wiley-Blackwell; 2011.

19. Dausset J. Iso-leuco-anticorps. *Acta Haematol*. 1958;20(1-4):156-166.

20. Geserick G, Wirth I. Genetic Kinship Investigation from Blood Groups to DNA Markers. *Transfus Med Hemotherapy*. 2012;39(3):163-175.

21. Wyman AR, White R. A highly polymorphic locus in human DNA. *Proc Natl Acad Sci*. 1980;77(11):6754-6758.

22. Jeffreys AJ, Wilson V, Thein SL. Hypervariable 'minisatellite' regions in human DNA. *Nature*. 1985;314(6006):67-73.

23. Jeffreys AJ, Wilson V, Thein SL. Individual-specific 'fingerprints' of human DNA. *Nature*. 1985;316(6023):76-79.

24. Jeffreys AJ, Brookfield JFY, Semeonoff R. Positive identification of an immigration test-case using human DNA fingerprints. *Nature*. 1985;317(6040):818-819.

25. Wanbaugh J. *The Bloodling*. Bantam Books; 1995.

26. Balazs I, Baird M, Clyne M, Meade E. Human population genetic studies of five hypervariable DNA loci. *Am J Hum Genet*. 1989;44(2):182-190.

27. Budowle B, Giusti AM, Waye JS, et al. Fixed-bin analysis for statistical evaluation of continuous distributions of allelic data from VNTR loci, for use in forensic comparisons. *Am J Hum Genet*. 1991;48(5):841-855.

28. Roewer L. DNA fingerprinting in forensics: past, present, future. *Investig Genet*. 2013;4(1):22.

29. Mullis KB. The Unusual Origin of the Polymerase Chain Reaction. *Sci Am*. 1990;262(4):56-65.

30. Bartlett JMS, Stirling D. A Short History of the Polymerase Chain Reaction. In: *PCR*

Protocols. Vol 226. Humana Press; 2003:3-6.

31. Mullis KB, Erlich HA, Arnheim N, Horn GT, Saiki RK, Scharf SJ. Process for amplifying, detecting, and/or cloning nucleic acid sequences. 1989.

32. Saiki R, Scharf S, Faloona F, et al. Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science*. 1985;230(4732):1350-1354.

33. Mullis KB, Faloona FA. Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods Enzymol*. 1987;155:335-350.

34. Mullis KB. The unusual origin of the polymerase chain reaction. *Sci Am*. 1990;262(4):56-61, 64-65.

35. Stoneking M, Hedgecock D, Higuchi RG, Vigilant L, Erlich HA. Population variation of human mtDNA control region sequences detected by enzymatic amplification and sequence-specific oligonucleotide probes. *Am J Hum Genet*. 1991;48(2):370-382.

36. Baird ML. Use of the AmpliType PM + HLA DQAI PCR Amplification and Typing Kits for Identity Testing. In: *Forensic DNA Profiling Protocols*. Vol 98. Humana Press; 1998:261-278.

37. Weber JL, May PE. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am J Hum Genet Hum Genet*. 1989;44(3):388-396.

38. Litt M, Luty JA. A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. *Am J Hum Genet*. 1989;44(3):397-401.

39. Frégeau CJ, Fourney RM. DNA typing with fluorescently tagged short tandem repeats: a sensitive and accurate approach to human identification. *Biotechniques*. 1993;15(1):100-119.

40. Hochmeister MN, Jung JM, Budowle B, Borer U V., Dirnhofer R. Swiss population data on three tetrameric short tandem repeat loci - VWA, HUMTH01, and F13A1 - derived using multiplex PCR and laser fluorescence detection. *Int J Legal Med*. 1994;107(1):34-36.

41. Oldroyd NJ, Urquhart AJ, Kimpton CP, et al. A highly discriminating octoplex short tandem repeat polymerase chain reaction system suitable for human individual identification. *Electrophoresis*. 1995;16(3):334-337.

42. Børsting C, Morling N. Next generation sequencing and its applications in forensic genetics. *Forensic Sci Int Genet*. 2015;18:78-89.

43. Bruijns B, Tiggelaar R, Gardeniers H. Massively parallel sequencing techniques for forensics: A review. *Electrophoresis*. 2018;39(21):2642-2654.

44. Redon R, Ishikawa S, Fitch KR, et al. Global variation in copy number in the human genome. *Nature*. 2006;444(7118):444-454.

45. Turnpenny PD, Ellard S. *Emery's Elements of Medical Genetics*. Elsevier; 2017.

46. Ellegren H. Microsatellites: simple sequences with complex evolution. *Nat Rev Genet*. 2004;5(6):435-445.

47. Collins JR, Stephens RM, Gold B, Long B, Dean M, Burt SK. An exhaustive DNA micro-satellite map of the human genome using high performance computing. *Genomics*. 2003;82(1):10-19.

48. Edwards A, Civitello A, Hammond HA, Caskey CT. DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. *Am J Hum Genet*. 1991;49(4):746-756.

49. Subramanian S, Mishra RK, Singh L. Genome-wide analysis of microsatellite repeats in humans: their abundance and density in specific genomic regions. *Genome Biol*. 2003;4(2):R13.

50. Urquhart A, Kimpton CP, Downes TJ, Gill P. Variation in Short Tandem Repeat sequences — a survey of twelve microsatellite loci for use as forensic identification markers. *Int J Legal Med*. 1994;107(1):13-20.

51. Bär W. DNA recommendations — 1994 report concerning further recommendations of the DNA Commission of the ISFH regarding PCR-based polymorphisms in STR (short tandem repeat) systems. *Forensic Sci Int*. 1994;69(2):103-104.

52. Chakraborty R, Kimmel M, Stivers DN, Davison LJ, Deka R. Relative mutation rates at di-, tri-, and tetranucleotide microsatellite loci. *Proc Natl Acad Sci*. 1997;94(3):1041-1046.

53. Li Y-C, Korol AB, Fahima T, Beiles A, Nevo E. Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. *Mol Ecol*. 2002;11(12):2453-2465.

54. Levinson G, Gutman GA. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol Biol Evol*. 1987;4(3):203-221.

55. Ellegren H. Microsatellite mutations in the germline: implications for evolutionary inference. *Trends Genet*. 2000;16(12):551-558.

56. Fan H, Chu J-Y. A Brief Review of Short Tandem Repeat Mutation. *Genomics Proteomics Bioinformatics*. 2007;5(1):7-14.

57. Budowle B, Giusti AM, Waye JS, et al. Fixed-bin analysis for statistical evaluation of continuous distributions of allelic data from VNTR loci, for use in forensic comparisons. *Am J Hum Genet*. 1991;48(5):841-855.

58. Leclercq S, Rivals E, Jarne P. DNA Slippage Occurs at Microsatellite Loci without Minimal Threshold Length in Humans: A Comparative Genomic Approach. *Genome Biol Evol*. 2010;2:325-335.

59. Bär W, Brinkmann B, Budowle B, et al. DNA recommendations. Further report of the DNA Commission of the ISFH regarding the use of short tandem repeat systems. International Society for Forensic Haemogenetics. *Int J Legal Med*. 1997;110(4):175-176.

60. Gill P. An assessment of the utility of single nucleotide polymorphisms (SNPs) for forensic purposes. *Int J Legal Med*. 2001;114(4-5):204-210.

61. Gusmão L, Butler JM, Carracedo A, et al. DNA Commission of the International Society of Forensic Genetics (ISFG): An update of the recommendations on the use of Y-STRs in forensic analysis. *Forensic Sci Int*. 2006;157(2-3):187-197.

62. Gettings KB, Aponte RA, Vallone PM, Butler JM. STR allele sequence variation: Current knowledge and future issues. *Forensic Sci Int Genet*. 2015;18:118-130.

63. Butler JM. *Advanced Topics in Forensic DNA Typing: Interpretation*. Elsevier; 2015.

64. Martin PD, Schmitter H, Schneider PM. A brief history of the formation of DNA databases in forensic science within Europe. *Forensic Sci Int*. 2001;119(2):225-231.

65. Asplen C, Lane SA. International perspectives on forensic DNA databases. *Forensic Sci Int*. 2004;146 Suppl:S119-21.

66. Corte-Real F. Forensic DNA databases. *Forensic Sci Int*. 2004;146:S143-S144.

67. Hares DR. Expanding the CODIS core loci in the United States. *Forensic Sci Int Genet*. 2012;6(1):e52-e54.

68. Schneider PM. Expansion of the European Standard Set of DNA Database Loci — the

Current Situation. *Profiles DNA*. 2009;12(1):6-7.

69. Budowle B, Moretti TR, Niezgoda SJ, Brown, B L. CODIS and PCR-based short tandem repeat loci: law enforcement tools. In: *Proceedings of the Second European Symposium on Human Identification*. Promega Corp; 1998:73-88.

70. Schneider PM, Martin PD. Criminal DNA databases: the European situation. *Forensic Sci Int*. 2001;119(2):232-238.

71. Hares DR. Selection and implementation of expanded CODIS core loci in the United States. *Forensic Sci Int Genet*. 2015;17:33-34.

72. Wallace HM, Jackson AR, Gruber J, Thibedeau AD. Forensic DNA databases–Ethical and legal standards: A global review. *Egypt J Forensic Sci*. 2014;4(3):57-63.

73. Ruitberg CM. STRBase: a short tandem repeat DNA database for the human identity testing community. *Nucleic Acids Res*. 2001;29(1):320-322.

74. Sobrino B, Brión M, Carracedo A. SNPs in forensic genetics: a review on SNP typing methodologies. *Forensic Sci Int*. 2005;154(2-3):181-194.

75. Visscher PM, Brown MA, McCarthy MI, Yang J. Five Years of GWAS Discovery. *Am J Hum Genet*. 2012;90(1):7-24.

76. Choudhury A, Hazelhurst S, Meintjes A, et al. Population-specific common SNPs reflect demographic histories and highlight regions of genomic plasticity with functional relevance. *BMC Genomics*. 2014;15(1):437.

77. Budowle B, van Daal A. Forensically relevant SNP classes. *Biotechniques*. 2008;44(5):603-610.

78. Scally A. The mutation rate in human evolution and demographic inference. *Curr Opin Genet Dev*. 2016;41:36-43.

79. Roach JC, Glusman G, Smit AFA, et al. Analysis of Genetic Inheritance in a Family Quartet by Whole-Genome Sequencing. *Science*. 2010;328(5978):636-639.

80. Børsting C, Morling N. Mutations and/or close relatives? Six case work examples where 49 autosomal SNPs were used as supplementary markers. *Forensic Sci Int Genet*. 2011;5(3):236-241.

81. Butler JM, Budowle B, Gill P, et al. Report on ISFG SNP Panel Discussion. *Forensic Sci Int*

Genet Suppl Ser. 2008;1(1):471-472.

82. Chakraborty R, Stivers DN, Su B, Zhong Y, Budowle B. The utility of short tandem repeat loci beyond human identification: Implications for development of new DNA typing systems. *Electrophoresis*. 1999;20(8):1682-1696.

83. de la Puente M, Phillips C, Santos C, Fondevila M, Carracedo Á, Lareu MV. Evaluation of the Qiagen 140-SNP forensic identification multiplex for massively parallel sequencing. *Forensic Sci Int Genet*. 2017;28:35-43.

84. Köcher S, Müller P, Berger B, et al. Inter-laboratory validation study of the ForenSeq™ DNA Signature Prep Kit. *Forensic Sci Int Genet*. 2018;36:77-85.

85. Mehta B, Daniel R, Phillips C, Doyle S, Elvidge G, McNevin D. Massively parallel sequencing of customised forensically informative SNP panels on the MiSeq. *Electrophoresis*. 2016;37(21):2832-2840.

86. Gill P, Werrett DJ, Budowle B, Guerrieri R. An assessment of whether SNPs will replace STRs in national DNA databases - Joint considerations of the DNA working group of the European Network of Forensic Science Institutes (ENFSI) and the Scientific Working Group on DNA Analysis Methods (SWGDAM). *Sci Justice - J Forensic Sci Soc*. 2004;44(1):51-53.

87. Butler JM, Coble MD, Vallone PM. STRs vs. SNPs: thoughts on the future of forensic DNA testing. *Forensic Sci Med Pathol*. 2007;3(3):200-205.

88. Mehta B, Daniel R, Phillips C, McNevin D. Forensically relevant SNaPshot® assays for human DNA SNP analysis: a review. *Int J Legal Med*. 2017;131(1):21-37.

89. Nei M. F-statistics and analysis of gene diversity in subdivided populations. *Ann Hum Genet*. 1977;41(2):225-233.

90. Wang Q, Fu L, Zhang X, et al. Expansion of a SNaPshot assay to a 55-SNP multiplex: assay enhancements, validation, and power in forensic science. *Electrophoresis*. 2016;37(10):1310-1317.

91. Sanchez JJ, Phillips C, Børsting C, et al. A multiplex assay with 52 single nucleotide polymorphisms for human identification. *Electrophoresis*. 2006;27(9):1713-1724.

92. Børsting C, Fordyce SL, Olofsson J, Mogensen HS, Morling N. Evaluation of the Ion Torrent™ HID SNP 169-plex: A SNP typing assay developed for human identification by

second generation sequencing. *Forensic Sci Int Genet*. 2014;12:144-154.

93. Ge J, Budowle B, Planz J V., Chakraborty R. Haplotype block: a new type of forensic DNA markers. *Int J Legal Med*. 2010;124(5):353-361.

94. Phillips C, Prieto L, Fondevila M, et al. Ancestry Analysis in the 11-M Madrid Bomb Attack Investigation. *PLoS One*. 2009;4(8):e6583.

95. Kayser M. Forensic use of Y-chromosome DNA: a general overview. *Hum Genet*. 2017;136(5):621-635.

96. Shriver MD, Parra EJ, Dios S, et al. Skin pigmentation, biogeographical ancestry and admixture mapping. *Hum Genet*. 2003;112(4):387-399.

97. Galanter JM, Fernandez-Lopez JC, Gignoux CR, et al. Development of a Panel of Genome-Wide Ancestry Informative Markers to Study Admixture Throughout the Americas. *PLoS Genet*. 2012;8(3):e1002554.

98. Santos C, Phillips C, Fondevila M, et al. Pacifiplex : an ancestry-informative SNP panel centred on Australia and the Pacific region. *Forensic Sci Int Genet*. 2016;20:71-80.

99. Phillips C, Aradas AF, Kriegel AK, et al. Eurasiaplex: A forensic SNP assay for differentiating European and South Asian ancestries. *Forensic Sci Int Genet*. 2013;7(3):359-366.

100. Phillips C, Parson W, Lundsberg B, et al. Building a forensic ancestry panel from the ground up: The EUROFORGEN Global AIM-SNP set. *Forensic Sci Int Genet*. 2014;11:13-25.

101. Silventoinen K, Sammalisto S, Perola M, et al. Heritability of Adult Body Height: A Comparative Study of Twin Cohorts in Eight Countries. *Twin Res*. 2003;6(5):399-408.

102. Clark P, Stark AE, Walsh RJ, Jardine R, Martin NG. A twin study of skin reflectance. *Ann Hum Biol*. 1981;8(6):529-541.

103. Sturm RA, Frudakis TN. Eye colour: portals into pigmentation genes and ancestry. *Trends Genet*. 2004;20(8):327-332.

104. Hart KL, Kimura SL, Mushailov V, Budimlija ZM, Prinz M, Wurmbach E. Improved eye- and skin-color prediction based on 8 SNPs. *Croat Med J*. 2013;54(3):248-256.

105. Walsh S, Lindenbergh A, Zuniga SB, et al. Developmental validation of the IrisPlex system: Determination of blue and brown iris colour for forensic intelligence. *Forensic Sci Int Genet*. 2011;5(5):464-471.

106. Walsh S, Liu F, Wollstein A, et al. The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA. *Forensic Sci Int Genet*. 2013;7(1):98-115.

107. Amigo J, Phillips C, Lareu M, Carracedo Á. The SNPforID browser: an online tool for query and display of frequency data from the SNPforID project. *Int J Legal Med*. 2008;122(5):435-440.

108. Jacobs PA, Strong JA. A case of human intersexuality having a possible XXY sex-determining mechanism. *Nature*. 1959;183(4657):302-303.

109. Tiepolo L, Zuffardi O. Localization of factors controlling spermatogenesis in the nonfluorescent portion of the human Y chromosome long arm. *Hum Genet*. 1976;34(2):119-124.

110. Jobling MA, Tyler-Smith C. Human Y-chromosome variation in the genome-sequencing era. *Nat Rev Genet*. 2017;18(8):485-497.

111. Xue Y, Tyler-Smith C. Past successes and future opportunities for the genetics of the human Y chromosome. *Hum Genet*. 2017;136(5):481-483.

112. Cortez D, Marin R, Toledo-Flores D, et al. Origins and functional evolution of Y chromosomes across mammals. *Nature*. 2014;508(7497):488-493.

113. Bellott DW, Hughes JF, Skaletsky H, et al. Mammalian Y chromosomes retain widely expressed dosage-sensitive regulators. *Nature*. 2014;508(7497):494-499.

114. Treangen TJ, Salzberg SL. Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nat Rev Genet*. 2012;13(1):36-46.

115. Hallast P, Jobling MA. The Y chromosomes of the great apes. *Hum Genet*. 2017;136(5):511-528.

116. Gubbay J, Collignon J, Koopman P, et al. A gene mapping to the sex-determining region of the mouse Y chromosome is a member of a novel family of embryonically expressed genes. *Nature*. 1990;346(6281):245-250.

117. Sinclair AH, Berta P, Palmer MS, et al. A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. *Nature*. 1990;346(6281):240-244.

118. Skaletsky H, Kuroda-Kawaguchi T, Minx PJ, et al. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature*. 2003;423(6942):825-837.

119. Burgoyne PS. Genetic homology and crossing over in the X and Y chromosomes of mammals. *Hum Genet*. 1982;61(2):85-90.

120. Polani PE. Pairing of X and Y chromosomes, non-inactivation of X-linked genes, and the maleness factor. *Hum Genet*. 1982;60(3):207-211.

121. Freije D, Helms C, Watson M, Donis-Keller H. Identification of a second pseudoautosomal region near the Xq and Yq telomeres. *Science*. 1992;258(5089):1784-1787.

122. Rappold GA. The pseudoautosomal regions of the human sex chromosomes. *Hum Genet*. 1993;92(4):315-324.

123. Graves JA, Wakefield MJ, Toder R. The origin and evolution of the pseudoautosomal regions of human sex chromosomes. *Hum Mol Genet*. 1998;7(13):1991-1996.

124. Helena Mangs A, Morris B. The Human Pseudoautosomal Region (PAR): Origin, Function and Future. *Curr Genomics*. 2007;8(2):129-136.

125. Rozen S, Skaletsky H, Marszalek JD, et al. Abundant gene conversion between arms of palindromes in human and ape Y chromosomes. *Nature*. 2003;423(6942):873-876.

126. Butler JM, Schoske R, Vallone PM, Kline MC, Redd AJ, Hammer MF. A novel multiplex for simultaneous amplification of 20 Y chromosome STR markers. *Forensic Sci Int*. 2002;129(1):10-24.

127. Hall A, Ballantyne J. Strategies for the Design and Assessment of Y-Short Tandem Repeat Multiplexes for Forensic Use. *Forensic Sci Rev*. 2003;15(2):137-149.

128. Page DC, Harper ME, Love J, Botstein D. Occurrence of a transposition from the X-chromosome long arm to the Y-chromosome short arm during human evolution. *Nature*. 1984;311(5982):119-123.

129. Poznik GD, Henn BM, Yee M-C, et al. Sequencing Y Chromosomes Resolves Discrepancy in Time to Common Ancestor of Males Versus Females. *Science*. 2013;341(6145):562-565.

130. Vogt PH, Edelmann A, Kirsch S, et al. Human Y chromosome azoospermia factors (AZF) mapped to different subregions in Yq11. *Hum Mol Genet*. 1996;5(7):933-943.

131. Repping S, Skaletsky H, Brown L, et al. Polymorphism for a 1.6-Mb deletion of the human Y chromosome persists through balance between recurrent mutation and haploid selection. *Nat Genet*. 2003;35(3):247-251.

132. Rosser ZH, Balaresque P, Jobling MA. Gene Conversion between the X Chromosome and the Male-Specific Region of the Y Chromosome at a Translocation Hotspot. *Am J Hum Genet*. 2009;85(1):130-134.

133. Hallast P, Balaresque P, Bowden GR, Ballereau S, Jobling MA. Recombination Dynamics of a Human Y-Chromosomal Palindrome: Rapid GC-Biased Gene Conversion, Multi-kilobase Conversion Tracts, and Rare Inversions. *PLoS Genet*. 2013;9(7):e1003666.

134. Balaresque P, King TE, Parkin EJ, et al. Gene Conversion Violates the Stepwise Mutation Model for Microsatellites in Y-Chromosomal Palindromic Repeats. *Hum Mutat*. 2014;35(5):609-617.

135. Trombetta B, Cruciani F, Underhill PA, Sellitto D, Scozzari R. Footprints of X-to-Y Gene Conversion in Recent Human Evolution. *Mol Biol Evol*. 2010;27(3):714-725.

136. Trombetta B, Sellitto D, Scozzari R, Cruciani F. Inter- and Intraspecies Phylogenetic Analyses Reveal Extensive X–Y Gene Conversion in the Evolution of Gametologous Sequences of Human Sex Chromosomes. *Mol Biol Evol*. 2014;31(8):2108-2123.

137. Trombetta B, Cruciani F. Y chromosome palindromes and gene conversion. *Hum Genet*. 2017;136(5):605-619.

138. Hammer MF. A recent insertion of an alu element on the Y chromosome is a useful marker for human population studies. *Mol Biol Evol*. 1994;11(5):749-761.

139. Gusmão L, Brion M, González-Neira A, Lareu M, Carracedo A. Y chromosome specific polymorphisms in forensic analysis. *Leg Med*. 1999;1(2):55-60.

140. Jobling MA, Tyler-Smith C. The human Y chromosome: an evolutionary marker comes of age. *Nat Rev Genet*. 2003;4(8):598-612.

141. Kittler R, Erler A, Brauer S, Stoneking M, Kayser M. Apparent intrachromosomal exchange on the human Y chromosome explained by population history. *Eur J Hum Genet*. 2003;11(4):304-314.

142. Redd AJ, Agellon AB, Kearney VA, et al. Forensic value of 14 novel STRs on the human Y chromosome. *Forensic Sci Int*. 2002;130(2-3):97-111.

143. Butler JM. Recent Developments in Y-Short Tandem Repeat and Y-Single Nucleotide Polymorphism Analysis. *Forensic Sci Rev*. 2003;15(2):91-111.

144. Schoske R. High-throughput Y-STR typing of U.S. populations with 27 regions of the Y

chromosome using two multiplex PCR assays. *Forensic Sci Int*. 2004;139(2-3):107-121.

145. Hanson EK, Ballantyne J. A highly discriminating 21 locus Y-STR "megaplex" system designed to augment the minimal haplotype loci for forensic casework. *J Forensic Sci*. 2004;49(1):40-51.

146. Hanson EK, Ballantyne J. An Ultra-High Discrimination Y Chromosome Short Tandem Repeat Multiplex DNA Typing System. *PLoS One*. 2007;2(8):e688.

147. Kayser M, Caglià A, Corach D, et al. Evaluation of Y-chromosomal STRs: a multicenter study. *Int J Legal Med*. 1997;110(3):125-133.

148. Vermeulen M, Wollstein A, van der Gaag K, et al. Improving global and regional resolution of male lineage differentiation by simple single-copy Y-chromosomal short tandem repeat polymorphisms. *Forensic Sci Int Genet*. 2009;3(4):205-213.

149. Ballantyne KN, Goedbloed M, Fang R, et al. Mutability of Y-Chromosomal Microsatellites: Rates, Characteristics, Molecular Bases, and Forensic Implications. *Am J Hum Genet*. 2010;87(3):341-353.

150. Ballantyne KN, Keerl V, Wollstein A, et al. A new future of forensic Y-chromosome analysis: Rapidly mutating Y-STRs for differentiating male relatives and paternal lineages. *Forensic Sci Int Genet*. 2012;6(2):208-218.

151. Ballantyne KN, Ralf A, Aboukhalid R, et al. Toward Male Individualization with Rapidly Mutating Y-Chromosomal Short Tandem Repeats. *Hum Mutat*. 2014;35(8):1021-1032.

152. Pascali VL, Dobosz M, Brinkmann B. Coordinating Y-chromosomal STR research for the Courts. *Int J Legal Med*. 1998;112(1):1-1.

153. Schneider PM, Meuser S, Waiyawuth W, Seo Y, Rittner C. Tandem repeat structure of the duplicated Y-chromosomal STR locus DYS385 and frequency studies in the German and three Asian populations. *Forensic Sci Int*. 1998;97(1):61-70.

154. Ayub Q. Identification and characterisation of novel human Y-chromosomal microsatellites from sequence database information. *Nucleic Acids Res*. 2000;28(2):8e-8.

155. Núñez C, Baeta M, Ibarbia N, et al. 17 to 23: A novel complementary mini Y-STR panel to extend the Y-STR databases from 17 to 23 markers for forensic purposes. *Electrophoresis*. 2017;38(7):1016-1021.

156. Alghafri R, Goodwin W, Hadi S. Rapidly mutating Y-STRs multiplex genotyping panel to

investigate UAE population. *Forensic Sci Int Genet Suppl Ser*. 2013;4(1):e200-e201.

157. Lim S-K, Xue Y, Parkin EJ, Tyler-Smith C. Variation of 52 new Y-STR loci in the Y Chromosome Consortium worldwide panel of 76 diverse individuals. *Int J Legal Med*. 2007;121(2):124-127.

158. Jacobs M, Janssen L, Vanderheyden N, Bekaert B, Van de Voorde W, Decorte R. Development and evaluation of multiplex Y-STR assays for application in molecular genealogy. *Forensic Sci Int Genet Suppl Ser*. 2009;2(1):57-59.

159. Calafell F, Larmuseau MHD. The Y chromosome as the most popular marker in genetic genealogy benefits interdisciplinary research. *Hum Genet*. 2017;136(5):559-573.

160. Larmuseau MHD, Ottoni C. Mediterranean Y-chromosome 2.0—why the Y in the Mediterranean is still relevant in the postgenomic era. *Ann Hum Biol*. 2018;45(1):20-33.

161. Underhill PA, Shen P, Lin AA, et al. Y chromosome sequence variation and the history of human populations. *Nat Genet*. 2000;26:358-361.

162. Athey W. Haplogroup Prediction from Y-STR Values Using a Bayesian-Allele- Frequency Approach. *J Genet Geneal*. 2006;2:34-39.

163. Núñez C, Geppert M, Baeta M, Roewer L, Martínez-Jarreta B. Y chromosome haplogroup diversity in a Mestizo population of Nicaragua. *Forensic Sci Int Genet*. 2012;6(6):e192-e195.

164. 1000 Genomes Project Consortium, Auton A, Brooks LD, et al. A global reference for human genetic variation. *Nature*. 2015;526(7571):68-74.

165. Poznik GD, Xue Y, Mendez FL, et al. Punctuated bursts in human male demography inferred from 1,244 worldwide Y-chromosome sequences. *Nat Genet*. 2016;48(6):593-599.

166. Scozzari R, Massaia A, Trombetta B, et al. An unbiased resource of novel SNP markers provides a new chronology for the human Y chromosome and reveals a deep phylogenetic structure in Africa. *Genome Res*. 2014;24(3):535-544.

167. Hallast P, Batini C, Zadik D, et al. The Y-Chromosome Tree Bursts into Leaf: 13,000 High-Confidence SNPs Covering the Majority of Known Clades. *Mol Biol Evol*. 2015;32(3):661-673.

168. Batini C, Hallast P, Zadik D, et al. Large-scale recent expansion of European patrilineages

shown by population resequencing. *Nat Commun*. 2015;6(1):7152.

169. Trombetta B, D'Atanasio E, Massaia A, et al. Phylogeographic Refinement and Large Scale Genotyping of Human Y Chromosome Haplogroup E Provide New Insights into the Dispersal of Early Pastoralists in the African Continent. *Genome Biol Evol*. 2015;7(7):1940-1950.

170. Y Chromosome Consortium. A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Res*. 2002;12(2):339-348.

171. van Oven M, Van Geystelen A, Kayser M, Decorte R, Larmuseau MH. Seeing the Wood for the Trees: A Minimal Reference Phylogeny for the Human Y Chromosome. *Hum Mutat*. 2014;35(2):187-191.

172. Fondevila M, Børsting C, Phillips C, et al. Forensic SNP genotyping with SNaPshot: Technical considerations for the development and optimization of multiplexed SNP assays. *Forensic Sci Rev*. 2017;29(1):57-76.

173. Sokolov BP. Primer extension technique for the detection of single nucleotide in genomic DNA. *Nucleic Acids Res*. 1990;18(12):3671.

174. Syvänen A-C, Aalto-Setälä K, Harju L, Kontula K, Söderlund H. A primer-guided nucleotide incorporation assay in the genotyping of apolipoprotein E. *Genomics*. 1990;8(4):684-692.

175. Morley JM, Bark JE, Evans CE, Perry JG, Hewitt CA, Tully G. Validation of mitochondrial DNA minisequencing for forensic casework. *Int J Legal Med*. 1999;112(4):241-248.

176. Brión M, Sanchez JJ, Balogh K, et al. Introduction of an single nucleodite polymorphism-based "Major Y-chromosome haplogroup typing kit" suitable for predicting the geographical origin of male lineages. *Electrophoresis*. 2005;26(23):4411-4420.

177. van Oven M, Ralf A, Kayser M. An efficient multiplex genotyping approach for detecting the major worldwide human Y-chromosome haplogroups. *Int J Legal Med*. 2011;125(6):879-885.

178. Geppert M, Baeta M, Núñez C, et al. Hierarchical Y-SNP assay to study the hidden diversity and phylogenetic relationship of native populations in South America. *Forensic Sci Int Genet*. 2011;5(2):100-104.

179. Valverde L, Köhnemann S, Cardoso S, Pfeiffer H, de Pancorbo MM. Improving the analysis of Y-SNP haplogroups by a single highly informative 16 SNP multiplex PCR-

minisequencing assay. *Electrophoresis*. 2013;34(4):605-612.

180. Onofri V, Alessandrini F, Turchi C, Pesaresi M, Buscemi L, Tagliabracci A. Development of multiplex PCRs for evolutionary and forensic applications of 37 human Y chromosome SNPs. *Forensic Sci Int*. 2006;157(1):23-35.

181. Park MJ, Lee HY, Kim NY, Lee EY, Yang WI, Shin K-J. Y-SNP miniplexes for East Asian Y-chromosomal haplogroup determination in degraded DNA. *Forensic Sci Int Genet*. 2013;7(1):75-81.

182. van Oven M, van den Tempel N, Kayser M. A multiplex SNP assay for the dissection of human Y-chromosome haplogroup O representing the major paternal lineage in East and Southeast Asia. *J Hum Genet*. 2012;57(1):65-69.

183. Bouakaze C, Keyser C, Amory S, Crubézy E, Ludes B. First successful assay of Y-SNP typing by SNaPshot minisequencing on ancient DNA. *Int J Legal Med*. 2007;121(6):493-499.

184. Mehta B, Daniel R, McNevin D. High resolution melting (HRM) of forensically informative SNPs. *Forensic Sci Int Genet Suppl Ser*. 2013;4(1):e376-e377.

185. Venables SJ, Mehta B, Daniel R, Walsh SJ, van Oorschot RAH, McNevin D. Assessment of high resolution melting analysis as a potential SNP genotyping technique in forensic casework. *Electrophoresis*. 2014;35(21-22):3036-3043.

186. Mehta B, Daniel R, McNevin D. HRM and SNaPshot as alternative forensic SNP genotyping methods. *Forensic Sci Med Pathol*. 2017;13(3):293-301.

187. Mao F, Leung W-Y, Xin X. Characterization of EvaGreen and the implication of its physicochemical properties for qPCR applications. *BMC Biotechnol*. 2007;7(1):76.

188. Wittwer CT, Reed GH, Gundry CN, Vandersteen JG, Pryor RJ. High-resolution genotyping by amplicon melting analysis using LCGreen. *Clin Chem*. 2003;49(6 Pt 1):853-860.

189. Ririe KM, Rasmussen RP, Wittwer CT. Product Differentiation by Analysis of DNA Melting Curves during the Polymerase Chain Reaction. *Anal Biochem*. 1997;245(2):154-160.

190. Akey JM, Sosnoski D, Parra E, et al. Melting curve analysis of SNPs (McSNP): a gel-free and inexpensive approach for SNP genotyping. *Biotechniques*. 2001;30(2):358-362, 364, 366-367.

191. Giglio S, Monis PT, Saint CP. Demonstration of preferential binding of SYBR Green I to specific DNA fragments in real-time multiplex PCR. *Nucleic Acids Res*. 2003;31(22):e136.

192. Dobrowolski SF, Gray J, Miller T, Sears M. Identifying sequence variants in the human mitochondrial genome using high-resolution melt (HRM) profiling. *Hum Mutat*. 2009;30(6):891-898.

193. Taylor CF. Mutation scanning using high-resolution melting. *Biochem Soc Trans*. 2009;37(2):433-437.

194. Liew M, Pryor R, Palais R, et al. Genotyping of single-nucleotide polymorphisms by high-resolution melting of small amplicons. *Clin Chem*. 2004;50(7):1156-1164.

195. Erali M, Voelkerding K V., Wittwer CT. High resolution melting applications for clinical laboratory medicine. *Exp Mol Pathol*. 2008;85(1):50-58.

196. Gundry CN, Dobrowolski SF, Martin YR, et al. Base-pair neutral homozygotes can be discriminated by calibrated high-resolution melting of small amplicons. *Nucleic Acids Res*. 2008;36(10):3401-3408.

197. Vossen RHAM, Aten E, Roos A, den Dunnen JT. High-resolution melting analysis (HRMA): more than just sequence variant screening. *Hum Mutat*. 2009;30(6):860-866.

198. Gundry CN. Amplicon Melting Analysis with Labeled Primers: A Closed-Tube Method for Differentiating Homozygotes and Heterozygotes. *Clin Chem*. 2003;49(3):396-406.

199. Seipp MT, Durtschi JD, Voelkerding K V., Wittwer CT. Multiplex amplicon genotyping by high-resolution melting. *J Biomol Tech*. 2009;20(3):160-164.

200. Geyer CN, Hanson ND. Multiplex high-resolution melting analysis as a diagnostic tool for detection of plasmid-mediated AmpC β-lactamase genes. *J Clin Microbiol*. 2014;52(4):1262-1265.

201. Holland PM, Abramson RD, Watson R, Gelfand DH. Detection of specific polymerase chain reaction product by utilizing the 5'----3' exonuclease activity of Thermus aquaticus DNA polymerase. *Proc Natl Acad Sci*. 1991;88(16):7276-7280.

202. Livak KJ. Allelic discrimination using fluorogenic probes and the 5' nuclease assay. *Genet Anal Biomol Eng*. 1999;14(5-6):143-149.

203. Alonso A, Garcia O. Real-time quantitative PCR in Forensic Science. In: *Molecular Forensics*. Willey; 2007:59-67.

204. Bass C, Nikou D, Vontas J, Donnelly MJ, Williamson MS, Field LM. The Vector Population Monitoring Tool (VPMT): High-Throughput DNA-Based Diagnostics for the Monitoring of

Mosquito Vector Populations. *Malar Res Treat*. 2010;2010:1-8.

205. Louhelainen J. SNP Arrays. *Microarrays*. 2016;5(4):27.

206. Heather JM, Chain B. The sequence of sequencers: The history of sequencing DNA. *Genomics*. 2016;107(1):1-8.

207. Ranjan K, Minakshi P, Gaya P. Application of Molecular and Serological Diagnostics in Veterinary Parasitology. *J Adv Parasitol*. 2015;2(4):80-99.

208. Jäger AC, Alvarez ML, Davis CP, et al. Developmental validation of the MiSeq FGx Forensic Genomics System for Targeted Next Generation Sequencing in Forensic DNA Casework and Database Laboratories. *Forensic Sci Int Genet*. 2017;28:52-70.

209. Guo F, Yu J, Zhang L, Li J. Massively parallel sequencing of forensic STRs and SNPs using the Illumina ® ForenSeq™ DNA Signature Prep Kit on the MiSeq FGx™ Forensic Genomics System. *Forensic Sci Int Genet*. 2017;31:135-148.

210. Meiklejohn KA, Robertson JM. Evaluation of the Precision ID Identity Panel for the Ion Torrent ™ PGM ™ sequencer. *Forensic Sci Int Genet*. 2017;31:48-56.

211. Alonso A, Müller P, Roewer L, Willuweit S, Budowle B, Parson W. European survey on forensic applications of massively parallel sequencing. *Forensic Sci Int Genet*. 2017;29:e23-e25.

212. Almalki N, Chow HY, Sharma V, Hart K, Siegel D, Wurmbach E. Systematic assessment of the performance of Illumina's MiSeq FGx™ forensic genomics system. *Electrophoresis*. 2017;38(6):846-854.

213. Churchill JD, Schmedes SE, King JL, Budowle B. Evaluation of the Illumina ® Beta Version ForenSeq™ DNA Signature Prep Kit for use in genetic profiling. *Forensic Sci Int Genet*. 2016;20:20-29.

214. Just RS, Moreno LI, Smerick JB, Irwin JA. Performance and concordance of the ForenSeq™ system for autosomal and Y chromosome short tandem repeat sequencing of reference-type specimens. *Forensic Sci Int Genet*. 2017;28:1-9.

215. Silvia AL, Shugarts N, Smith J. A preliminary assessment of the ForenSeq™ FGx System: next generation sequencing of an STR and SNP multiplex. *Int J Legal Med*. 2017;131(1):73-86.

216. Xavier C, Parson W. Evaluation of the Illumina ForenSeq™ DNA Signature Prep Kit – MPS

forensic application for the MiSeq FGx™ benchtop sequencer. *Forensic Sci Int Genet*. 2017;28:188-194.

217. Eduardoff M, Santos C, de la Puente M, et al. Inter-laboratory evaluation of SNP-based forensic identification by massively parallel sequencing using the Ion PGM™. *Forensic Sci Int Genet*. 2015;17:110-121.

218. Elena S, Alessandro A, Ignazio C, Sharon W, Luigi R, Andrea B. Revealing the challenges of low template DNA analysis with the prototype Ion AmpliSeq™ Identity panel v2.3 on the PGM™ Sequencer. *Forensic Sci Int Genet*. 2016;22:25-36.

219. Guo F, Zhou Y, Song H, et al. Next generation sequencing of SNPs using the HID-Ion AmpliSeq™ Identity Panel on the Ion Torrent PGM™ platform. *Forensic Sci Int Genet*. 2016;25:73-84.

220. Churchill JD, Chang J, Ge J, et al. Blind study evaluation illustrates utility of the Ion PGM™ system for use in human identity DNA typing. *Croat Med J*. 2015;56(3):218-229.

221. Zhao X, Li H, Wang Z, Ma K, Cao Y, Liu W. Massively parallel sequencing of 10 autosomal STRs in Chinese using the ion torrent personal genome machine (PGM). *Forensic Sci Int Genet*. 2016;25:34-38.

222. de Knijff P. From next generation sequencing to now generation sequencing in forensics. *Forensic Sci Int Genet*. 2019;38:175-180.

223. Willuweit S, Roewer L. The new Y Chromosome Haplotype Reference Database. *Forensic Sci Int Genet*. 2015;15:43-48.

224. Rocca RA, Magoon G, Reynolds DF, et al. Discovery of Western European R1b1a2 Y Chromosome Variants in 1000 Genomes Project Data: An Online Community Approach. *PLoS One*. 2012;7(7):e41634.

225. Underhill PA, Kivisild T. Use of Y Chromosome and Mitochondrial DNA Population Structure in Tracing Human Migrations. *Annu Rev Genet*. 2007;41(1):539-564.

226. Prinz M, Sansone M. Y chromosome-specific short tandem repeats in forensic casework. *Croat Med J*. 2001;42(3):288-291.

227. Larmuseau MHD, Van Geystelen A, van Oven M, Decorte R. Genetic genealogy comes of age: Perspectives on the use of deep-rooted pedigrees in human population genetics. *Am J Phys Anthropol*. 2013;150(4):505-511.

228. Larmuseau MH, Ottoni C, Raeymaekers JA, Vanderheyden N, Larmuseau HF, Decorte R. Temporal differentiation across a West-European Y-chromosomal cline: genealogy as a tool in human population genetics. *Eur J Hum Genet*. 2012;20(4):434-440.

229. Wilkins JF. Unraveling male and female histories from human genetic data. *Curr Opin Genet Dev*. 2006;16(6):611-617.

230. Lippold S, Xu H, Ko A, et al. Human paternal and maternal demographic histories: insights from high-resolution Y chromosome and mtDNA sequences. *Investig Genet*. 2014;5(1):13.

231. Karmin M, Saag L, Vicente M, et al. A recent bottleneck of Y chromosome diversity coincides with a global change in culture. *Genome Res*. 2015;25(4):459-466.

232. Zerjal T, Dashnyam B, Pandya A, et al. Genetic relationships of Asians and Northern Europeans, revealed by Y-chromosomal DNA analysis. *Am J Hum Genet*. 1997;60(5):1174-1183.

233. Wei W, Ayub Q, Xue Y, Tyler-Smith C. A comparison of Y-chromosomal lineage dating using either resequencing or Y-SNP plus Y-STR genotyping. *Forensic Sci Int Genet*. 2013;7(6):568-572.

234. Hughes JF, Page DC. The Biology and Evolution of Mammalian Y Chromosomes. *Annu Rev Genet*. 2015;49(1):507-527.

235. Mensah MA, Hestand MS, Larmuseau MHD, et al. Pseudoautosomal Region 1 Length Polymorphism in the Human Population. *PLoS Genet*. 2014;10(11):e1004578.

236. van Geystelen A, Wenseleers T, Decorte R, Caspers MJL, Larmuseau MHD. In silico detection of phylogenetic informative Y-chromosomal single nucleotide polymorphisms from whole genome sequencing data. *Electrophoresis*. 2014;35(21-22):3102-3110.

237. Sykes B, Irven C. Surnames and the Y Chromosome. *Am J Hum Genet*. 2000;66(4):1417-1419.

238. King TE, Jobling MA. Founders, Drift, and Infidelity: The Relationship between Y Chromosome Diversity and Patrilineal Surnames. *Mol Biol Evol*. 2009;26(5):1093-1102.

239. Martinez-Cadenas C, Blanco-Verea A, Hernando B, et al. The relationship between surname frequency and Y chromosome variation in Spain. *Eur J Hum Genet*. 2016;24(1):120-128.

240. Solé-Morata N, Bertranpetit J, Comas D, Calafell F. Y-chromosome diversity in Catalan

surname samples: insights into surname origin and frequency. *Eur J Hum Genet*. 2015;23(11):1549-1557.

241. McEvoy B, Bradley DG. Y-chromosomes and the extent of patrilineal ancestry in Irish surnames. *Hum Genet*. 2006;119(1-2):212-219.

242. Larmuseau MHD, Matthijs K, Wenseleers T. Cuckolded Fathers Rare in Human Populations. *Trends Ecol Evol*. 2016;31(5):327-329.

243. Hayward AD, Lummaa V, Bazykin GA. Fitness Consequences of Advanced Ancestral Age over Three Generations in Humans. *PLoS One*. 2015;10(6):e0128197.

244. Bolund E, Hayward A, Pettay JE, Lummaa V. Effects of the demographic transition on the genetic variances and covariances of human life-history traits. *Evolution*. 2015;69(3):747-755.

245. Pinhasi R, Thomas MG, Hofreiter M, Currat M, Burger J. The genetic history of Europeans. *Trends Genet*. 2012;28(10):496-505.

246. Lacan M, Keyser C, Crubézy E, Ludes B. Ancestry of modern Europeans: contributions of ancient DNA. *Cell Mol Life Sci*. 2013;70(14):2473-2487.

247. Semino O, Passarino G, Oefner PJ, et al. The genetic legacy of Paleolithic Homo sapiens sapiens in extant Europeans: a Y chromosome perspective. *Science*. 2000;290(5494):1155-1159.

248. Menozzi P, Piazza A, Cavalli-Sforza L. Synthetic maps of human gene frequencies in Europeans. *Science*. 1978;201(4358):786-792.

249. Cavalli-Sforza LL, Menozzi P, Piazza A. *The History and Geography of Human Genes*. Princeton: Princeton University Press; 1994.

250. Chikhi L, Destro-Bisol G, Bertorelle G, Pascali V, Barbujani G. Clines of nuclear DNA markers suggest a largely Neolithic ancestry of the European gene pool. *Proc Natl Acad Sci*. 1998;95(15):9053-9058.

251. Belle EMS, Landry P-A, Barbujani G. Origins and evolution of the Europeans' genome: evidence from multiple microsatellite loci. *Proc R Soc B Biol Sci*. 2006;273(1594):1595-1602.

252. Rosser ZH, Zerjal T, Hurles ME, et al. Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language. *Am J Hum Genet*.

2000;67:1526-1543.

253. Underhill PA, Poznik GD, Rootsi S, et al. The phylogenetic and geographic structure of Y-chromosome haplogroup R1a. *Eur J Hum Genet*. 2015.

254. Kayser M, Lao O, Anslinger K, et al. Significant genetic differentiation between Poland and Germany follows present-day political borders, as revealed by Y-chromosome analysis. *Hum Genet*. 2005;117(5):428-443.

255. Underhill P. A. Inferring Human History: Clues from Y-Chromosome Haplotypes. *Cold Spring Harb Symp Quant Biol*. 2003;68:487-494.

256. Alonso S, Flores C, Cabrera V, et al. The place of the Basques in the European Y-chromosome diversity landscape. *Eur J Hum Genet*. 2005;13(12):1293-1302.

257. Balaresque P, Bowden GR, Adams SM, et al. A Predominantly Neolithic Origin for European Paternal Lineages. *PLoS Biol*. 2010;8(1):e1000285.

258. Myres NM, Rootsi S, Lin AA, et al. A major Y-chromosome haplogroup R1b Holocene era founder effect in Central and Western Europe. *Eur J Hum Genet*. 2011;19(1):95-101.

259. Busby GBJ, Brisighelli F, Sánchez-Diz P, et al. The peopling of Europe and the cautionary tale of Y chromosome lineage R-M269. *Proc Biol Sci*. 2012;279(1730):884-892.

260. Cruciani F, Trombetta B, Antonelli C, et al. Strong intra- and inter-continental differentiation revealed by Y chromosome SNPs M269, U106 and U152. *Forensic Sci Int Genet*. 2011;5(3):e49-e52.

261. Semino O, Magri C, Benuzzi G, et al. Origin, Diffusion, and Differentiation of Y-Chromosome Haplogroups E and J: Inferences on the Neolithization of Europe and Later Migratory Events in the Mediterranean Area. *Am J Hum Genet*. 2004;74(5):1023-1034.

262. Rootsi S, Kivisild T, Benuzzi G, et al. Phylogeography of Y-Chromosome Haplogroup I Reveals Distinct Domains of Prehistoric Gene Flow in Europe. *Am J Hum Genet*. 2004;75(1):128-137.

263. Pakendorf B, Stoneking M. Mitochondrial DNA and human evolution. *Annu Rev Genomics Hum Genet*. 2005;6(1):165-183.

264. Haak W, Brandt G, Jong HN d., et al. Ancient DNA, Strontium isotopes, and osteological analyses shed light on social and kinship organization of the Later Stone Age. *Proc Natl Acad Sci*. 2008;105(47):18226-18231.

265. Skoglund P, Malmstrom H, Raghavan M, et al. Origins and Genetic Legacy of Neolithic Farmers and Hunter-Gatherers in Europe. *Science*. 2012;336(6080):466-469.

266. Brandt G, Szécsényi-Nagy A, Roth C, Alt KW, Haak W. Human paleogenetics of Europe – The known knowns and the known unknowns. *J Hum Evol*. 2015;79:73-92.

267. Lazaridis I, Patterson N, Mittnik A, et al. Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature*. 2014;513(7518):409-413.

268. Fu Q, Posth C, Hajdinjak M, et al. The genetic history of Ice Age Europe. *Nature*. 2016;534(7606):200-205.

269. Kivisild T. The study of human Y chromosome variation through ancient DNA. *Hum Genet*. 2017;136(5):529-546.

270. Bramanti B, Thomas MG, Haak W, et al. Genetic Discontinuity Between Local Hunter-Gatherers and Central Europe's First Farmers. *Science*. 2009;326(5949):137-140.

271. Krause J, Briggs AW, Kircher M, et al. A Complete mtDNA Genome of an Early Modern Human from Kostenki, Russia. *Curr Biol*. 2010;20(3):231-236.

272. Hervella M, Izagirre N, Alonso S, et al. Ancient DNA from Hunter-Gatherer and Farmer Groups from Northern Spain Supports a Random Dispersion Model for the Neolithic Expansion into Europe. *PLoS One*. 2012;7(4):e34417.

273. Sánchez-Quinto F, Schroeder H, Ramirez O, et al. Genomic Affinities of Two 7,000-Year-Old Iberian Hunter-Gatherers. *Curr Biol*. 2012;22(16):1494-1499.

274. Fu Q, Mittnik A, Johnson PLF, et al. A Revised Timescale for Human Evolution Based on Ancient Mitochondrial Genomes. *Curr Biol*. 2013;23(7):553-559.

275. Olalde I, Allentoft ME, Sánchez-Quinto F, et al. Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European. *Nature*. 2014;507(7491):225-228.

276. Raghavan M, Skoglund P, Graf KE, et al. Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature*. 2014;505(7481):87-91.

277. Whittle AWR. Europe in The Neolithic : The Creation Of New Worlds. *Cambridge*. 1996.

278. T. Douglas Price, Ruth Tringham, Marek Zvelebil, Malcolm Lillie, William K. Barnett, Didier Binder, João Zilhão, Michael Jochim, Peter Bogucki PW. *Europe's First Farmers*.

Cambridge University Press; 2000.

279. Whittle AWR, Cummings V. *Going Over: The Mesolithic-Neolithic Transition in North-West Europe*. Oxford: Oxford University Press; 2007.

280. Haak W, Balanovsky O, Sanchez JJ, et al. Ancient DNA from European Early Neolithic Farmers Reveals Their Near Eastern Affinities. *PLoS Biol*. 2010;8(11):e1000536.

281. Lacan M, Keyser C, Ricaut F-X, et al. Ancient DNA reveals male diffusion through the Neolithic Mediterranean route. *Proc Natl Acad Sci*. 2011;108(24):9788-9791.

282. Keller A, Graefen A, Ball M, et al. New insights into the Tyrolean Iceman's origin and phenotype as inferred by whole-genome sequencing. *Nat Commun*. 2012;3(1):698.

283. Szecsenyi-Nagy A, Brandt G, Haak W, et al. Tracing the genetic origin of Europe's first farmers reveals insights into their social organization. *Proc R Soc B Biol Sci*. 2015;282(1805):20150339-20150339.

284. Lacan M, Keyser C, Ricaut F-X, et al. Ancient DNA suggests the leading role played by men in the Neolithic dissemination. *Proc Natl Acad Sci*. 2011;108(45):18255-18259.

285. Gamba C, Jones ER, Teasdale MD, et al. Genome flux and stasis in a five millennium transect of European prehistory. *Nat Commun*. 2014;5(1):5257.

286. Mathieson I, Lazaridis I, Rohland N, et al. Genome-wide patterns of selection in 230 ancient Eurasians. *Nature*. 2015;528(7583):499-503.

287. Gamba C, Fernández E, Tirado M, et al. Ancient DNA from an Early Neolithic Iberian population supports a pioneer colonization by first farmers. *Mol Ecol*. 2012;21(1):45-56.

288. Lee EJ, Makarewicz C, Renneberg R, et al. Emerging genetic patterns of the european neolithic: Perspectives from a late neolithic bell beaker burial site in Germany. *Am J Phys Anthropol*. 2012;148(4):571-579.

289. Brandt G, Haak W, Adler CJ, et al. Ancient DNA Reveals Key Stages in the Formation of Central European Mitochondrial Genetic Diversity. *Science*. 2013;342(6155):257-261.

290. Chandler H, Sykes B, Zilhão J. Using ancient DNA to examine genetic continuity at the Mesolithic-Neolithic transition in Portugal. In: *Actas Del III Congreso Del Neolítico En La Península Ibérica. Monografías Del Instituto Internacional de Investigaciones Prehistóricas de Cantabria*. ; 2005:781-786.

291. Haak W, Forster P, Bramanti B, et al. Ancient DNA from the first European farmers in 7500-year-old Neolithic sites. *Science*. 2005;310(5750):1016-1018.

292. Bramanti B. Ancient DNA: genetic analysis of aDNA from sixteen skeletons of the Vedrovice. *Anthropologie*. 2008;46:153-160.

293. Cunliffe B. *Europe Between the Oceans: 9000 BC-AD 1000*. Yale University Press; 2011.

294. Soares P, Achilli A, Semino O, et al. The Archaeogenetics of Europe. *Curr Biol*. 2010;20(4):R174-R183.

295. Morelli L, Contu D, Santoni F, Whalen MB, Francalacci P, Cucca F. A Comparison of Y-Chromosome Variation in Sardinia and Anatolia Is More Consistent with Cultural Rather than Demic Diffusion of Agriculture. *PLoS One*. 2010;5(4):e10419.

296. Cardoso S, Valverde L, Alfonso-Sánchez MA, et al. The Expanded mtDNA Phylogeny of the Franco-Cantabrian Region Upholds the Pre-Neolithic Genetic Substrate of Basques. *PLoS One*. 2013;8(7):e67835.

297. Cinnioglu C, King R, Kivisild T, et al. Excavating Y-chromosome haplotype strata in Anatolia. *Hum Genet*. 2004;114(2):127-148.

298. Klyosov AA. Ancient History of the Arbins, Bearers of Haplogroup R1b, from Central Asia to Europe, 16,000 to 1500 Years before Present. *Adv Anthropol*. 2012;02(02):87-105.

299. Jobling MA, Hurles ME, Tyler-Smith C. *Human Evolutionary Genetics: Origins, Peoples & Disease*. Garland Science; 2004.

300. Balanovsky O. Toward a consensus on SNP and STR mutation rates on the human Y-chromosome. *Hum Genet*. 2017;136(5):575-590.

301. Stumpf MPH. Genealogical and Evolutionary Inference with the Human Y Chromosome. *Science*. 2001;291(5509):1738-1742.

302. Forster P, Harding R, Torroni A, Bandelt HJ. Origin and evolution of Native American mtDNA variation: a reappraisal. *Am J Hum Genet*. 1996;59(4):935-945.

303. Goldstein DB, Ruiz Linares A, Cavalli-Sforza LL, Feldman MW. An evaluation of genetic distances for use with microsatellite loci. *Genetics*. 1995;139(1):463-471.

304. Goldstein DB, Ruiz Linares A, Cavalli-Sforza LL, Feldman MW. Genetic absolute dating based on microsatellites and the origin of modern humans. *Proc Natl Acad Sci*.

1995;92(15):6723-6727.

305. Kingman JFC. On the genealogy of large populations. *J Appl Probab*. 1982;19(A):27-43.

306. Hudson RR. Gene genealogies and the coalescent process. *Oxford Surv Evol Biol*. 1990;7:1-44.

307. Wilson IJ, Weale ME, Balding DJ. Inferences from DNA data: population histories, evolutionary processes and forensic match probabilities. *J R Stat Soc Ser A - Statistics Soc*. 2003;166(2):155-188.

308. Bahlo M, Griffiths RC. Inference from Gene Trees in a Subdivided Population. *Theor Popul Biol*. 2000;57(2):79-95.

309. Bouckaert R, Heled J, Kühnert D, et al. BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. *PLoS Comput Biol*. 2014;10(4):e1003537.

310. Zhivotovsky LA, Underhill PA, Cinnioğlu C, et al. The Effective Mutation Rate at Y Chromosome Short Tandem Repeats, with Application to Human Population-Divergence Time. *Am J Hum Genet*. 2004;74(1):50-61.

311. Zhivotovsky LA, Underhill PA, Feldman MW. Difference between Evolutionarily Effective and Germ line Mutation Rate Due to Stochastically Varying Haplogroup Size. *Mol Biol Evol*. 2006;23(12):2268-2270.

312. Di Giacomo F, Luca F, Popa LO, et al. Y chromosomal haplogroup J as a signature of the post-neolithic colonization of Europe. *Hum Genet*. 2004;115(5):357-371.

313. Marshall OJ. PerlPrimer: cross-platform, graphical primer design for standard, bisulphite and real-time PCR. *Bioinformatics*. 2004;20(15):2471-2472.

314. Vallone PM, Butler JM. AutoDimer: a screening tool for primer-dimer and hairpin structures. *Biotechniques*. 2004;37(2):226-231.

315. Butler JM, Decker AE, Vallone PM, Kline MC. Allele frequencies for 27 Y-STR loci with U.S. Caucasian, African American, and Hispanic samples. *Forensic Sci Int*. 2006;156(2-3):250-260.

316. Sánchez-Diz P, Gusmão L, Beleza S, et al. Results of the GEP-ISFG collaborative study on two Y-STRs tetraplexes: GEPY I (DYS461, GATA C4, DYS437 and DYS438) and GEPY II (DYS460, GATA A10, GATA H4 and DYS439). *Forensic Sci Int*. 2003;135(2):158-162.

317. Nebel A, Filon D, Hohoff C, Faerman M, Brinkmann B, Oppenheim A. Haplogroup-specific deviation from the stepwise mutation model at the microsatellite loci DYS388 and DYS392. *Eur J Hum Genet*. 2001;9(1):22-26.

318. Sanchez JJ, Børsting C, Morling N. Typing of Y chromosome SNPs with multiplex PCR methods. *Methods Mol Biol*. 2005;297:209-228.

319. Palencia-Madrid L, Cardoso S, Castro-Maestre F, Baroja-Careaga I, Rocandio AM, de Pancorbo MM. Development of a new screening method to determine the main 52 mitochondrial haplogroups through a single minisequencing reaction. *Mitochondrion*. February 2018.

320. SWGDAM. *SWGDAM Validation Guidelines for DNA Analysis Methods*.; 2016.

321. Excoffier L, Lischer HEL. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour*. 2010;10(3):564-567.

322. Nei M. *Molecular Evolutionary Genetics*. Columbia University Press; 1987.

323. Rice WR. Analyzing Tables of Statistical Tests. *Evolution*. 1989;43(1):223.

324. Hammer Ø, Harper DAT, Ryan PD. Paleontological statistics software package for education and data analysis. *Palaeontol Electron*. 2001;4:9-18.

325. Adler D, Murdoch D. rgl: 3D visualization device (OpenGL). *R package version*. 2013:http://cran.r-project.org/web/packages/rgl/index.h.

326. Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F. GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. *Lab génome, Popul Interact CNRS Umr 5171*. 2004.

327. Jombart T. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*. 2008;24(11):1403-1405.

328. Jombart T, Devillard S, Dufour A-B, Pontier D. Revealing cryptic spatial patterns in genetic variability by a new multivariate method. *Heredity*. 2008;101(1):92-103.

329. Bandelt HJ, Forster P, Rohl A. Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol*. 1999;16(1):37-48.

330. Bandelt HJ, Forster P, Rohl A. Median-joining networks for inferring intraspecific

phylogenies. *Mol Biol Evol*. 1999;16(1):37-48.

331. Klyosov AA, Kilin V V. Kilin-Klyosov TMRCA Calculator for Time Spans up to Millions of Years. *Adv Anthropol*. 2016;06(03):51-71.

332. Helgason A, Einarsson AW, Guðmundsdóttir VB, et al. The Y-chromosome point mutation rate in humans. *Nat Genet*. 2015;47:453-457.

333. Saillard J, Forster P, Lynnerup N, Bandelt H-J, Nørby S. mtDNA Variation among Greenland Eskimos: The Edge of the Beringian Expansion. *Am J Hum Genet*. 2000;67(3):718-726.

334. Goldstein DB, Ruiz Linares A, Cavalli-Sforza LL, Feldman MW. Genetic absolute dating based on microsatellites and the origin of modern humans. *Proc Natl Acad Sci*. 1995;92(15):6723-6727.

335. Goldstein DB, Ruiz Linares A, Cavalli-Sforza LL, Feldman MW. An evaluation of genetic distances for use with microsatellite loci. *Genetics*. 1995;139(1):463-471.

336. Lintusaari J, Gutmann MU, Dutta R, Kaski S, Corander J. Fundamentals and Recent Developments in Approximate Bayesian Computation. *Syst Biol*. 2017;66(1):e66-e82.

337. Excoffier L, Foll M. fastsimcoal: a continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. *Bioinformatics*. 2011;27(9):1332-1334.

338. Excoffier L, Dupanloup I, Huerta-Sánchez E, Sousa VC, Foll M. Robust Demographic Inference from Genomic and SNP Data. *PLoS Genet*. 2013;9(10):e1003905.

339. Pritchard JK, Seielstad MT, Perez-Lezaun A, Feldman MW. Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *Mol Biol Evol*. 1999;16(12):1791-1798.

340. Beaumont MA. Joint determination of topology, divergence time, and immigration in population trees. In: *Simulations, Genetics and Human Prehistory*. McDonald Institute for Archaelogical Research; 2008:135-154.

341. Vai S, Ghirotto S, Pilli E, et al. Genealogical Relationships between Early Medieval and Modern Inhabitants of Piedmont. *PLoS One*. 2015;10(1):e0116801.

342. Beaumont MA, Zhang W, Balding DJ. Approximate Bayesian computation in population genetics. *Genetics*. 2002;162(4):2025-2035.

343. Hamilton G, Currat M, Ray N, Heckel G, Beaumont M, Excoffier L. Bayesian estimation of recent migration rates after a spatial expansion. *Genetics*. 2005;170(1):409-417.

344. Chiaroni J, Underhill PA, Cavalli-Sforza LL. Y chromosome diversity, human expansion, drift, and cultural evolution. *Proc Natl Acad Sci*. 2009;106(48):20174-20179.

345. Larmuseau MHD, Vanderheyden N, Jacobs M, Coomans M, Larno L, Decorte R. Micro-geographic distribution of Y-chromosomal variation in the central-western European region Brabant. *Forensic Sci Int Genet*. 2011;5(2):95-99.

346. Olalde I, Brace S, Allentoft ME, et al. The Beaker phenomenon and the genomic transformation of northwest Europe. *Nature*. 2018;555(7695):190-196.

347. Semino O, Passarino G, Oefner PJ, et al. The genetic legacy of Paleolithic Homo sapiens sapiens in extant Europeans: a Y chromosome perspective. *Science*. 2000;290(5494):1155-1159.

348. Sjödin P, François O. Wave-of-Advance Models of the Diffusion of the Y Chromosome Haplogroup R1b1b2 in Europe. *PLoS One*. 2011;6(6):e21592.

349. Shi W, Ayub Q, Vermeulen M, et al. A Worldwide Survey of Human Male Demographic History Based on Y-SNP and Y-STR Data from the HGDP-CEPH Populations. *Mol Biol Evol*. 2010;27(2):385-393.

350. Haak W, Lazaridis I, Patterson N, et al. Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature*. 2015;522:207-211.

351. Allentoft ME, Sikora M, Sjögren K-G, et al. Population genomics of Bronze Age Eurasia. *Nature*. 2015;522(7555):167-172.

352. Martiniano R, Cassidy LM, Ó'Maoldúin R, et al. The population genomics of archaeological transition in west Iberia: Investigation of ancient substructure using imputation and haplotype-based methods. *PLOS Genet*. 2017;13(7):e1006852.

353. Adams SM, Bosch E, Balaresque PL, et al. The Genetic Legacy of Religious Diversity and Intolerance: Paternal Lineages of Christians, Jews, and Muslims in the Iberian Peninsula. *Am J Hum Genet*. 2008;83(6):725-736.

354. Flores C, Maca-Meyer N, González AM, et al. Reduced genetic structure of the Iberian peninsula revealed by Y-chromosome analysis: implications for population demography. *Eur J Hum Genet*. 2004;12(10):855-863.

355. Rey-González D, Gelabert-Besada M, Cruz R, et al. Micro and macro geographical analysis of Y-chromosome lineages in South Iberia. *Forensic Sci Int Genet*. 2017;29:e9-e15.

356. Regueiro M, Garcia-Bertrand R, Fadhlaoui-Zid K, Álvarez J, Herrera RJ. From Arabia to Iberia: A Y chromosome perspective. *Gene*. 2015;564(2):141-152.

357. Durbin RM, Altshuler DL, Durbin RM, et al. A map of human genome variation from population-scale sequencing. *Nature*. 2010;467(7319):1061-1073.

358. Martínez-Cruz B, Harmant C, Platt DE, et al. Evidence of Pre-Roman Tribal Genetic Structure in Basques from Uniparentally Inherited Markers. *Mol Biol Evol*. 2012;29(9):2211-2222.

359. Belenguer E. *Jaime I y Su Reinado*. Editorial Milenio; 2008.

360. Rębała K, Martínez-Cruz B, Tönjes A, et al. Contemporary paternal genetic landscape of Polish and German populations: from early medieval Slavic expansion to post-World War II resettlements. *Eur J Hum Genet*. 2013;21(4):415-422.

361. Larmuseau MHD, Vanderheyden N, Van Geystelen A, van Oven M, de Knijff P, Decorte R. Recent Radiation within Y-chromosomal Haplogroup R-M269 Resulted in High Y-STR Haplotype Resemblance. *Ann Hum Genet*. 2014;78(2):92-103.

362. Boattini A, Martinez-Cruz B, Sarno S, et al. Uniparental Markers in Italy Reveal a Sex-Biased Genetic Structure and Different Historical Strata. *PLoS One*. 2013;8(5):e65441.

363. Beleza S, Gusmao L, Lopes A, et al. Micro-Phylogeographic and Demographic History of Portuguese Male Lineages. *Ann Hum Genet*. 2006;70(2):181-194.

364. Villar Liébana M. *Indo-Europoeos y No Indo-Europeos En La Hispania Prerromana*. Universidad de Salamanca; 2000.

365. Salinas de Frías M. *Los Pueblos Prerromanos de La Península Ibérica*. Akal; 2006.

366. McEvedy C, Jones R. *Atlas of World Population History*. Penguin; 1978.

367. Hazard HW. Chapter XII: The Spanish and Portiguese Reconquest, 1095-1492. In: *A History of the Crusades, Volume III, The Fourteenth and Fifteenth Centuries*. Madison, Wisconsin: University of Wisconsin Press; 1975:396-456.

368. Larmuseau MHD, Vanoverbeke J, Gielis G, Vanderheyden N, Larmuseau HFM, Decorte R. In the name of the migrant father—Analysis of surname origins identifies genetic

admixture events undetectable from genealogical records. *Heredity*. 2012;109(2):90-95.

369. Larmuseau MHD, Calafell F, Princen SA, Decorte R, Soen V. The black legend on the Spanish presence in the low countries: Verifying shared beliefs on genetic ancestry. *Am J Phys Anthropol*. 2018;166(1):219-227.

370. Soen V. Más allá de la leyenda negra. Léon van der Essen y la historiografía reciente en torno al castigo de las ciudades rebeldes en los Países Bajos (siglos XIV a XVI). In: *El Ejército Español En Flandes 1567–1584*. Yuste: Academia de Yuste; 2008:45-72.

371. Pipkin A. *Rape in the Republic, 1609–1725: Formulating Dutch Identity*. Brill; 2013.

372. Solé-Morata N, Bertranpetit J, Comas D, Calafell F. Recent Radiation of R-M269 and High Y-STR Haplotype Resemblance Confirmed. *Ann Hum Genet*. 2014;78(4):253-254.

373. Zhou J, Teo Y-Y. Estimating time to the most recent common ancestor (TMRCA): comparison and application of eight methods. *Eur J Hum Genet*. 2016;24(8):1195-1201.

374. Kingman JF. Origins of the coalescent. 1974-1982. *Genetics*. 2000;156(4):1461-1463.

375. Tavaré S, Balding DJ, Griffiths RC, Donnelly P. Inferring coalescence times from DNA sequence data. *Genetics*. 1997;145(2):505-518.

376. Meligkotsidou L, Fearnhead P. Maximum-likelihood estimation of coalescence times in genealogical trees. *Genetics*. 2005;171(4):2073-2084.

377. Cox MP. Accuracy of molecular dating with the rho statistic: deviations from coalescent expectations under a range of demographic models. *Hum Biol*. 2008;80(4):335-357.

378. Hu Q, Liu Y, Yi S, Huang D. A comparison of four methods for PCR inhibitor removal. *Forensic Sci Int Genet*. 2015;16:94-97.

379. Tebbe CC, Vahjen W. Interference of humic acids and DNA extracted directly from soil in detection and transformation of recombinant DNA from bacteria and a yeast. *Appl Environ Microbiol*. 1993;59(8):2657-2665.

380. Akane A, Matsubara K, Nakamura H, Takahashi S, Kimura K. Identification of the heme compound copurified with deoxyribonucleic acid (DNA) from bloodstains, a major inhibitor of polymerase chain reaction (PCR) amplification. *J Forensic Sci*. 1994;39(2):362-372.

381. Roewer L, Nothnagel M, Gusmão L, et al. Continent-Wide Decoupling of Y-Chromosomal

Genetic Variation from Language and Geography in Native South Americans. *PLoS Genet*. 2013;9(4):e1003460.

382. Muzzio M, Ramallo V, Motti JMB, Santos MR, López Camelo JS, Bailliet G. Software for Y-haplogroup predictions: a word of caution. *Int J Legal Med*. 2011;125(1):143-147.

8. Appendix

Attached Table 1. 17 YSTR-YSNP data from the Basque sample of population. Corresponds to Supplementary Table S3 from *Study Number 1.*

ID	Characteristics	Final Haplogroup	M269	L11	U106	S116	DF27	U152	M529	L238	DF19	DYS19	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	DYS385	DYS438	DYS439	DYS437	DYS448	DYS456	DYS458	DYS635	YGATAH4
1	Urban_non native	L11*	1	1	0	0	0	0	0	0	0	14	12	28	24	10	13	13	11,14	12	12	15	19	16	17	23	12
2	Urban_non native	L11*	1	1	0	0	0	0	0	0	0	14	13	29	23	11	13	13	11,15	12	13	16	19	15	18	25	12
3	Rural_Native	U106	1	1	1	0	0	0	0	0	0	14	14	30	23	13	13	13	11,14	12	12	15	19	16	17	23	13
4	Urban_native	U106	1	1	1	0	0	0	0	0	0	14	13	29	23	11	13	13	11,14	12	11	15	19	16	17	23	12
5	Urban_native	U106	1	1	1	0	0	0	0	0	0	14	12	28	24	11	13	13	11,13	12	12	15	19	15	17	25	13
6	Urban_non native	U106	1	1	1	0	0	0	0	0	0	14	13	29	23	11	13	13	11,14	12	12	15	19	15	18	23	11
7	Urban_non native	U106	1	1	1	0	0	0	0	0	0	14	13	30	25	10	13	13	11,14	11	13	15	19	15	18	23	12
8	Rural_Native	U152	1	1	0	1	0	1	0	0	0	14	12	28	24	10	13	13	11,14	12	11	15	20	16	17	24	12
9	Rural_Native	U152	1	1	0	1	0	1	0	0	0	14	12	28	24	10	13	13	11,14	12	11	15	20	16	17	24	12
10	Rural_Native	U152	1	1	0	1	0	1	0	0	0	13	12	28	24	11	13	13	11,14	12	12	15	19	15	15	23	13
11	Rural_Native	U152	1	1	0	1	0	1	0	0	0	14	13	29	24	11	13	13	11,14	12	12	14	19	15	17	24	12
12	Rural_Native	U152	1	1	0	1	0	1	0	0	0	14	13	29	24	11	13	13	12,14	12	12	15	19	15	18	24	11
13	Urban_non native	U152	1	1	0	1	0	1	0	0	0	14	14	30	24	11	13	13	11,14	12	12	15	19	16	17	23	11
14	Urban_non native	U152	1	1	0	1	0	1	0	0	0	15	13	29	24	10	13	13	11,16	13	13	15	19	14	17	23	12
15	Rural_Native	M529	1	1	0	1	0	0	1	0	0	15	13	29	24	11	13	13	12,14	12	11	15	19	16	17	23	12
16	Urban_native	M529	1	1	0	1	0	0	1	0	0	14	14	30	24	11	13	13	11,14	12	11	15	19	15	19	23	11
17	Urban_native	M529	1	1	0	1	0	0	1	0	0	14	13	29	24	11	13	13	12,15	12	11	15	19	16	15	23	12
18	Urban_non native	M529	1	1	0	1	0	0	1	0	0	14	13	29	23	11	13	13	11,14	12	12	15	19	15	17	24	12
19	Urban_non native	M529	1	1	0	1	0	0	1	0	0	14	14	30	24	11	13	13	10,14	12	12	15	19	15	18	23	12
20	Rural_Native	S116*	1	1	0	1	0	0	0	0	0	14	13	30	24	10	13	13	11,14	12	13	15	20	15	16	23	12
21	Rural_Native	S116*	1	1	0	1	0	0	0	0	0	14	14	30	24	11	13	13	11,15	12	11	15	19	15	16	23	12
22	Rural_Native	S116*	1	1	0	1	0	0	0	0	0	14	14	30	24	10	13	13	11,14	12	11	14	19	16	20	24	12
23	Rural_Native	S116*	1	1	0	1	0	0	0	0	0	14	13	29	24	11	13	13	12,14	12	11	15	19	15	17	23	12
24	Rural_Native	S116*	1	1	0	1	0	0	0	0	0	14	13	29	24	11	13	13	11,15	12	11	15	19	15	17	23	12

Attached Table 1. Continuation.

ID	Characteristics	Final Haplogroup	M269	L11	U106	S116	DF27	U152	M529	L238	DF19	DYS19	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	DYS385	DYS438	DYS439	DYS437	DYS448	DYS456	DYS458	DYS635	YGATAH4
25	Rural_Native	S116*	1	1	0	1	0	0	0	0	0	14	14	30	24	11	13	13	11,17	12	11	15	19	15	17	23	12
26	Rural_Native	S116*	1	1	0	1	0	0	0	0	0	14	13	29	24	11	13	13	12,14	12	11	15	19	15	17	23	12
27	Rural_Native	S116*	1	1	0	1	0	0	0	0	0	14	13	29	24	10	13	13	12,14	12	12	15	19	16	17	23	12
28	Rural_Native	S116*	1	1	0	1	0	0	0	0	0	14	14	31	24	10	13	13	11,14	12	12	14	19	16	17	23	12
29	Rural_Native	S116*	1	1	0	1	0	0	0	0	0	14	14	30	24	11	13	13	11,14	12	11	15	19	15	18	23	11
30	Rural_Native	S116*	1	1	0	1	0	0	0	0	0	14	14	30	24	11	13	13	11,14	12	11	15	19	15	18	23	11
31	Rural_Native	S116*	1	1	0	1	0	0	0	0	0	14	14	30	24	11	13	13	11,14	12	11	15	19	15	19	23	11
32	Rural_Native	S116*	1	1	0	1	0	0	0	0	0	14	14	30	24	11	13	13	11,15	12	11	15	19	15	17	23	12
33	Rural_Native	S116*	1	1	0	1	0	0	0	0	0	14	14	30	24	10	13	13	11,14	12	11	15	19	15	18	23	11
34	Rural_Native	S116*	1	1	0	1	0	0	0	0	0	14	14	31	24	11	13	13	14,14	12	13	15	19	16	18	24	13
35	Rural_Native	S116*	1	1	0	1	0	0	0	0	0	15	13	29	24	11	13	13	12,14	12	11	15	19	16	16	23	12
36	Rural_Native	S116*	1	1	0	1	0	0	0	0	0	14	14	30	24	10	13	13	11,14	12	14	15	20	16	17	23	12
37	Rural_Native	S116*	1	1	0	1	0	0	0	0	0	14	14	31	24	11	13	13	11,13	12	11	15	19	15	17	24	12
38	Rural_Native	S116*	1	1	0	1	0	0	0	0	0	14	13	29	24	11	13	13	11,14	12	11	15	19	15	18	23	12
39	Rural_Native	S116*	1	1	0	1	0	0	0	0	0	14	13	29	24	11	13	13	11,14	12	11	15	19	15	18	23	13
40	Rural_Native	S116*	1	1	0	1	0	0	0	0	0	14	14	30	24	11	13	13	11,14	12	11	15	19	15	19	23	11
41	Rural_Native	S116*	1	1	0	1	0	0	0	0	0	14	14	30	24	10	13	13	11,14	12	12	15	19	15	17	24	12
42	Rural_Native	S116*	1	1	0	1	0	0	0	0	0	14	15	31	24	11	13	13	11,14	12	11	15	19	15	18	23	12
43	Rural_Native	S116*	1	1	0	1	0	0	0	0	0	14	13	29	25	11	14	13	11,13	12	13	15	18	17	17	23	12
44	Rural_Native	S116*	1	1	0	1	0	0	0	0	0	14	14	30	24	10	13	13	11,14	12	12	15	19	16	17	23	12
45	Rural_Native	S116*	1	1	0	1	0	0	0	0	0	14	14	30	24	11	13	13	11,14	12	11	15	19	15	18	23	11
46	Rural_Native	S116*	1	1	0	1	0	0	0	0	0	14	14	30	24	11	13	13	11,15	12	11	15	19	15	18	24	11
47	Rural_Native	S116*	1	1	0	1	0	0	0	0	0	14	14	30	24	11	13	13	11,14	12	11	15	19	15	19	23	11
48	Rural_Native	S116*	1	1	0	1	0	0	0	0	0	14	13	29	24	11	13	13	12,14	12	11	15	19	15	17	23	12

Attached Table 1. Continuation.

ID	Characteristics	Final Haplogroup	M269	L11	U106	S116	DF27	U152	M529	L238	DF19	DYS19	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	DYS385	DYS438	DYS439	DYS437	DYS448	DYS456	DYS458	DYS635	YGATAH4
49	Rural_Native	S116*	1	1	0	1	0	0	0	0	0	14	14	30	24	11	13	13	11,14	12	11	15	19	15	19	23	11
50	Urban_native	S116*	1	1	0	1	0	0	0	0	0	14	15	31	24	11	13	13	12,17	12	11	15	19	15	19	23	12
51	Urban_native	S116*	1	1	0	1	0	0	0	0	0	14	14	30	24	11	13	13	11,14	12	11	15	19	15	18	23	12
52	Urban_native	S116*	1	1	0	1	0	0	0	0	0	14	13	29	24	11	13	13	11,14	12	12	15	19	17	17	23	12
53	Urban_native	S116*	1	1	0	1	0	0	0	0	0	14	13	29	24	11	13	13	12,14	12	11	15	19	15	17	23	12
54	Urban_native	S116*	1	1	0	1	0	0	0	0	0	14	14	31	24	11	13	12	11,14	12	11	15	19	15	17	23	12
55	Urban_non native	S116*	1	1	0	1	0	0	0	0	0	13	14	30	23	10	11	14	17,17	10	13	14	21	15	16	21	11
56	Urban_non native	S116*	1	1	0	1	0	0	0	0	0	14	13	29	25	11	13	13	11,14	13	12	15	19	16	18	23	12
57	Urban_non native	S116*	1	1	0	1	0	0	0	0	0	14	13	29	24	11	13	13	13,14	11	12	14	19	16	17	23	12
58	Urban_non native	S116*	1	1	0	1	0	0	0	0	0	14	14	31	25	11	13	13	11,14	12	12	15	19	15	17	23	12
59	Urban_non native	S116*	1	1	0	1	0	0	0	0	0	14	13	29	25	11	13	13	11,14	12	12	14	18	17	17	24	12
60	Urban_non native	S116*	1	1	0	1	0	0	0	0	0	14	15	31	24	11	14	13	11,14	12	11	15	19	15	17	23	13
61	Urban_non native	S116*	1	1	0	1	0	0	0	0	0	16	13	29	24	11	13	13	10,15	12	12	15	18	16	17	23	12
62	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	13	30	24	10	13	13	11,14	12	11	15	19	16	16	23	12
63	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	24	11	13	13	11,14	12	12	15	19	15	17	23	12
64	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	23	11	13	13	12,14	12	11	15	19	16	17	23	13
65	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	24	11	13	13	11,14	12	12	15	19	16	16	23	12
66	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	25	11	13	13	11,14	12	12	15	19	15	17	23	12
67	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	12	28	24	11	13	13	11,14	12	12	15	19	16	16	23	12
68	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	12	28	24	11	13	13	11,14	12	12	15	19	16	16	23	12
69	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	14	30	24	11	13	13	11,14	12	12	15	19	15	17	23	12
70	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	12	28	25	11	13	13	11,14	12	12	15	19	15	17	23	12
71	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	15	14	30	23	10	13	13	11,11	12	13	15	19	18	18	23	13
72	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	25	10	13	12	11,14	12	13	15	19	15	17	23	13

Attached Table 1. Continuation.

ID	Characteristics	Final Haplogroup	M269	L11	U106	S116	DF27	U152	M529	L238	DF19	DYS19	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	DYS385	DYS438	DYS439	DYS437	DYS448	DYS456	DYS458	DYS635	YGATAH4
73	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	13	30	24	10	13	13	11,14	12	12	14	19	16	17	23	13
74	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	14	30	24	10	13	13	11,11	12	12	15	19	18	17	23	12
75	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	24	11	13	13	11,14	12	13	15	19	15	17	23	12
76	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	13	30	24	11	13	13	11,14	12	12	15	19	16	17	23	13
77	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	24	11	13	13	11,14	12	12	15	19	16	16	23	12
78	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	24	11	13	13	11,14	12	12	15	19	15	17	23	12
79	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	24	10	13	12	11,13	12	13	15	19	15	17	23	12
80	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	24	10	13	13	11,14	12	13	15	19	15	16	23	12
81	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	24	11	13	13	11,14	12	12	15	19	15	17	23	12
82	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	24	11	13	13	11,14	12	11	15	19	16	17	23	12
83	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	24	11	13	13	11,14	12	12	15	19	16	16	23	12
84	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	12	29	24	11	13	13	11,14	12	13	15	19	16	16	23	12
85	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	12	29	24	11	13	13	11,14	12	13	15	19	16	16	23	12
86	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	24	11	13	13	11,14	12	12	14	19	15	16	23	12
87	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	12	28	24	11	13	13	11,14	12	12	15	19	15	16	23	12
88	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	14	30	25	11	13	13	11,14	12	12	15	19	15	17	23	11
89	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	24	10	14	13	11,14	12	11	15	19	16	18	23	12
90	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	12	28	24	11	13	13	10,14	12	12	15	19	16	16	23	12
91	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	13	30	24	11	13	13	11,14	12	12	15	19	16	18	23	12
92	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	13	30	24	11	13	13	11,14	12	11	15	19	16	17	23	12
93	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	12	28	24	11	13	13	11,14	12	12	15	19	16	16	23	12
94	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	24	11	13	12	11,13	12	13	15	19	15	17	23	11
95	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	24	11	13	13	11,14	10	12	15	19	16	15	23	12
96	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	13	31	23	10	13	13	12,14	12	12	15	19	14	19	26	12

Attached Table 1. Continuation.

ID	Characteristics	Final Haplogroup	M269	L11	U106	S116	DF27	U152	M529	L238	DF19	DYS19	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	DYS385	DYS438	DYS439	DYS437	DYS448	DYS456	DYS458	DYS635	YGATAH4
97	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	13	31	23	10	13	13	12,14	12	12	15	19	14	19	26	12
98	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	24	10	13	13	11,14	12	12	15	19	16	17	23	12
99	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	12	28	24	11	13	13	10,14	12	11	15	19	16	16	23	12
100	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	12	28	24	11	14	14	11,14	12	12	14	19	17	16	23	12
101	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	13	30	24	11	13	13	11,14	12	12	15	19	16	17	23	13
102	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	13	30	24	11	13	13	11,14	12	12	15	19	16	17	23	13
103	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	24	11	13	13	12,14	12	12	15	19	15	17	23	12
104	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	12	28	23	11	13	13	11,14	12	12	15	19	16	16	23	12
105	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	23	11	13	13	11,14	12	12	15	19	16	16	23	12
106	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	24	11	13	13	11,14	12	12	15	19	16	17	23	12
107	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	14	30	23	10	13	13	11,11	12	13	15	19	16	18	23	12
108	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	24	11	13	13	11,14	12	12	15	19	15	18	23	12
109	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	24	11	13	13	11,14	12	12	15	19	15	17	23	12
110	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	25	11	11	13	11,14	12	12	15	19	17	19	23	12
111	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	13	30	24	11	13	13	11,14	12	12	15	19	15	16	23	12
112	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	13	30	24	10	13	13	11,14	12	12	15	19	16	17	23	13
113	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	13	30	24	11	13	13	11,14	12	12	14	19	16	17	23	12
114	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	13	30	24	11	13	13	11,14	12	11	15	19	16	17	23	12
115	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	12	28	24	11	13	13	10,14	12	12	15	19	16	16	23	11
116	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	14	30	23	10	13	13	11,11	12	11	15	19	17	19	23	12
117	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	14	30	23	10	13	13	11,11	12	13	15	19	17	18	23	12
118	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	14	30	23	10	13	13	11,11	12	13	15	19	17	18	23	12
119	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	24	11	13	13	11,14	12	12	15	19	15	17	23	12
120	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	25	10	13	12	11,13	12	13	15	20	16	17	23	12

Attached Table 1. Continuation.

ID	Characteristics	Final Haplogroup	M269	L11	U106	S116	DF27	U152	M529	L238	DF19	DYS19	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	DYS385	DYS438	DYS439	DYS437	DYS448	DYS456	DYS458	DYS635	YGATAH4
121	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	24	11	13	13	11,14	12	12	15	19	15	17	23	12
122	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	14	31	25	10	13	12	11,14	12	12	15	19	15	17	23	12
123	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	16	13	29	23	11	13	13	11,14	12	13	14	18	16	19	23	12
125	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	12	27	24	11	13	13	12,14	12	12	15	19	15	16	23	12
126	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	24	11	13	13	11,14	13	13	15	19	16	15	23	12
127	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	14	32	24	11	13	13	11,14	12	12	15	20	16	17	23	12
128	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	14	30	23	11	13	14	11,15	12	12	14	19	17	17	23	12
129	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	14	31	24	11	13	13	11,14	12	12	14	18	15	17	23	11
130	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	24	10	13	13	11,14	12	11	15	19	16	17	22	12
131	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	14	30	24	11	13	13	11,15	12	12	15	19	15	17	23	12
132	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	24	11	13	13	12,14	12	12	15	19	16	16	23	11
133	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	13	32	24	11	13	13	11,14	12	12	15	19	15	16	23	12
134	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	24	11	13	13	12,14	12	12	15	19	16	16	23	11
135	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	24	11	13	13	12,14	12	12	15	19	16	16	23	11
136	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	24	11	13	13	11,14	13	12	15	19	15	19	23	12
137	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	24	10	13	13	11,14	12	12	15	19	15	18	23	12
138	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	15	13	29	24	11	13	13	11,15	12	13	15	18	14	18	23	13
139	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	14	30	24	11	13	13	11,14	12	12	14	18	15	18	24	11
140	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	24	11	13	13	11,15	12	11	14	18	16	17	24	11
141	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	24	11	13	13	11,14	12	12	14	18	16	17	23	12
142	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	14	31	24	10	13	13	11,14	12	11	14	18	15	17	23	11
143	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	14	30	23	11	13	14	11,14	12	13	14	18	17	17	23	11
144	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	14	30	24	11	13	13	11,14	12	11	14	18	16	18	23	11
145	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	24	11	13	13	13,14	12	12	14	18	15	17	23	11

Attached Table 1. Continuation.

ID	Characteristics	Final Haplogroup	M269	L11	U106	S116	DF27	U152	M529	L238	DF19	DYS19	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	DYS385	DYS438	DYS439	DYS437	DYS448	DYS456	DYS458	DYS635	YGATAH4
146	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	23	11	13	14	11,14	12	13	14	18	16	18	23	11
147	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	14	31	24	11	13	13	11,14	12	12	14	18	16	17	23	11
148	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	14	30	24	11	13	13	11,14	12	12	14	18	15	17	23	11
149	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	14	30	24	11	13	13	11,15	12	13	14	18	15	16	24	11
150	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	14	32	24	10	13	13	11,14	12	12	15	18	15	18	24	11
151	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	15	15	32	25	10	13	13	11,14	12	12	14	18	15	17	23	11
152	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	15	31	24	11	13	13	12,14	12	12	14	18	15	17	23	11
153	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	15	14	30	25	11	13	13	11,14	12	12	14	18	15	17	23	11
154	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	14	30	23	11	13	14	11,14	12	13	14	18	16	17	23	11
155	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	12	28	23	11	13	14	11,15	12	13	14	18	16	17	23	11
156	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	23	10	13	14	11,14	12	14	14	18	16	19	23	11
157	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	23	11	13	14	11,14	12	12	14	18	15	17	23	11
158	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	14	31	24	11	13	13	11,14	12	12	14	18	16	17	23	11
159	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	15	15	32	25	11	14	13	11,13	12	12	14	18	15	17	23	11
160	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	14	31	24	11	13	13	11,14	12	12	14	18	16	17	23	11
161	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	24	11	13	13	11,15	12	12	14	18	16	17	23	11
162	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	24	11	13	14	11,14	12	13	14	18	16	17	23	12
163	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	14	31	23	11	13	13	11,14	12	12	14	18	16	17	23	11
164	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	23	11	13	14	11,14	12	14	14	18	16	18	23	11
165	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	24	10	13	13	11,14	12	13	14	18	16	18	23	11
166	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	24	10	13	13	11,14	12	12	14	18	15	17	24	12
167	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	14	31	25	11	13	13	11,14	12	12	14	18	15	16	23	11
168	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	14	30	24	11	13	13	12,14	12	11	14	18	15	16	23	11
169	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	24	10	13	13	11,14	12	12	14	18	15	17	23	12

Attached Table 1. Continuation.

ID	Characteristics	Final Haplogroup	M269	L11	U106	S116	DF27	U152	M529	L238	DF19	DYS19	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	DYS385	DYS438	DYS439	DYS437	DYS448	DYS456	DYS458	DYS635	YGATAH4
170	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	14	30	24	11	14	13	11,15	12	12	14	18	15	18	23	11
171	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	13	30	24	11	14	13	11,14	12	13	14	18	16	16	24	11
172	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	14	30	24	11	13	13	12,14	12	14	14	18	15	17	23	11
173	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	15	14	30	24	11	12	13	11,16	12	13	14	17	16	17	23	11
174	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	14	31	24	11	13	12	12,14	12	12	14	18	15	16	23	11
175	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	24	11	13	13	11,13	12	12	14	18	15	16	23	12
176	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	23	10	13	14	11,14	12	14	14	18	16	18	23	11
177	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	23	11	13	14	11,14	12	14	14	18	16	18	23	11
178	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	23	11	13	14	11,14	12	14	14	18	16	18	23	11
179	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	14	30	24	10	13	13	11,14	12	13	14	18	15	16	23	11
180	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	14	31	24	11	13	13	11,14	12	12	14	18	15	17	23	11
181	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	14	30	24	11	12	13	12,15	12	11	14	18	15	17	23	11
182	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	14	30	24	10	13	13	11,14	12	11	14	18	15	17	23	11
183	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	25	11	13	14	11,14	12	12	14	18	15	17	23	12
184	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	13	30	24	11	13	13	11,14	12	12	14	18	15	17	23	12
185	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	23	11	13	14	11,14	12	12	14	18	16	18	23	12
186	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	14	30	24	11	13	13	11,14	12	11	14	18	15	17	23	11
187	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	24	11	13	13	11,14	12	11	14	18	16	17	23	12
188	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	24	11	13	13	11,14	12	11	14	18	16	17	23	12
189	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	24	11	13	13	12,14	12	11	14	18	16	17	23	12
190	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	24	10	13	13	11,14	12	12	14	18	16	17	23	12
191	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	24	11	13	14	11,14	12	12	14	18	16	17	23	11
192	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	24	11	13	14	11,14	12	12	14	18	16	17	23	12
193	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	24	11	13	14	11,14	12	12	14	18	16	17	23	12

Attached Table 1. Continuation.

ID	Characteristics	Final Haplogroup	M269	L11	U106	S116	DF27	U152	M529	L238	DF19	DYS19	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	DYS385	DYS438	DYS439	DYS437	DYS448	DYS456	DYS458	DYS635	YGATAH4
194	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	14	30	24	11	13	13	11,14	12	11	14	18	16	17	23	11
195	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	14	30	24	11	13	13	11,14	12	11	14	18	17	17	23	11
197	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	13	30	23	11	13	13	11,15	12	12	15	19	15	18	21	12
198	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	13	30	24	11	13	13	11,14	12	12	15	19	16	17	23	13
199	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	12	28	23	11	13	14	11,14	12	13	14	18	16	17	23	11
200	Rural_Native	DF27	1	1	0	1	1	0	0	0	0	14	14	30	25	11	13	13	11,15	12	11	14	18	17	18	23	11
201	Urban_native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	24	11	13	13	13,14	12	11	15	19	17	17	23	12
202	Urban_native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	24	11	13	13	11,14	12	11	15	19	16	16	23	12
203	Urban_native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	24	11	13	14	12,14	12	12	15	19	15	16	23	13
204	Urban_native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	24	11	13	13	11,14	12	12	15	19	15	17	23	12
205	Urban_native	DF27	1	1	0	1	1	0	0	0	0	14	14	30	25	10	13	12	11,13	12	13	15	19	15	17	23	12
206	Urban_native	DF27	1	1	0	1	1	0	0	0	0	14	12	29	24	11	13	13	11,14	12	13	15	19	16	16	23	12
207	Urban_native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	24	11	13	13	11,14	12	13	15	19	15	17	23	12
208	Urban_native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	24	10	13	12	11,13	12	13	15	19	15	18	23	12
209	Urban_native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	24	11	13	13	11,14	12	12	15	19	16	16	23	12
210	Urban_native	DF27	1	1	0	1	1	0	0	0	0	14	14	30	23	11	13	13	11,11	12	13	14	18	16	16	23	11
211	Urban_native	DF27	1	1	0	1	1	0	0	0	0	13	13	30	26	11	13	13	11,14	12	13	15	19	16	17	25	12
212	Urban_native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	24	10	13	14	11,15	12	11	15	19	15	17	23	13
213	Urban_native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	24	11	13	13	11,14	12	12	15	19	16	16	23	12
214	Urban_native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	24	11	13	13	11,15	12	12	15	19	15	18	23	11
215	Urban_native	DF27	1	1	0	1	1	0	0	0	0	14	14	30	24	11	13	13	13,14	12	12	14	18	16	16	23	11
216	Urban_native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	24	11	13	13	12,14	12	12	15	19	16	16	23	11
217	Urban_native	DF27	1	1	0	1	1	0	0	0	0	14	12	28	24	11	13	13	11,15	12	11	15	19	15	17	23	13
218	Urban_native	DF27	1	1	0	1	1	0	0	0	0	14	15	31	23	11	13	14	11,14	12	13	14	18	16	17	23	11

Attached Table 1. Continuation.

ID	Characteristics	Final Haplogroup	M269	L11	U106	S116	DF27	U152	M529	L238	DF19	DYS19	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	DYS385	DYS438	DYS439	DYS437	DYS448	DYS456	DYS458	DYS635	YGATAH4
219	Urban_native	DF27	1	1	0	1	1	0	0	0	0	14	14	30	23	11	13	14	11,14	12	13	14	18	16	17	23	11
220	Urban_native	DF27	1	1	0	1	1	0	0	0	0	14	14	30	24	11	13	13	11,14	12	12	14	18	16	17	23	11
221	Urban_native	DF27	1	1	0	1	1	0	0	0	0	14	14	30	24	11	13	13	11,14	12	11	14	18	15	17	23	11
222	Urban_native	DF27	1	1	0	1	1	0	0	0	0	14	14	30	24	10	13	13	11,14	12	11	14	18	15	17	23	11
223	Urban_native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	24	11	13	13	11,14	12	12	14	18	15	17	24	12
224	Urban_native	DF27	1	1	0	1	1	0	0	0	0	14	14	30	24	11	13	13	11,15	12	11	14	18	16	17	23	11
225	Urban_native	DF27	1	1	0	1	1	0	0	0	0	14	12	28	24	11	13	13	11,14	12	12	15	18	16	17	25	12
226	Urban_non native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	24	11	13	14	11,14	12	12	15	19	16	16	23	12
227	Urban_non native	DF27	1	1	0	1	1	0	0	0	0	14	14	30	24	10	13	13	11,14	12	11	15	19	15	17	23	11
228	Urban_non native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	24	11	13	13	11,14	12	12	15	18	15	17	23	12
229	Urban_non native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	24	10	13	13	11,14	10	11	14	19	16	17	23	12
230	Urban_non native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	23	11	13	13	11,14	12	12	15	19	16	16	23	12
231	Urban_non native	DF27	1	1	0	1	1	0	0	0	0	14	14	30	23	11	13	13	11,14	12	12	15	19	15	17	23	12
232	Urban_non native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	24	11	13	13	11,15	12	13	14	19	14	17	23	12
233	Urban_non native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	24	11	13	13	11,14	12	13	15	19	15	15	23	12
234	Urban_non native	DF27	1	1	0	1	1	0	0	0	0	14	12	28	24	11	13	13	10,14	12	12	15	19	16	16	23	12
235	Urban_non native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	24	11	13	12	11,14	12	12	15	19	15	18	23	12
236	Urban_non native	DF27	1	1	0	1	1	0	0	0	0	15	14	31	24	11	13	14	11,14	12	11	15	20	15	17	23	12
237	Urban_non native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	24	10	13	13	11,15	12	12	15	19	15	17	23	12
238	Urban_non native	DF27	1	1	0	1	1	0	0	0	0	14	13	30	24	11	13	13	11,16	12	11	14	18	15	17	23	11
239	Urban_non native	DF27	1	1	0	1	1	0	0	0	0	14	14	30	23	10	13	13	11,11	12	13	15	19	17	17	23	12
240	Urban_non native	DF27	1	1	0	1	1	0	0	0	0	14	14	30	24	11	13	13	11,14	12	13	15	18	16	19	23	12
241	Urban_non native	DF27	1	1	0	1	1	0	0	0	0	14	14	30	24	10	13	13	11,14	12	11	15	19	16	17	23	11
242	Urban_non native	DF27	1	1	0	1	1	0	0	0	0	14	13	30	23	11	13	13	11,14	12	11	15	19	15	17	24	12

Attached Table 1. Continuation.

ID	Characteristics	Final Haplogroup	M269	L11	U106	S116	DF27	U152	M529	L238	DF19	DYS19	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	DYS385	DYS438	DYS439	DYS437	DYS448	DYS456	DYS458	DYS635	YGATAH4
243	Urban_non native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	23	11	13	14	11,16	12	12	15	19	15	16	23	12
244	Urban_non native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	25	11	13	13	11,14	12	12	15	19	15	17	23	11
245	Urban_non native	DF27	1	1	0	1	1	0	0	0	0	14	14	30	23	10	13	13	11,11	12	13	15	19	18	17	23	12
246	Urban_non native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	24	10	13	13	11,15	12	12	15	19	15	17	23	13
247	Urban_non native	DF27	1	1	0	1	1	0	0	0	0	14	12	27	24	10	13	13	10,16	12	12	15	19	17	17	23	11
248	Urban_non native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	24	11	13	13	11,14	12	12	15	19	16	17	23	13
249	Urban_non native	DF27	1	1	0	1	1	0	0	0	0	14	14	30	24	13	13	13	11,14	12	13	15	19	15	18	23	12
250	Urban_non native	DF27	1	1	0	1	1	0	0	0	0	15	14	31	24	11	13	13	11,13	12	12	15	19	16	18	23	11
251	Urban_non native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	24	10	13	13	11,14	12	11	15	19	16	18	23	12
252	Urban_non native	DF27	1	1	0	1	1	0	0	0	0	14	14	30	25	11	13	13	11,14	12	12	14	18	15	18	23	11
253	Urban_non native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	24	11	13	13	11,14	12	11	15	19	16	16	23	12
254	Urban_non native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	24	10	13	13	11,14	12	12	15	19	15	15	23	12
255	Urban_non native	DF27	1	1	0	1	1	0	0	0	0	14	13	28	24	11	13	13	12,14	12	13	15	19	16	18	24	13
256	Urban_non native	DF27	1	1	0	1	1	0	0	0	0	14	13	28	24	10	13	13	11,13	12	12	15	19	16	17	23	12
257	Urban_non native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	24	10	13	15	12,15	12	12	15	19	16	18	23	12
258	Urban_non native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	24	11	13	13	11,14	11	11	15	19	15	16	23	12
259	Urban_non native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	24	11	13	13	11,15	12	11	15	19	15	18	23	12
260	Urban_non native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	24	10	13	13	11,14	12	11	15	19	16	16	24	11
261	Urban_non native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	24	11	13	13	11,14	12	12	15	19	16	17	23	12
262	Urban_non native	DF27	1	1	0	1	1	0	0	0	0	14	14	30	24	10	13	13	11,13	12	12	14	18	16	17	24	11
263	Urban_non native	DF27	1	1	0	1	1	0	0	0	0	14	15	32	24	11	13	13	11,14	12	12	14	18	15	17	24	11
264	Urban_non native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	24	11	13	13	11,16	12	13	14	18	16	16	23	11
265	Urban_non native	DF27	1	1	0	1	1	0	0	0	0	14	14	30	22	11	14	13	11,14	12	12	14	18	15	17	23	11
266	Urban_non native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	24	12	13	13	11,14	12	13	15	18	16	18	24	11

Attached Table 1. Continuation.

ID	Characteristics	Final Haplogroup	M269	L11	U106	S116	DF27	U152	M529	L238	DF19	DYS19	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	DYS385	DYS438	DYS439	DYS437	DYS448	DYS456	DYS458	DYS635	YGATAH4
267	Urban_non native	DF27	1	1	0	1	1	0	0	0	0	14	13	30	24	11	13	13	11,14	12	13	14	18	15	17	23	11
268	Urban_non native	DF27	1	1	0	1	1	0	0	0	0	14	14	30	24	10	13	13	11,15	12	12	14	18	16	18	24	11
269	Urban_non native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	24	12	13	13	11,14	12	13	15	18	16	18	24	11
270	Urban_non native	DF27	1	1	0	1	1	0	0	0	0	14	14	31	24	11	13	13	11,13	12	12	14	18	15	17	23	11
271	Urban_non native	DF27	1	1	0	1	1	0	0	0	0	14	14	30	24	11	13	13	12,15	12	11	14	18	15	17	23	11
272	Urban_non native	DF27	1	1	0	1	1	0	0	0	0	14	13	29	24	11	12	13	11,12	12	12	14	18	16	16	23	11
273	Urban_non native	DF27	1	1	0	1	1	0	0	0	0	14	14	31	24	11	13	13	11,14	12	12	14	18	16	16	23	11
274	Urban_non native	DF27	1	1	0	1	1	0	0	0	0	14	14	30	24	10	14	13	12,14	12	12	14	18	15	17	23	12
275	Urban_non native	DF27	1	1	0	1	1	0	0	0	0	14	13	30	23	11	13	13	11,14	11	11	15	19	16	18	23	12

Attached Table 2. Y-SNP and Y-STR haplotypes for all the analyzed samples of population. Corresponds to Supplementary Table S3 from *Study Number 2*.

ID	Original code	Population	S116	DF27	L617	L881	DF17	Z195	Z196	L176.2	S68	M167	Z220	Z278	M153	Final Haplogroup	DYS19	DYS389	389I	389II	DYS390	DYS391	DYS392	DYS393	DYS438	DYS439	DYS437	DYS448	DYS456	DYS458	DYS635	GATAH4
S1	AST1	Asturias	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	12	28	24	11	13	13	12	12	15	19	16	15	23	12
S2	AST10	Asturias	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	13	29	24	12	13	13	13	14	14	20	15	17	23	12
S3	AST16	Asturias	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	13	29	24	11	13	13	12	12	15	18	16	15	23	12
S4	AST2	Asturias	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	13	29	24	10	13	13	12	12	14	19	16	17	23	13
S5	AST20	Asturias	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	17	13	30	24	11	13	13	12	13	15	19	16	16	23	12
S6	AST21	Asturias	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	14	30	24	11	13	13	12	12	15	19	16	17	23	12
S7	AST23	Asturias	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	13	29	23	10	13	13	12	11	15	18	15	17	23	12
S8	AST24	Asturias	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	14	30	25	10	13	13	12	13	15	19	16	17	23	12
S9	AST36	Asturias	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	13	29	24	11	13	13	12	12	15	19	16	16	23	11
S10	AST46	Asturias	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	15	13	28	23	11	13	13	12	12	15	19	15	18	23	12
S11	AST51	Asturias	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	15	16	13	29	24	11	12	14	12	11	15	19	15	17	23	12
S12	AST53	Asturias	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	17	13	30	22	11	13	13	13	13	15	19	15	17	23	12
S13	AST54	Asturias	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	13	16	13	29	24	11	13	13	12	12	15	19	16	17	23	13
S14	AST58	Asturias	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	14	30	24	11	13	13	11	13	14	19	16	17	23	12
S15	AST60	Asturias	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	13	29	24	11	13	13	12	12	15	19	16	16	23	12
S16	AST61	Asturias	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	13	29	24	11	13	13	12	12	15	20	15	17	23	12
S17	AST64	Asturias	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	17	13	30	24	11	13	13	12	12	15	19	15	17	23	12
S18	AST8	Asturias	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	15	13	28	24	10	13	13	12	13	14	19	15	16	24	12
S19	AST9	Asturias	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	13	29	24	11	13	13	12	12	15	19	15	17	23	12
S20	AST30	Asturias	1	1	0	0	0	1	1	0	0	0	0	0	0	Z196p	14	16	13	29	24	11	13	13	12	13	15	19	16	16	23	11
S21	AST35	Asturias	1	1	0	0	0	1	1	0	0	0	0	0	0	Z196p	14	17	14	31	24	11	13	13	12	11	15	19	15	17	23	12
S22	AST37	Asturias	1	1	0	0	0	1	1	0	0	0	0	0	0	Z196p	14	16	13	29	25	11	13	13	12	12	15	21	16	17	23	12
S23	AST55	Asturias	1	1	0	0	0	1	1	0	0	0	0	0	0	Z196p	14	15	12	27	24	11	13	13	12	13	15	19	15	16	23	12
S24	AST25	Asturias	1	1	0	0	0	1	1	0	0	0	1	0	0	Z220	14	16	14	30	24	11	13	13	12	12	14	18	16	20	23	11
S25	AST62	Asturias	1	1	0	0	0	1	1	0	0	0	1	1	0	Z278	14	17	14	31	24	11	13	13	12	12	14	18	15	15	23	11
S26	AST65	Asturias	1	1	0	0	0	1	1	0	0	0	1	1	0	Z278	14	16	14	30	24	11	13	13	12	12	14	18	15	15	23	11
	AST5	Asturias	1	1	0	0	0	1	1	0	0	0	0	0	0																	
S289	AST11	Asturias	0	0	0	0	0	0	0	0	0	0	0	0	0		15	16	23	11	11	12	9	12	16	20	13	18	14	16	21	11
S290	AST12	Asturias	0	0	0	0	0	0	0	0	0	0	0	0	0		17	15	25	9	11	13	10	13	15	21	14	16	12	12	22	11
S291	AST13	Asturias	0	0	0	0	0	0	0	0	0	0	0	0	0		15	17	23	10	13	13	9	11	14	19	16	17	14	16	21	11

Attached Table 2. Continuation.

ID	Original code	Population	S116	DF27	L617	L881	DF17	Z195	Z196	L176.2	S68	M167	Z220	Z278	M153	Final Haplogroup	DYS19	DYS389	389I	389II	DYS390	DYS391	DYS392	DYS393	DYS438	DYS439	DYS437	DYS448	DYS456	DYS458	DYS635	GATAH4
S292	AST14	Asturias	0	0	0	0	0	0	0	0	0	0	0	0	0		14	16	22	10	14	11	10	12	15	20	15	15	13	17	22	12
S293	AST17	Asturias	0	0	0	0	0	0	0	0	0	0	0	0	0		13	16	24	9	11	13	10	10	14	20	16	18	12	14	21	12
S294	AST18	Asturias	0	0	0	0	0	0	0	0	0	0	0	0	0		14	16	23	10	11	12	9	11	14	21	15	17	13	17	21	11
S295	AST19	Asturias	0	0	0	0	0	0	0	0	0	0	0	0	0		13	17	23	10	11	13	10	13	14	19	16	16	16	16	21	11
S296	AST26	Asturias	0	0	0	0	0	0	0	0	0	0	0	0	0		14	16	22	10	11	13	10	12	16	20	15	15	13	14	22	11
S297	AST27	Asturias	0	0	0	0	0	0	0	0	0	0	0	0	0		15	17	22	10	11	14	10	11	16	22	16	21	13	15	21	11
S298	AST28	Asturias	0	0	0	0	0	0	0	0	0	0	0	0	0		15	16	24	9	11	12	9	11	14	21	16	15	12	17	23	12
S299	AST29	Asturias	0	0	0	0	0	0	0	0	0	0	0	0	0		14	16	23	10	11	12	9	11	14	21	15	18	12	17	22	11
S300	AST3	Asturias	0	0	0	0	0	0	0	0	0	0	0	0	0		13	18	24	10	11	13	10	12	14	20	15	17	16	18	24	11
S301	AST31	Asturias	0	0	0	0	0	0	0	0	0	0	0	0	0		15	17	23	10	14	13	9	11	14	20	15	17	14	18	21	12
S302	AST34	Asturias	0	0	0	0	0	0	0	0	0	0	0	0	0		13	17	24	10	11	13	10	12	14	20	16	17	18	18	21	12
S303	AST38	Asturias	0	0	0	0	0	0	0	0	0	0	0	0	0		13	17	24	11	11	12	11	12	14	20	15	19	16	17	25	12
S304	AST39	Asturias	0	0	0	0	0	0	0	0	0	0	0	0	0		15	17	23	10	11	12	9	11	15	20	14	16	13	14	22	13
S305	AST4	Asturias	0	0	0	0	0	0	0	0	0	0	0	0	0		17	15	23	10	11	13	10	12	15	21	14	18	12	12	22	12
S306	AST41	Asturias	0	0	0	0	0	0	0	0	0	0	0	0	0		13	16	23	10	11	14	10	13	14	21	15	15	17	17	21	11
	AST42	Asturias	0	0	0	0	0	0	0	0	0	0	0	0	0																	
S307	AST43	Asturias	0	0	0	0	0	0	0	0	0	0	0	0	0		15	16	22	10	11	12	9	12	15	21	15	15	13	15	21	11
S308	AST44	Asturias	0	0	0	0	0	0	0	0	0	0	0	0	0		14	16	23	10	11	12	10	11	16	20	14	15	13	15	21	11
S309	AST48	Asturias	0	0	0	0	0	0	0	0	0	0	0	0	0		14	16	24	10	11	13	10	11	16	19	14	16	13	14	22	10
S310	AST49	Asturias	0	0	0	0	0	0	0	0	0	0	0	0	0		15	17	23	10	11	14	10	11	16	21	15	14	14	15	23	12
S311	AST50	Asturias	0	0	0	0	0	0	0	0	0	0	0	0	0		15	17	21	10	11	14	10	11	16	21	16	16	13	15	21	11
S312	AST52	Asturias	0	0	0	0	0	0	0	0	0	0	0	0	0		15	16	24	10	11	12	9	11	15	19	13	18	14	14	21	11
S313	AST56	Asturias	0	0	0	0	0	0	0	0	0	0	0	0	0		15	18	23	10	13	13	9	11	14	19	15	17	14	16	21	11
S314	AST7	Asturias	0	0	0	0	0	0	0	0	0	0	0	0	0		15	18	24	10	11	13	10	11	14	20	18	15	19	19	20	12
S315	AST15	Asturias	1	0	0	0	0	0	0	0	0	0	0	0	0		14	16	24	12	13	13	12	11	14	18	15	17	12	14	23	12
S316	AST22	Asturias	1	0	0	0	0	0	0	0	0	0	0	0	0		14	16	25	10	13	14	12	12	15	19	16	17	11	14	23	12
S317	AST32	Asturias	1	0	0	0	0	0	0	0	0	0	0	0	0		15	17	23	11	13	13	12	11	15	19	15	17	11	13	23	12
S318	AST33	Asturias	1	0	0	0	0	0	0	0	0	0	0	0	0		14	15	24	11	13	13	12	12	15	19	15	17	11	14	22	12
S319	AST47	Asturias	1	0	0	0	0	0	0	0	0	0	0	0	0		14	16	25	10	13	14	12	12	15	19	16	17	11	14	23	12
S320	AST57	Asturias	1	0	0	0	0	0	0	0	0	0	0	0	0		14	16	23	11	13	13	12	12	15	19	15	15	11	14	24	13

Attached Table 2. Continuation.

ID	Original code	Population	S116	DF27	L617	L881	DF17	Z195	Z196	L176.2	S68	M167	Z220	Z278	M153	Final Haplogroup	DYS19	DYS389	389I	389II	DYS390	DYS391	DYS392	DYS393	DYS438	DYS439	DYS437	DYS448	DYS456	DYS458	DYS635	GATAH4
S321	AST59	Asturias	1	0	0	0	0	0	0	0	0	0	0	0	0		14	16	23	11	13	13	12	13	15	19	15	15	11	14	24	12
S322	AST6	Asturias	1	0	0	0	0	0	0	0	0	0	0	0	0		14	16	24	11	13	13	12	11	14	18	15	17	12	14	23	12
S323	AST63	Asturias	1	0	0	0	0	0	0	0	0	0	0	0	0		14	16	24	11	13	13	12	12	15	20	14	16	11	14	24	12
C27	291x	Cantabria	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	14	30	23	12	13	13	12	12	15	19	15	17	22	12
C28	308x	Cantabria	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	13	29	24	10	13	13	11	11	15	19	15	17	23	12
C29	a390	Cantabria	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	13	29	25	10	13	13	12	12	15	19	16	19	23	11
C30	a395	Cantabria	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	15	16	13	29	24	11	13	13	12	12	15	19	19	18	24	11
C31	a455	Cantabria	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	17	14	31	24	11	13	13	12	13	15	19	15	17	23	12
C32	a457	Cantabria	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	13	29	23	10	14	13	12	12	15	19	15	17	23	12
C33	a464	Cantabria	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	15	16	13	29	25	11	13	13	12	12	14	18	15	17	23	12
C34	a465	Cantabria	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	14	30	24	10	13	13	12	12	14	18	16	17	24	11
C35	a466	Cantabria	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	13	29	24	11	13	14	12	13	15	18	16	12	23	11
C36	a471	Cantabria	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	13	29	22	10	13	12	12	13	15	19	15	17	24	13
C37	a477	Cantabria	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	13	29	24	11	13	13	12	12	15	19	15	18	23	12
C38	a478	Cantabria	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	17	14	31	24	10	13	13	12	11	14	18	15	18	23	11
C39	a480	Cantabria	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	17	16	12	28	23	11	13	13	12	13	15	19	16	16	23	11
C40	a491	Cantabria	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	12	28	24	11	13	13	12	12	14	18	16	18	23	11
C41	a492	Cantabria	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	14	30	25	11	13	13	12	12	15	18	15	17	23	12
C42	a493	Cantabria	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	13	29	25	11	13	13	12	12	14	18	16	16	24	11
C43	a496	Cantabria	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	13	29	24	11	13	14	12	14	15	19	17	17	23	12
C44	a521	Cantabria	1	1	0	0	0	1	1	1	0	0	0	0	0	L176.2	14	16	13	29	24	11	13	13	12	14	15	19	16	17	23	12
C45	a498	Cantabria	1	1	0	0	0	1	1	0	0	0	1	0	0	Z220	14	17	13	30	23	11	13	13	12	13	15	20	16	17	23	11
C46	a501	Cantabria	1	1	0	0	0	1	1	0	0	0	1	1	0	Z278	14	16	13	29	24	10	13	13	12	11	15	20	16	18	23	12
C47	a503	Cantabria	1	1	0	0	0	1	1	0	0	0	1	0	0	Z220	14	16	13	29	25	11	13	13	12	12	15	19	16	19	24	11
C48	a504	Cantabria	1	1	0	0	0	1	1	0	0	0	1	1	0	Z278	15	16	13	29	24	12	13	13	12	13	15	18	15	16	24	12
C49	a505	Cantabria	1	1	0	0	0	1	1	0	0	0	1	1	0	Z278	14	16	13	29	25	10	13	13	12	12	15	19	16	**17**	24	12
C50	a514	Cantabria	1	1	0	0	0	1	1	0	0	0	1	1	0	Z278	14	16	14	30	25	11	13	13	12	12	14	18	15	18	23	11
C51	a515	Cantabria	1	1	0	0	0	1	1	0	0	0	1	0	0	Z220	14	16	14	30	25	11	12	13	12	11	14	18	15	17	23	12
C52	a518	Cantabria	1	1	0	0	0	1	1	0	0	0	1	0	0	Z220	14	16	13	29	24	12	13	13	12	12	15	20	16	17	23	12
C53	a520	Cantabria	1	1	0	0	0	1	1	0	0	0	1	1	0	Z278	14	16	14	30	24	10	13	13	12	12	14	18	16	17	25	11

Attached Table 2. Continuation.

ID	Original code	Population	S116	DF27	L617	L881	DF17	Z195	Z196	L176.2	S68	M167	Z220	Z278	M153	Final Haplogroup	DYS19	DYS389	389I	389II	DYS390	DYS391	DYS392	DYS393	DYS438	DYS439	DYS437	DYS448	DYS456	DYS458	DYS635	GATAH4
C54	a523	Cantabria	1	1	0	0	0	1	1	0	0	0	1	1	0	Z278	14	16	13	29	24	11	13	13	12	12	15	19	15	18	23	12
C55	a541	Cantabria	1	1	0	0	0	1	1	0	0	0	1	1	0	Z278	14	16	13	29	24	10	13	13	12	12	15	19	15	18	23	12
C56	a548	Cantabria	1	1	0	0	0	1	1	0	0	0	1	1	0	Z278	14	16	13	29	23	11	13	13	12	12	14	18	15	17	23	11
	a468	Cantabria	1	1	0	0	0	0	0	0	0	0	0	0	0																	
	a484	Cantabria	1	1	0	0	0	0	0	0	0	0	0	0	0																	
	a513	Cantabria	1	1	0	0	0	1	1	0	0	0	0	0	0																	
	a516	Cantabria	1	1	0	0	0	1	1	0	0	0	0	0	0																	
	a508	Cantabria	1	1	0	0	0	1	1	1	0	1	0	0	0																	
	a522	Cantabria	1	1	0	0	0	1	1	1	0	1	0	0	0																	
	a500	Cantabria	1	1	0	0	0	1	1	0	0	0	1	1	0																	
	a502	Cantabria	1	1	0	0	0	1	1	0	0	0	1	1	0																	
	a519	Cantabria	1	1	0	0	0	1	1	0	0	0	1	1	0																	
	a532	Cantabria	1	1	0	0	0	1	1	0	0	0	1	1	0																	
C324	a226	Cantabria	0	0	0	0	0	0	0	0	0	0	0	0	0		14	16	22	10	11	13	10	11	16	20	12	15	13	15	22	11
C325	a237	Cantabria	0	0	0	0	0	0	0	0	0	0	0	0	0		13	17	23	10	11	13	10	13	14	20	17	15	16	19	22	12
C326	a294	Cantabria	0	0	0	0	0	0	0	0	0	0	0	0	0		14	17	24	10	13	12	10	11	14	20	17	17.2	17	18	23	10
C327	a301	Cantabria	0	0	0	0	0	0	0	0	0	0	0	0	0		13	16	24	9	11	13	10	10	14	20	16	17	14	16	21	12
C328	a335	Cantabria	0	0	0	0	0	0	0	0	0	0	0	0	0		14	16	22	10	11	13	10	11	16	20	14	15	14	14	21	11
C329	a337	Cantabria	0	0	0	0	0	0	0	0	0	0	0	0	0		15	17	23	10	13	13	9	11	14	19	16	18	14	16	21	11
C330	a340	Cantabria	0	0	0	0	0	0	0	0	0	0	0	0	0		14	16	22	11	11	13	10	11	16	20	14	15	12	13	22	11
C331	a347	Cantabria	0	0	0	0	0	0	0	0	0	0	0	0	0		15	17	25	10	11	12	11	10	14	20	16	15	11	14	23	12
C332	a350	Cantabria	0	0	0	0	0	0	0	0	0	0	0	0	0		13	16	24	9	11	13	10	10	14	19	16	19	13	14	21	12
C333	a354	Cantabria	0	0	0	0	0	0	0	0	0	0	0	0	0		13	17	24	10	11	12	10	12	14	20	15	17.2	12	17	21	10
	a358	Cantabria	0	0	0	0	0	0	0	0	0	0	0	0	0																	
	a366	Cantabria	0	0	0	0	0	0	0	0	0	0	0	0	0																	
	a394	Cantabria	1	1	0	0	0	0	0	0	0	0	0	0	0																	
	a499	Cantabria	1	1	0	0	0	0	0	0	0	0	0	0	0																	
	a506	Cantabria	1	1	0	0	0	0	0	0	0	0	0	0	0																	
C339	a290	Cantabria	0	0	0	0	0	0	0	0	0	0	0	0	0		14	16	24	10	13	13	12	12	16	18.2	15	18	11	14	23	13
C340	a302	Cantabria	0	0	0	0	0	0	0	0	0	0	0	0	0		15	16	24	11	13	13	12	12	15	19	15	16	11	14	23	12

Attached Table 2. Continuation.

ID	Original code	Population	S116	DF27	L617	L881	DF17	Z195	Z196	L176.2	S68	M167	Z220	Z278	M153	Final Haplogroup	DYS19	DYS389	389I	389II	DYS390	DYS391	DYS392	DYS393	DYS438	DYS439	DYS437	DYS448	DYS456	DYS458	DYS635	GATAH4
C341	a345	Cantabria	0	0	0	0	0	0	0	0	0	0	0	0	0		14	16	24	11	13	13	12	13	15	19	15	17	11	15	23	12
C342	a351	Cantabria	0	0	0	0	0	0	0	0	0	0	0	0	0		14	16	24	10	13	13	12	11	15	19	16	17	11	14	23	12
C343	a355	Cantabria	0	0	0	0	0	0	0	0	0	0	0	0	0		15	17	23	11	12	14	10	11	14	20	15	15	15	16	21	11
C344	a360	Cantabria	0	0	0	0	0	0	0	0	0	0	0	0	0		14	17	24	11	13	13	12	11	14	18	15	17	11	15	23	11
C345	a361	Cantabria	0	0	0	0	0	0	0	0	0	0	0	0	0		13	17	24	10	11	13	10	12	14	21	17	15	16	17	24	12
C346	a369	Cantabria	0	0	0	0	0	0	0	0	0	0	0	0	0		14	17	24	10	13	13	12	11	15	19	15	18	11	15	23	12
C347	a370	Cantabria	0	0	0	0	0	0	0	0	0	0	0	0	0		13	16	24	9	11	13	10	10	14	20	14	19	13	14	21	11
C348	a352	Cantabria	1	0	0	0	0	0	0	0	0	0	0	0	0		14	16	24	11	13	13	12	13	15	19	15	18	11	14	23	10
C349	a386	Cantabria	1	0	0	0	0	0	0	0	0	0	0	0	0		14	17	24	11	13	13	12	12	14	18	16	20	11	14	23	12
C350	a389	Cantabria	1	0	0	0	0	0	0	0	0	0	0	0	0		14	16	24	11	13	13	12	12	14	18	15	17	12	14	23	11
C351	a392	Cantabria	1	0	0	0	0	0	0	0	0	0	0	0	0		15	15	24	10	11	13	10	12	15	20	15	18	12	12	20	13
C352	a396	Cantabria	1	0	0	0	0	0	0	0	0	0	0	0	0		17	15	23	10	11	13	10	11	15	21	14	17	12	12	22	12
C353	a453	Cantabria	1	0	0	0	0	0	0	0	0	0	0	0	0		14	17	24	11	13	13	12	12	15	19	15	17	11	14	23	13
C354	a454	Cantabria	1	0	0	0	0	0	0	0	0	0	0	0	0		15	17	23	10	13	13	9	11	14	19	15	18	14	16	21	11
C355	a463	Cantabria	1	0	0	0	0	0	0	0	0	0	0	0	0		15	17	22	10	11	14	10	12	16	21	17	16	14	15	21	12
C356	a469	Cantabria	1	0	0	0	0	0	0	0	0	0	0	0	0		13	16	24	11	13	13	12	12	15	20	17	18	11	14	23	12
C357	a470	Cantabria	1	0	0	0	0	0	0	0	0	0	0	0	0		14	16	24	10	11	13	12	11	15	19	17	18	11	14	24	14
C358	a475	Cantabria	1	0	0	0	0	0	0	0	0	0	0	0	0		14	15	24	11	13	13	12	12	15	19	16	17	11	15	23	12
C359	a476	Cantabria	1	0	0	0	0	0	0	0	0	0	0	0	0		14	16	24	12	13	13	12	12	14	18	17	18	11	15	23	11
C360	a483	Cantabria	1	0	0	0	0	0	0	0	0	0	0	0	0		14	17	24	11	13	13	12	12	15	19	15	13	11	14	24	12
C361	a485	Cantabria	1	0	0	0	0	0	0	0	0	0	0	0	0		15	17	24	11	13	12	12	11	15	19	15	17	11	14	23	12
C362	a636	Cantabria	1	0	0	0	0	0	0	0	0	0	0	0	0		14	17	24	10	13	13	12	12	15	19	17	17	11	14	23	12
	a292	Cantabria	0	0	0	0	0	0	0	0	0	0	0	0	0																	
	a348	Cantabria	0	0	0	0	0	0	0	0	0	0	0	0	0																	
	a349	Cantabria	0	0	0	0	0	0	0	0	0	0	0	0	0																	
	a353	Cantabria	0	0	0	0	0	0	0	0	0	0	0	0	0																	
	a356	Cantabria	0	0	0	0	0	0	0	0	0	0	0	0	0																	
	a367	Cantabria	0	0	0	0	0	0	0	0	0	0	0	0	0																	
	a368	Cantabria	0	0	0	0	0	0	0	0	0	0	0	0	0																	
	a371	Cantabria	0	0	0	0	0	0	0	0	0	0	0	0	0																	

Attached Table 2. Continuation.

ID	Original code	Population	S116	DF27	L617	L881	DF17	Z195	Z196	L176.2	S68	M167	Z220	Z278	M153	Final Haplogroup	DYS19	DYS389	389I	389II	DYS390	DYS391	DYS392	DYS393	DYS438	DYS439	DYS437	DYS448	DYS456	DYS458	DYS635	GATAH4
	a372	Cantabria	0	0	0	0	0	0	0	0	0	0	0	0	0																	
	a373	Cantabria	0	0	0	0	0	0	0	0	0	0	0	0	0																	
	a374	Cantabria	0	0	0	0	0	0	0	0	0	0	0	0	0																	
	a376	Cantabria	0	0	0	0	0	0	0	0	0	0	0	0	0																	
	a377	Cantabria	0	0	0	0	0	0	0	0	0	0	0	0	0																	
	a378	Cantabria	0	0	0	0	0	0	0	0	0	0	0	0	0																	
	a380	Cantabria	0	0	0	0	0	0	0	0	0	0	0	0	0																	
	a456	Cantabria	1	0	0	0	0	0	0	0	0	0	0	0	0																	
	a517	Cantabria	1	0	0	0	0	0	0	0	0	0	0	0	0																	
V57	1ftvAPa	Native Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	17	13	30	24	11	13	13	12	11	15	19	16	17	23	12
V58	3vf2sPa	Native Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	13	29	24	11	13	13	12	12	15	19	16	16	23	12
V59	5s7zmPa	Native Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	13	29	25	11	13	13	12	12	15	19	15	17	23	12
V60	94783Pa	Native Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	12	28	24	11	13	13	12	12	15	19	16	16	23	11
V61	ART008	Native Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	12	28	24	11	13	13	12	12	15	19	16	16	23	12
V62	ART017	Native Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	12	28	24	11	13	13	12	12	15	19	16	16	23	12
V63	ART020	Native Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	14	30	24	11	13	13	12	12	15	19	15	17	23	12
V64	ART027	Native Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	12	28	25	11	13	13	12	12	15	19	15	17	23	12
V65	eLr02Pa	Native Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	14	30	23	10	13	13	12	11	15	19	17	19	23	12
V66	ERE003	Native Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	15	16	14	30	23	10	13	13	12	13	15	19	18	18	23	13
V67	ERE015	Native Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	13	29	25	10	13	12	12	13	15	19	15	17	23	13
V68	ERE030	Native Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	17	13	30	24	10	13	13	12	12	14	19	16	17	23	13
V69	ERE033	Native Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	14	30	24	10	13	13	12	12	15	19	18	17	23	12
V70	G85e 108	Native Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	13	29	24	11	13	13	12	11	15	19	17	17	23	12
V71	G85e 14	Native Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	13	29	24	10	13	13	12	12	15	19	15	16	23	11
V72	G85e 168	Native Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	13	29	24	11	13	14	12	12	15	19	15	16	23	13
V73	G85e 171	Native Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	14	30	25	10	13	12	12	13	15	19	15	17	23	12
V74	G85e 211	Native Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	17	12	29	24	11	13	13	12	13	15	19	16	16	23	12
V75	G85e 212	Native Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	13	29	24	11	13	13	12	13	15	19	15	17	23	12
V76	G85e 77	Native Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	13	29	24	10	13	12	12	13	15	19	15	18	23	12
V77	G85e 90	Native Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	14	30	23	11	13	13	12	13	14	18	16	16	23	11

Attached Table 2. Continuation.

ID	Original code	Population	S116	DF27	L617	L881	DF17	Z195	Z196	L176.2	S68	M167	Z220	Z278	M153	Final Haplogroup	DYS19	DYS389	389I	389II	DYS390	DYS391	DYS392	DYS393	DYS438	DYS439	DYS437	DYS448	DYS456	DYS458	DYS635	GATAH4
V78	G85e 96	Native Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	13	17	13	30	26	11	13	13	12	13	15	19	16	17	25	12
V79	G85e 98	Native Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	13	29	24	10	13	14	12	11	15	19	15	17	23	13
V80	GhXOGPa	Native Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	14	30	23	10	13	13	12	13	15	19	17	18	23	12
V81	gK6XgPa	Native Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	14	30	23	10	13	13	12	13	15	19	17	18	23	12
V82	gou13	Native Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	13	29	24	11	13	13	12	12	15	19	15	17	23	12
V83	JvehsPa	Native Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	13	29	24	11	13	13	12	13	15	19	15	17	23	12
V84	KOR004	Native Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	17	13	30	24	11	13	13	12	12	15	19	16	17	23	13
V85	KOR019	Native Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	13	29	24	11	13	13	12	12	15	19	16	16	23	12
V86	KOR026	Native Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	13	29	24	11	13	13	12	12	15	19	15	17	23	12
V87	KOR027	Native Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	13	29	24	10	13	12	12	13	15	19	15	17	23	12
V88	KOR037	Native Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	13	29	24	10	13	13	12	13	15	19	15	16	23	12
V89	kWFiT	Native Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	13	29	24	11	13	13	12	12	15	19	15	17	23	12
V90	NAB002	Native Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	13	29	24	11	13	13	12	12	15	19	16	16	23	12
V91	NAB005	Native Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	17	12	29	24	11	13	13	12	13	15	19	16	16	23	12
V92	NAB006	Native Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	17	12	29	24	11	13	13	12	13	15	19	16	16	23	12
V93	NAB014	Native Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	13	29	24	11	13	13	12	12	14	19	15	16	23	12
V94	NAB022	Native Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	12	28	24	11	13	13	12	12	15	19	15	16	23	12
V95	P7Uz5Pa	Native Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	13	29	25	10	13	12	12	13	15	20	16	17	23	12
V96	PrG3PPa	Native Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	13	29	24	11	13	13	12	12	15	19	15	17	23	12
V97	q6hKQPa	Native Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	14	30	25	11	13	13	12	12	15	19	15	17	23	11
V98	qBln4	Native Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	13	29	24	11	13	13	12	12	15	19	15	17	23	12
V99	TX 2405	Native Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	13	29	24	11	13	13	12	12	15	19	16	16	23	12
V100	TX 429	Native Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	13	29	24	11	13	13	12	12	15	19	15	18	23	11
V101	USA002CHIDV0	Native Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	13	29	24	10	14	13	12	11	15	19	16	18	23	12
V102	USA002RENDV0	Native Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	12	28	24	11	13	13	12	12	15	19	16	16	23	12
V103	USA003BOIDV0	Native Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	17	13	30	24	11	13	13	12	12	15	19	16	18	23	12
V104	USA022CHIDV0	Native Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	12	28	24	11	13	13	12	12	15	19	16	16	23	12
V105	USA031RENDV0	Native Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	13	29	24	11	13	13	10	12	15	19	16	15	23	12
V106	USA038RENDV0	Native Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	18	13	31	23	10	13	13	12	12	15	19	14	19	26	12
V107	USA040B0IDV0	Native Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	13	29	24	10	13	13	12	12	15	19	16	17	23	12

Attached Table 2. Continuation.

ID	Original code	Population	S116	DF27	L617	L881	DF17	Z195	Z196	L176.2	S68	M167	Z220	Z278	M153	Final Haplogroup	DYS19	DYS389	389I	389II	DYS390	DYS391	DYS392	DYS393	DYS438	DYS439	DYS437	DYS448	DYS456	DYS458	DYS635	GATAH4
V108	USA052BOIDV0	Native Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	12	28	24	11	13	13	12	11	15	19	16	16	23	12
V109	USA053BOIDV0	Native Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	12	28	24	11	14	14	12	12	14	19	17	16	23	12
V110	USA081BOIDV0	Native Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	17	13	30	24	11	13	13	12	12	15	19	16	17	23	13
V111	USA087BOIDV0	Native Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	17	13	30	24	11	13	13	12	12	15	19	16	17	23	13
V112	USA094BOIDV0	Native Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	13	29	24	11	13	13	12	12	15	19	15	17	23	12
V113	USA096BOIDV0	Native Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	12	28	23	11	13	13	12	12	15	19	16	16	23	12
V114	USA110BOIDV0	Native Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	13	29	23	11	13	13	12	12	15	19	16	16	23	12
V115	USA117BOIDV0	Native Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	13	29	24	11	13	13	12	12	15	19	16	17	23	12
V116	USA126BOIDV0	Native Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	14	30	23	10	13	13	12	13	15	19	16	18	23	12
V117	USA143BOIDV0	Native Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	13	29	24	11	13	13	12	12	15	19	15	18	23	12
V118	USA154BOIDV0	Native Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	13	29	24	11	13	13	12	12	15	19	15	17	23	12
V119	USA172BOIDV0	Native Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	13	29	25	11	11	13	12	12	15	19	17	19	23	12
V120	USA182BOIDV0	Native Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	17	13	30	24	11	13	13	12	12	15	19	15	16	23	12
V121	USA185BOIDV0	Native Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	17	13	30	24	10	13	13	12	12	15	19	16	17	23	13
V122	USA188BOIDV0	Native Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	17	13	30	24	11	13	13	12	12	14	19	16	17	23	12
V123	Y3TVGPa	Native Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	17	14	31	25	10	13	12	12	12	15	19	15	17	23	12
V124	ART024	Native Basques	1	1	0	0	0	1	1	0	0	0	0	0	0	Z196p	14	17	13	30	23	11	13	13	12	12	15	19	15	18	21	12
V125	ERE018	Native Basques	1	1	0	0	1	1	1	0	0	0	0	0	0	DF17	16	16	13	29	23	11	13	13	12	13	14	18	16	19	23	12
V126	G85e 103	Native Basques	1	1	0	0	0	1	1	0	0	0	0	0	0	Z196p	14	16	14	30	24	11	13	13	12	12	14	18	16	16	23	11
V127	G85e 222	Native Basques	1	1	0	0	0	1	1	0	0	0	0	0	0	Z196p	14	16	13	29	24	11	13	13	12	12	15	19	16	16	23	11
V128	NAB010	Native Basques	1	1	0	0	0	1	1	0	0	0	0	0	0	Z196p	14	17	13	30	24	11	13	13	12	12	15	19	16	17	23	13
V129	USA016RENDV0	Native Basques	1	1	0	0	0	1	1	0	0	0	0	0	0	Z196p	14	15	12	27	24	11	13	13	12	12	15	19	15	16	23	12
V130	USA139BOIDVO	Native Basques	1	1	0	0	0	1	1	0	0	0	0	0	0	Z196p	14	16	13	29	24	11	13	13	12	12	15	19	16	16	23	12
V131	G85e 91	Native Basques	1	1	0	0	0	1	1	1	0	0	0	0	0	L176.2p	14	16	12	28	24	11	13	13	12	11	15	19	15	17	23	13
V132	HSCEPPa	Native Basques	1	1	0	0	0	1	1	1	0	0	0	0	0	L176.2p	14	16	13	29	24	11	13	13	13	13	15	19	16	15	23	12
V133	juegPa	Native Basques	1	1	0	0	0	1	1	1	0	0	0	0	0	L176.2p	14	18	14	32	24	11	13	13	12	12	15	20	16	17	23	12
V134	USA026BOIDV0	Native Basques	1	1	0	0	0	1	1	1	0	0	0	0	0	L176.2p	14	16	14	30	23	11	13	14	12	12	14	19	17	17	23	12
V135	USA079BOIDV0	Native Basques	1	1	0	0	0	1	1	1	1	0	0	0	0	S68	14	16	13	29	24	10	13	13	12	12	15	19	15	18	23	12
V136	USA158BOIDV0	Native Basques	1	1	0	0	0	1	1	1	0	0	0	0	0	L176.2p	14	17	14	31	24	11	13	13	12	12	14	18	15	17	23	11
V137	USA190BOIDV0	Native Basques	1	1	0	0	0	1	1	1	1	0	0	0	0	S68	15	16	13	29	24	11	13	13	12	13	15	18	14	18	23	13

Attached Table 2. Continuation.

ID	Original code	Population	S116	DF27	L617	L881	DF17	Z195	Z196	L176.2	S68	M167	Z220	Z278	M153	Final Haplogroup	DYS19	DYS389	389I	389II	DYS390	DYS391	DYS392	DYS393	DYS438	DYS439	DYS437	DYS448	DYS456	DYS458	DYS635	GATAH4
V138	0nw9jPa	Native Basques	1	1	0	0	0	1	1	1	0	1	0	0	0	M167	14	16	13	29	24	10	13	13	12	11	15	19	16	17	22	12
V139	ERE016	Native Basques	1	1	0	0	0	1	1	1	0	1	0	0	0	M167	14	16	14	30	24	11	13	13	12	12	15	19	15	17	23	12
V140	KOR014	Native Basques	1	1	0	0	0	1	1	1	0	1	0	0	0	M167	14	16	13	29	24	11	13	13	12	12	15	19	16	16	23	11
V141	KOR025	Native Basques	1	1	0	0	0	1	1	1	0	1	0	0	0	M167	14	19	13	32	24	11	13	13	12	12	15	19	15	16	23	12
V142	KOR035	Native Basques	1	1	0	0	0	1	1	1	0	1	0	0	0	M167	14	16	13	29	24	11	13	13	12	12	15	19	16	16	23	11
V143	KOR041	Native Basques	1	1	0	0	0	1	1	1	0	1	0	0	0	M167	14	16	13	29	24	11	13	13	12	12	15	19	16	16	23	11
V144	NAB023	Native Basques	1	1	0	0	0	1	1	1	0	1	0	0	0	M167	14	16	13	29	24	11	13	13	13	12	15	19	15	19	23	12
V145	0Y3TPa	Native Basques	1	1	0	0	0	1	1	0	0	0	1	1	0	Z278	14	16	13	29	24	11	13	13	12	11	14	18	16	17	24	11
V146	34adjPa	Native Basques	1	1	0	0	0	1	1	0	0	0	1	1	0	Z278	14	16	13	29	24	11	13	13	12	12	14	18	16	17	23	12
V147	3v22sPa	Native Basques	1	1	0	0	0	1	1	0	0	0	1	1	0	Z278	14	17	14	31	24	10	13	13	12	11	14	18	15	17	23	11
V148	5Q235Pa	Native Basques	1	1	0	0	0	1	1	0	0	0	1	0	0	Z220	14	16	14	30	23	11	13	14	12	13	14	18	17	17	23	11
V149	7e560Pa	Native Basques	1	1	0	0	0	1	1	0	0	0	1	1	0	Z278	14	16	14	30	24	11	12	13	12	11	14	18	15	17	23	11
V150	8T10QPa	Native Basques	1	1	0	0	0	1	1	0	0	0	1	1	0	Z278	14	16	14	30	24	11	13	13	12	11	14	18	16	18	23	11
V151	99psYPa	Native Basques	1	1	0	0	0	1	1	0	0	0	1	1	0	Z278	14	16	14	30	24	10	13	13	12	11	14	18	15	17	23	11
V152	9k782Pa	Native Basques	1	1	0	0	0	1	1	0	0	0	1	1	0	Z278	14	16	13	29	24	11	13	13	12	12	14	18	15	17	23	11
V153	ART003	Native Basques	1	1	0	0	0	1	1	0	0	0	1	1	0	Z278	14	16	13	29	23	11	13	14	12	13	14	18	16	18	23	11
V154	ART006	Native Basques	1	1	0	0	0	1	1	0	0	0	1	1	0	Z278	14	17	14	31	24	11	13	13	12	12	14	18	16	17	23	11
V155	ART018	Native Basques	1	1	0	0	0	1	1	0	0	0	1	1	0	Z278	14	16	14	30	24	11	13	13	12	12	14	18	15	17	23	11
V156	ART023	Native Basques	1	1	0	0	0	1	1	0	0	0	1	1	0	Z278	14	16	14	30	24	11	13	13	12	13	14	18	15	16	24	11
V157	ART025	Native Basques	1	1	0	0	0	1	1	0	0	0	1	1	0	Z278	14	18	14	32	24	10	13	13	12	12	15	18	15	18	24	11
V158	B2n7BPa	Native Basques	1	1	0	0	0	1	1	0	0	0	1	1	0	Z278	15	17	15	32	25	10	13	13	12	12	14	18	15	17	23	11
V159	DaMWyPa	Native Basques	1	1	0	0	0	1	1	0	0	0	1	1	0	Z278	14	16	14	30	24	11	13	13	12	12	14	18	15	18	24	11
V160	ERE006	Native Basques	1	1	0	0	0	1	1	0	0	0	1	1	0	Z278	14	16	15	31	24	11	13	13	12	12	14	18	15	17	23	11
V161	ERE008	Native Basques	1	1	0	0	0	1	1	0	0	0	1	1	0	Z278	15	16	14	30	25	11	13	13	12	12	14	18	15	17	23	11
V162	ERE020	Native Basques	1	1	0	0	0	1	1	0	0	0	1	0	0	Z220	14	16	14	30	23	11	13	14	12	13	14	18	16	17	23	11
V163	ERE026	Native Basques	1	1	0	0	0	1	1	0	0	0	1	0	0	Z220	14	16	12	28	23	11	13	14	12	13	14	18	16	17	23	11
V164	G5TcJPa	Native Basques	1	1	0	0	0	1	1	0	0	0	1	0	0	Z220	14	16	13	29	23	10	13	14	12	14	14	18	16	19	23	11
V165	G85e 141	Native Basques	1	1	0	0	0	1	1	0	0	0	1	0	0	Z220	14	16	15	31	23	11	13	14	12	13	14	18	16	17	23	11
V166	G85e 150	Native Basques	1	1	0	0	0	1	1	0	0	0	1	0	0	Z220	14	16	14	30	23	11	13	14	12	13	14	18	16	17	23	11
V167	KOR015	Native Basques	1	1	0	0	0	1	1	0	0	0	1	0	0	Z220	14	16	13	29	23	11	13	14	12	12	14	18	15	17	23	11

Attached Table 2. Continuation.

ID	Original code	Population	S116	DF27	L617	L881	DF17	Z195	Z196	L176.2	S68	M167	Z220	Z278	M153	Final Haplogroup	DYS19	DYS389	389I	389II	DYS390	DYS391	DYS392	DYS393	DYS438	DYS439	DYS437	DYS448	DYS456	DYS458	DYS635	GATAH4
V168	KOR018	Native Basques	1	1	0	0	0	1	1	0	0	0	1	1	0	Z278	14	17	14	31	24	11	13	13	12	12	14	18	16	17	23	11
V169	KOR044	Native Basques	1	1	0	0	0	1	1	0	0	0	1	1	0	Z278	15	17	15	32	25	11	14	13	12	12	14	18	15	17	23	11
V170	KOR049	Native Basques	1	1	0	0	0	1	1	0	0	0	1	1	0	Z278	14	17	14	31	24	11	13	13	12	12	14	18	16	17	23	11
V171	KOR053	Native Basques	1	1	0	0	0	1	1	0	0	0	1	0	0	Z220	14	16	13	29	24	11	13	13	12	12	14	18	16	17	23	11
V172	KOR056	Native Basques	1	1	0	0	0	1	1	0	0	0	1	1	0	Z278	14	16	13	29	24	11	13	14	12	13	14	18	16	17	23	12
V173	n6fioPa	Native Basques	1	1	0	0	0	1	1	0	0	0	1	1	0	Z278	14	17	14	31	23	11	13	13	12	12	14	18	16	17	23	11
V174	NAB017	Native Basques	1	1	0	0	0	1	1	0	0	0	1	0	0	Z220	14	16	13	29	23	11	13	14	12	14	14	18	16	18	23	11
V175	NAB024	Native Basques	1	1	0	0	0	1	1	0	0	0	1	0	0	Z220	14	16	13	29	24	10	13	13	12	13	14	18	16	18	23	11
V176	S1NQDPa	Native Basques	1	1	0	0	0	1	1	0	0	0	1	1	0	Z278	14	16	13	29	25	11	13	14	12	12	14	18	15	17	23	12
V177	USA001RENDV0	Native Basques	1	1	0	0	0	1	1	0	0	0	1	1	0	Z278	14	16	13	29	24	10	13	13	12	12	14	18	15	17	24	12
V178	USA007CHIDV0	Native Basques	1	1	0	0	0	1	1	0	0	0	1	1	0	Z278	14	17	14	31	25	11	13	13	12	12	14	18	15	16	23	11
V179	USA015CHIDV0	Native Basques	1	1	0	0	0	1	1	0	0	0	1	1	0	Z278	14	16	14	30	24	11	13	13	12	11	14	18	15	16	23	11
V180	USA016CHIDV0	Native Basques	1	1	0	0	0	1	1	0	0	0	1	1	0	Z278	14	16	13	29	24	10	13	13	12	12	14	18	15	17	23	12
V181	USA018RENDV0	Native Basques	1	1	0	0	0	1	1	0	0	0	1	1	0	Z278	14	16	14	30	24	11	14	13	12	12	14	18	15	18	23	11
V182	USA021RENDV0	Native Basques	1	1	0	0	0	1	1	0	0	0	1	0	0	Z220	14	17	13	30	24	11	14	13	12	13	14	18	16	16	24	11
V183	USA050CHIDV0	Native Basques	1	1	0	0	0	1	1	0	0	0	1	1	0	Z278	14	16	14	30	24	11	13	13	12	14	14	18	15	17	23	11
V184	USA069BOIDV0	Native Basques	1	1	0	0	0	1	1	0	0	0	1	1	0	Z278	15	16	14	30	24	11	12	13	12	13	14	17	16	17	23	11
V185	USA083BOIDV0	Native Basques	1	1	0	0	0	1	1	0	0	0	1	1	0	Z278	14	17	14	31	24	11	13	12	12	12	14	18	15	16	23	11
V186	USA101BOIDV0	Native Basques	1	1	0	0	0	1	1	0	0	0	1	0	0	Z220	14	16	13	29	23	10	13	14	12	14	14	18	16	18	23	11
V187	USA109BOIDV0	Native Basques	1	1	0	0	0	1	1	0	0	0	1	0	0	Z220	14	16	13	29	23	11	13	14	12	14	14	18	16	18	23	11
V188	USA127BOIDV0	Native Basques	1	1	0	0	0	1	1	0	0	0	1	0	0	Z220	14	16	13	29	23	11	13	14	12	14	14	18	16	18	23	11
V189	USA131BOIDV0	Native Basques	1	1	0	0	0	1	1	0	0	0	1	0	0	Z220	14	16	14	30	24	10	13	13	12	13	14	18	15	16	23	11
V190	USA173BOIDV0	Native Basques	1	1	0	0	0	1	1	0	0	0	1	1	0	Z278	14	16	14	30	24	10	13	13	12	11	14	18	17	18	23	11
V191	USA175BOIDV0	Native Basques	1	1	0	0	0	1	1	0	0	0	1	0	0	Z220	14	17	14	31	24	11	13	13	12	12	14	18	15	17	23	11
V192	6b45yPa	Native Basques	1	1	0	0	0	1	1	0	0	0	1	1	1	M153	14	17	13	30	24	11	13	13	12	12	14	18	15	17	23	12
V193	ART021	Native Basques	1	1	0	0	0	1	1	0	0	0	1	1	1	M153	14	16	13	29	23	11	13	14	12	12	14	18	16	18	23	12
V194	ERE005	Native Basques	1	1	0	0	0	1	1	0	0	0	1	1	1	M153	14	16	14	30	24	11	13	13	12	11	14	18	15	17	23	11
V195	G85e 115	Native Basques	1	1	0	0	0	1	1	0	0	0	1	1	1	M153	14	16	14	30	24	11	13	13	12	11	14	18	15	17	23	11
V196	G85e 175	Native Basques	1	1	0	0	0	1	1	0	0	0	1	1	1	M153	14	16	14	30	24	10	13	13	12	11	14	18	15	17	23	11
V197	G85e 20	Native Basques	1	1	0	0	0	1	1	0	0	0	1	1	1	M153	14	16	13	29	24	11	13	13	12	12	14	18	15	17	24	12

Attached Table 2. Continuation.

ID	Original code	Population	S116	DF27	L617	L881	DF17	Z195	Z196	L176.2	S68	M167	Z220	Z278	M153	Final Haplogroup	DYS19	DYS389	389I	389II	DYS390	DYS391	DYS392	DYS393	DYS438	DYS439	DYS437	DYS448	DYS456	DYS458	DYS635	GATAH4
V198	G85e 42	Native Basques	1	1	0	0	0	1	1	0	0	0	1	1	1	M153	14	16	14	30	24	11	13	13	12	11	14	18	16	17	23	11
V199	KOR002	Native Basques	1	1	0	0	0	1	1	0	0	0	1	1	1	M153	14	16	13	29	24	11	13	13	12	11	14	18	16	17	23	12
V200	NAB020	Native Basques	1	1	0	0	0	1	1	0	0	0	1	1	1	M153	14	16	13	29	24	11	13	13	12	11	14	18	16	17	23	12
V201	USA056BOIDV0	Native Basques	1	1	0	0	0	1	1	0	0	0	1	1	1	M153	14	16	13	29	24	11	13	13	12	11	14	18	16	17	23	12
V202	USA057BOIDV0	Native Basques	1	1	0	0	0	1	1	0	0	0	1	1	1	M153	14	16	13	29	24	10	13	13	12	12	14	18	16	17	23	12
V203	USA059BOIDV0	Native Basques	1	1	0	0	0	1	1	0	0	0	1	1	1	M153	14	16	13	29	24	11	13	14	12	12	14	18	16	17	23	11
V204	USA105BOIDV0	Native Basques	1	1	0	0	0	1	1	0	0	0	1	1	1	M153	14	16	13	29	24	11	13	14	12	12	14	18	16	17	23	12
V205	USA161BOIDV0	Native Basques	1	1	0	0	0	1	1	0	0	0	1	1	1	M153	14	16	13	29	24	11	13	14	12	12	14	18	16	17	23	12
V206	USA167BOIDV0	Native Basques	1	1	0	0	0	1	1	0	0	0	1	1	1	M153	14	16	14	30	24	11	13	13	12	11	14	18	16	17	23	11
V207	6vDj6Pa	Native Basques	1	1	1	0	0	0	0	0	0	0	0	0	0	L617	14	17	13	30	24	10	13	13	12	11	15	19	16	16	23	12
V208	mUz36Pa	Native Basques	1	1	1	0	0	0	0	0	0	0	0	0	0	L617	14	16	13	29	24	11	13	13	12	11	15	19	16	17	23	12
V209	USA005RENDV0	Native Basques	1	1	1	0	0	0	0	0	0	0	0	0	0	L617	14	17	13	30	24	11	13	13	12	11	15	19	16	17	23	12
V210	Z2MvcPa	Native Basques	1	1	1	0	0	0	0	0	0	0	0	0	0	L617	14	16	13	29	23	11	13	13	12	11	15	19	16	17	23	13
	G85e 123	Native Basques	1	1	0	0	0	0	0	0	0	0	0	0	0																	
	G85e 169	Native Basques	1	1	0	0	0	0	0	0	0	0	0	0	0																	
	G85e 81	Native Basques	1	1	0	0	0	0	0	0	0	0	0	0	0																	
	USA039RENDV0	Native Basques	1	1	0	0	0	0	0	0	0	0	0	0	0																	
	KOR022	Native Basques	1	1	0	0	0	1	1	0	0	0	0	0	0																	
	G85e 43	Native Basques	1	1	0	0	0	1	1	0	0	0	1	0	0																	
	KOR003	Native Basques	1	1	0	0	0	1	1	0	0	0	1	1	0																	
	USA085BOIDV0	Native Basques	1	1	0	0	0	1	1	0	0	0	1	0	0																	
V363	ART005	Native Basques	0	0	0	0	0	0	0	0	0	0	0	0	0		15	17	22	11	11	14	10	12	16	21	15	16	14	14	20	12
V364	G85e 47	Native Basques	0	0	0	0	0	0	0	0	0	0	0	0	0		17	15	23	10	11	13	10	12	15	21	14	16	12	12	22	11
V365	G85e 97	Native Basques	0	0	0	0	0	0	0	0	0	0	0	0	0		15	16	25	10	11	12	10	12	14	19	14	16	13	16	22	11
V366	Hhw9CPa	Native Basques	0	0	0	0	0	0	0	0	0	0	0	0	0		15	18	23	10	12	14	10	11	14	20	13	14	14	15	21	10
V367	KOR001	Native Basques	0	0	0	0	0	0	0	0	0	0	0	0	0		14	17	23	10	12	12	10	11	15	20	15	17.2	13	15	20	11
V368	KOR045	Native Basques	0	0	0	0	0	0	0	0	0	0	0	0	0		14	16	22	10	11	13	10	11	16	20	13	15	13	14	21	11
V369	TX 30	Native Basques	0	0	0	0	0	0	0	0	0	0	0	0	0		16	19	25	11	11	13	11	10	14	21	15	15	11	15	23	13
V370	USA008BOIDV0	Native Basques	0	0	0	0	0	0	0	0	0	0	0	0	0		16	16	23	11	12	13	10	11	15	20	17	15	14	15	20	11
V371	USA017CHIDV0	Native Basques	0	0	0	0	0	0	0	0	0	0	0	0	0		16	15	24	10	11	13	10	11	15	21	14	17	12	12	22	12

Attached Table 2. Continuation.

ID	Original code	Population	S116	DF27	L617	L881	DF17	Z195	Z196	L176.2	S68	M167	Z220	Z278	M153	Final Haplogroup	DYS19	DYS389	389I	389II	DYS390	DYS391	DYS392	DYS393	DYS438	DYS439	DYS437	DYS448	DYS456	DYS458	DYS635	GATAH4
V372	USA017RENDV0	Native Basques	0	0	0	0	0	0	0	0	0	0	0	0	0		14	16	22	10	11	13	10	11	16	20	15	17	13	13	21	14
V373	USA035BOIDV0	Native Basques	0	0	0	0	0	0	0	0	0	0	0	0	0		16	19	25	11	11	13	11	10	14	20	15	15	11	15	23	13
V374	USA038CHIDV0	Native Basques	0	0	0	0	0	0	0	0	0	0	0	0	0		13	16	24	9	11	13	10	10	14	20	18	17	13	14	21	12
V375	USA042RENDV0	Native Basques	0	0	0	0	0	0	0	0	0	0	0	0	0		14	17	23	10	12	13	9	11	14	18	16	17	15	15	21	11
V376	USA089BOIDV0	Native Basques	0	0	0	0	0	0	0	0	0	0	0	0	0		13	16	24	9	11	13	10	10	14	20	16	18	13	14	21	12
V377	USA137BOIDV0	Native Basques	0	0	0	0	0	0	0	0	0	0	0	0	0		15	17	23	11	12	15	10	11	14	20	15	15	15	15	21	11
V378	USA176BOIDV0	Native Basques	0	0	0	0	0	0	0	0	0	0	0	0	0		16	16	23	9	11	12	9	12	15	21	16	15	13	16	22	12
V379	USA180BOIDV0	Native Basques	0	0	0	0	0	0	0	0	0	0	0	0	0		15	17	21	10	11	15	10	11	16	22	15	17	13	15	21	11
V380	wAc27Pa	Native Basques	0	0	0	0	0	0	0	0	0	0	0	0	0		17	15	24	9	11	13	10	13	14	21	14	16	12	12	20	11
V381	G85e 230	Native Basques	0	0	0	0	0	0	0	0	0	0	0	0	0		14	16	23	11	13	13	12	11	15	19	16	17	11	14	23	12
V382	TX 502	Native Basques	0	0	0	0	0	0	0	0	0	0	0	0	0		14	16	24	11	13	13	12	12	15	19	15	17	11	13	25	13
V383	USA054BOIDV0	Native Basques	0	0	0	0	0	0	0	0	0	0	0	0	0		14	16	23	13	13	13	12	12	15	19	16	17	11	14	23	13
V384	36uCIPa	Native Basques	1	0	0	0	0	0	0	0	0	0	0	0	0		13	16	24	11	13	13	12	12	15	19	15	15	11	14	23	13
V385	6qY33Pa	Native Basques	1	0	0	0	0	0	0	0	0	0	0	0	0		14	16	24	11	13	13	12	11	15	19	15	18	11	14	23	11
V386	7eo6bPa	Native Basques	1	0	0	0	0	0	0	0	0	0	0	0	0		14	16	24	10	13	13	12	11	15	20	16	17	11	14	24	12
V387	ART011	Native Basques	1	0	0	0	0	0	0	0	0	0	0	0	0		14	16	24	11	13	13	12	11	15	19	15	16	11	15	23	12
V388	ART013	Native Basques	1	0	0	0	0	0	0	0	0	0	0	0	0		14	16	24	10	13	13	12	11	14	19	16	20	11	14	24	12
V389	ART022	Native Basques	1	0	0	0	0	0	0	0	0	0	0	0	0		14	16	24	11	13	13	12	11	15	19	15	17	12	14	23	12
V390	ERE001	Native Basques	1	0	0	0	0	0	0	0	0	0	0	0	0		14	16	24	11	13	13	12	11	15	19	15	17	11	15	23	12
V391	ERE011	Native Basques	1	0	0	0	0	0	0	0	0	0	0	0	0		14	16	24	11	13	13	12	11	15	19	15	17	11	17	23	12
V392	ERE027	Native Basques	1	0	0	0	0	0	0	0	0	0	0	0	0		14	16	24	11	13	13	12	11	15	19	15	17	12	14	23	12
V393	ERE034	Native Basques	1	0	0	0	0	0	0	0	0	0	0	0	0		14	16	24	10	13	13	12	12	15	19	16	17	12	14	23	12
V394	F60C5Pa	Native Basques	1	0	0	0	0	0	0	0	0	0	0	0	0		14	16	24	11	13	13	12	11	15	19	15	18	11	15	24	11
V395	G85e 102	Native Basques	1	0	0	0	0	0	0	0	0	0	0	0	0		14	16	24	11	13	13	12	11	15	19	15	19	12	17	23	12
V396	G85e 113	Native Basques	1	0	0	0	0	0	0	0	0	0	0	0	0		14	16	24	11	13	13	12	11	15	19	15	19	11	14	23	11
V397	G85e 116	Native Basques	1	0	0	0	0	0	0	0	0	0	0	0	0		14	16	24	11	13	13	12	11	15	19	15	18	11	14	23	12
V398	G85e 120	Native Basques	1	0	0	0	0	0	0	0	0	0	0	0	0		14	16	24	11	13	13	12	12	15	19	17	17	11	14	23	12
V399	G85e 154	Native Basques	1	0	0	0	0	0	0	0	0	0	0	0	0		14	16	24	11	13	13	12	11	15	19	15	17	12	14	23	12
V400	G85e 174	Native Basques	1	0	0	0	0	0	0	0	0	0	0	0	0		14	16	24	11	13	13	12	11	15	19	16	15	12	15	23	12
V401	G85e 214	Native Basques	1	0	0	0	0	0	0	0	0	0	0	0	0		14	17	24	11	13	12	12	11	15	19	15	17	11	14	23	12

Attached Table 2. Continuation.

ID	Original code	Population	S116	DF27	L617	L881	DF17	Z195	Z196	L176.2	S68	M167	Z220	Z278	M153	Final Haplogroup	DYS19	DYS389	389I	389II	DYS390	DYS391	DYS392	DYS393	DYS438	DYS439	DYS437	DYS448	DYS456	DYS458	DYS635	GATAH4
V402	J7EG6Pa	Native Basques	1	0	0	0	0	0	0	0	0	0	0	0	0		14	17	24	10	13	13	12	13	15	20	15	16	11	14	23	12
V403	KOR005	Native Basques	1	0	0	0	0	0	0	0	0	0	0	0	0		14	17	24	10	13	13	12	12	14	19	16	17	11	14	23	12
V404	KOR040	Native Basques	1	0	0	0	0	0	0	0	0	0	0	0	0		14	16	24	11	13	13	12	11	15	19	15	18	11	14	23	11
V405	NAB021	Native Basques	1	0	0	0	0	0	0	0	0	0	0	0	0		14	16	24	11	13	13	12	11	15	19	15	18	11	14	23	11
V406	nVB3q	Native Basques	1	0	0	0	0	0	0	0	0	0	0	0	0		14	16	24	11	13	13	12	11	15	19	15	19	11	14	23	11
V407	p5z34Pa	Native Basques	1	0	0	0	0	0	0	0	0	0	0	0	0		14	16	24	11	13	13	12	11	15	19	15	19	11	14	23	11
V408	r7ilrPa	Native Basques	1	0	0	0	0	0	0	0	0	0	0	0	0		14	16	25	11	13	13	12	12	15	19	15	17	11	14	23	11
V409	USA008CHIDV0	Native Basques	1	0	0	0	0	0	0	0	0	0	0	0	0		14	16	24	10	13	13	12	11	15	19	15	18	11	14	23	11
V410	USA008RENDV0	Native Basques	1	0	0	0	0	0	0	0	0	0	0	0	0		14	17	24	11	13	13	12	13	15	19	16	18	14	14	24	13
V411	USA011RENDV0	Native Basques	1	0	0	0	0	0	0	0	0	0	0	0	0		14	16	24	10	13	13	12	12	15	18	16	16	11	12	23	13
V412	USA013RENDV0	Native Basques	1	0	0	0	0	0	0	0	0	0	0	0	0		15	16	24	11	13	13	12	11	15	19	16	16	12	14	23	12
V413	USA014BOIDV0	Native Basques	1	0	0	0	0	0	0	0	0	0	0	0	0		14	16	24	10	13	13	12	14	15	20	16	17	11	14	23	12
V414	USA014CHIDV0	Native Basques	1	0	0	0	0	0	0	0	0	0	0	0	0		14	16	24	11	13	13	12	12	14	19	15	17	11	14	24	12
V415	USA024RENDV0	Native Basques	1	0	0	0	0	0	0	0	0	0	0	0	0		14	17	24	11	13	13	12	11	15	19	15	17	11	13	24	12
V416	USA032CHIDV0	Native Basques	1	0	0	0	0	0	0	0	0	0	0	0	0		14	16	24	11	13	13	12	11	15	19	15	18	11	14	23	12
V417	USA041BOIDV0	Native Basques	1	0	0	0	0	0	0	0	0	0	0	0	0		14	16	24	11	13	13	12	11	15	19	15	18	11	14	23	13
V418	USA046BOIDV0	Native Basques	1	0	0	0	0	0	0	0	0	0	0	0	0		14	16	24	11	13	13	12	11	15	19	15	19	11	14	23	11
V419	USA050RENDV0	Native Basques	1	0	0	0	0	0	0	0	0	0	0	0	0		16	17	23	10	12	15	10	10	14	20	14	16	15	17	18.3	14
V420	USA056CHIDV0	Native Basques	1	0	0	0	0	0	0	0	0	0	0	0	0		14	16	24	11	13	13	12	12	15	18	16	17	11	13	23	12
V421	USA066BOIDV0	Native Basques	1	0	0	0	0	0	0	0	0	0	0	0	0		14	16	24	10	13	13	12	12	15	19	15	17	11	14	24	12
V422	USA068BOIDVO	Native Basques	1	0	0	0	0	0	0	0	0	0	0	0	0		15	16	24	11	13	13	12	11	15	19	16	17	12	14	23	12
V423	USA086BOIDV0	Native Basques	1	0	0	0	0	0	0	0	0	0	0	0	0		14	16	24	11	13	13	12	11	15	19	15	18	11	14	23	12
V424	USA093BOIDV0	Native Basques	1	0	0	0	0	0	0	0	0	0	0	0	0		14	16	25	11	14	13	12	13	15	18	17	17	11	13	23	12
V425	USA099BOIDV0	Native Basques	1	0	0	0	0	0	0	0	0	0	0	0	0		14	16	24	11	13	13	12	12	15	19	15	18	12	14	24	11
V426	USA106BOIDV0	Native Basques	1	0	0	0	0	0	0	0	0	0	0	0	0		14	16	24	10	13	13	12	12	15	19	16	17	11	14	23	12
V427	yaPSYPa	Native Basques	1	0	0	0	0	0	0	0	0	0	0	0	0		14	16	24	11	13	13	12	11	15	19	15	17	12	14	23	12
V428	yKtvgPa	Native Basques	1	0	0	0	0	0	0	0	0	0	0	0	0		14	16	24	10	13	13	12	11	15	20	16	17	11	14	24	12
V429	yOpszPa	Native Basques	1	0	0	0	0	0	0	0	0	0	0	0	0		14	16	24	11	13	13	12	11	15	19	15	19	11	14	23	11
F211	A 3752	Resident Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	13	29	24	11	13	14	12	12	15	19	16	16	23	12
F212	A 3762	Resident Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	14	30	24	10	13	13	12	11	15	19	15	17	23	11

Attached Table 2. Continuation.

ID	Original code	Population	S116	DF27	L617	L881	DF17	Z195	Z196	L176.2	S68	M167	Z220	Z278	M153	Final Haplogroup	DYS19	DYS389	389I	389II	DYS390	DYS391	DYS392	DYS393	DYS438	DYS439	DYS437	DYS448	DYS456	DYS458	DYS635	GATAH4
F213	A 3765	Resident Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	13	29	24	11	13	13	12	12	15	18	15	17	23	12
F214	A 3803	Resident Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	13	29	24	10	13	13	10	11	14	19	16	17	23	12
F215	A 3827	Resident Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	13	29	23	11	13	13	12	12	15	19	16	16	23	12
F216	A 3871	Resident Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	14	30	23	11	13	13	12	12	15	19	15	17	23	12
F217	A 3927	Resident Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	13	29	24	10	13	13	12	12	15	19	15	15	23	12
F218	G85e 109	Resident Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	12	28	24	11	13	13	12	12	15	19	16	16	23	13
F219	G85e 142	Resident Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	12	28	24	11	13	13	12	12	15	19	16	16	23	12
F220	G85e 143	Resident Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	13	29	24	11	13	12	12	12	15	19	15	18	23	12
F221	G85e 177	Resident Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	15	17	14	31	24	11	13	14	12	11	15	20	15	17	23	12
F222	G85e 235	Resident Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	13	29	24	10	13	13	12	12	15	19	15	17	23	12
F223	G85e 52	Resident Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	17	13	30	24	11	13	13	12	11	14	18	15	17	23	11
F224	G85e 60	Resident Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	14	30	23	10	13	13	12	13	15	19	17	17	23	12
F225	G85e 79	Resident Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	14	30	24	11	13	13	12	13	15	18	16	19	23	12
F226	G85e 8	Resident Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	12	28	24	11	13	13	12	13	15	19	17	16	23	12
F227	G85e 92	Resident Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	14	30	24	10	13	13	12	11	15	19	16	17	23	11
F228	G85e 94	Resident Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	17	13	30	23	11	13	13	12	11	15	19	15	17	24	12
F229	TX 109	Resident Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	13	29	23	11	13	14	12	12	15	19	15	16	23	12
F230	TX 2425	Resident Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	13	29	25	11	13	13	12	12	15	19	15	17	23	11
F231	TX 529	Resident Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	14	30	23	10	13	13	12	13	15	19	18	17	23	12
F232	TX 62	Resident Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	13	29	24	10	13	13	12	12	15	19	15	17	23	13
F233	TX 67	Resident Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	15	12	27	24	10	13	13	12	12	15	19	17	17	23	11
F234	TX 87	Resident Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	13	29	24	11	13	13	12	12	15	19	16	17	23	13
F235	TX 96	Resident Basques	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	14	30	24	13	13	13	12	13	15	19	15	18	23	12
F236	A 3736	Resident Basques	1	1	0	0	0	1	1	0	0	0	0	0	0	Z196p	15	17	14	31	24	11	13	13	12	12	15	19	16	18	23	11
F237	A 3790	Resident Basques	1	1	0	0	0	1	1	0	0	0	0	0	0	Z196p	14	16	13	29	24	10	13	13	12	11	15	19	16	18	23	12
F238	A 3961	Resident Basques	1	1	0	0	0	1	1	0	0	0	0	0	0	Z196p	14	16	14	30	25	11	13	13	12	12	14	18	15	18	23	11
F239	G85e 178	Resident Basques	1	1	0	0	0	1	1	0	0	0	0	0	0	Z196p	14	17	13	30	23	11	13	13	11	11	15	19	16	18	23	12
F240	G85e 93	Resident Basques	1	1	0	0	0	1	1	0	0	0	0	0	0	Z196p	14	16	13	29	24	11	13	13	12	11	15	19	16	16	23	12
F241	A 3918	Resident Basques	1	1	0	0	0	1	1	1	0	0	0	0	0	L176.2	14	16	14	30	24	10	13	13	12	12	14	18	16	18	24	11
F242	A 3737	Resident Basques	1	1	0	0	0	1	1	1	0	1	0	0	0	M167	14	15	13	28	24	11	13	13	12	13	15	19	16	18	24	13

Attached Table 2. Continuation.

ID	Original code	Population	S116	DF27	L617	L881	DF17	Z195	Z196	L176.2	S68	M167	Z220	Z278	M153	Final Haplogroup	DYS19	DYS389	389I	389II	DYS390	DYS391	DYS392	DYS393	DYS438	DYS439	DYS437	DYS448	DYS456	DYS458	DYS635	GATAH4
F243	A 3815	Resident Basques	1	1	0	0	0	1	1	1	0	1	0	0	0	M167	14	15	13	28	24	10	13	13	12	12	15	19	16	17	23	12
F244	A 3880	Resident Basques	1	1	0	0	0	1	1	1	0	1	0	0	0	M167	14	16	13	29	24	10	13	15	12	12	15	19	16	18	23	12
F245	G85e 15	Resident Basques	1	1	0	0	0	1	1	1	0	1	0	0	0	M167	14	16	13	29	24	11	13	13	12	11	15	19	15	18	23	12
F246	TX 309	Resident Basques	1	1	0	0	0	1	1	1	0	1	0	0	0	M167	14	16	13	29	24	10	13	13	12	11	15	19	16	16	24	11
F247	TX 518	Resident Basques	1	1	0	0	0	1	1	1	0	1	0	0	0	M167	14	16	13	29	24	11	13	13	12	12	15	19	16	17	23	12
F248	A 3753	Resident Basques	1	1	0	0	0	1	1	0	0	0	1	1	0	Z278	14	16	14	30	24	10	13	13	12	12	14	18	16	17	24	11
F249	A 3796	Resident Basques	1	1	0	0	0	1	1	0	0	0	1	1	0	Z278	14	17	15	32	24	11	13	13	12	12	14	18	15	17	24	11
F250	A 3824	Resident Basques	1	1	0	0	0	1	1	0	0	0	1	0	0	Z220	14	16	13	29	24	11	13	13	12	13	14	18	16	16	23	11
F251	A 3842	Resident Basques	1	1	0	0	0	1	1	0	0	0	1	1	0	Z278	14	16	14	30	22	11	14	13	12	12	14	18	15	17	23	11
F252	A 3872	Resident Basques	1	1	0	0	0	1	1	0	0	0	1	0	0	Z220	14	16	13	29	24	12	13	13	12	13	15	18	16	18	24	11
F253	A 3890	Resident Basques	1	1	0	0	0	1	1	0	0	0	1	1	0	Z278	14	17	13	30	24	11	13	13	12	13	14	18	15	17	23	11
F254	A 3922	Resident Basques	1	1	0	0	0	1	1	0	0	0	1	1	0	Z278	14	16	13	29	24	10	13	13	12	12	15	20	16	18	24	12
F255	A 3758	Resident Basques	1	1	0	0	0	1	1	0	0	0	1	1	0	Z278	14	17	14	31	24	11	13	13	12	12	14	18	15	17	23	11
F256	TX 2099	Resident Basques	1	1	0	0	0	1	1	0	0	0	1	1	0	Z278	14	16	14	30	24	11	13	13	12	11	14	18	15	17	23	11
F257	TX 2538	Resident Basques	1	1	0	0	0	1	1	0	0	0	1	1	0	Z278	14	16	13	29	24	11	12	13	12	12	14	18	16	16	23	11
F258	TX 59	Resident Basques	1	1	0	0	0	1	1	0	0	0	1	1	0	Z278	14	17	14	31	24	11	13	13	12	12	14	18	16	16	23	11
F259	A 3951	Resident Basques	1	1	0	0	0	1	1	0	0	0	1	1	1	M153	14	16	13	29	24	11	13	13	12	13	15	19	15	15	23	12
	A 3940	Resident Basques	1	1	0	0	0	0	0	0	0	0	0	0	0																	
	G85e 167	Resident Basques	1	1	0	0	0	0	0	0	0	0	0	0	0																	
	G85e 130	Resident Basques	1	1	0	0	0	1	1	1	0	1	0	0	0																	
	A 3912	Resident Basques	1	1	0	0	0	1	1	0	0	0	1	1	0																	
F430	A 3759	Resident Basques	0	0	0	0	0	0	0	0	0	0	0	0	0		16	17	25	11	11	13	11	10	14	20	15	15	11	14	23	12
F431	A 3768	Resident Basques	0	0	0	0	0	0	0	0	0	0	0	0	0		16	18	25	11	11	13	11	10	14	20	15	15	11	15	23	12
F432	A 3771	Resident Basques	0	0	0	0	0	0	0	0	0	0	0	0	0		16	14	25	11	11	13	10	11	15	20	14	16	13	17	23	10
F433	A 3785	Resident Basques	0	0	0	0	0	0	0	0	0	0	0	0	0		15	16	22	10	11	12	10	13	14	18	15	13	12	12	24	11
F434	A 3797	Resident Basques	0	0	0	0	0	0	0	0	0	0	0	0	0		15	17	21	10	11	14	10	11	16	21	15	16	14	15	22	11
F435	A 3896	Resident Basques	0	0	0	0	0	0	0	0	0	0	0	0	0		16	17	25	11	11	13	11	10	14	20	16	15	12	14	23	12
F436	A 3933	Resident Basques	0	0	0	0	0	0	0	0	0	0	0	0	0		14	16	22	10	11	13	10	11	16	20	15	15	12	14	22	11
F437	A 3958	Resident Basques	0	0	0	0	0	0	0	0	0	0	0	0	0		15	17	23	10	13	14	9	11	14	19	15	18	14	16	21	11
F438	G85e 112	Resident Basques	0	0	0	0	0	0	0	0	0	0	0	0	0		13	16	24	9	11	13	9	10	14	20	15	18	13	14	21	11

Attached Table 2. Continuation.

ID	Original code	Population	S116	DF27	L617	L881	DF17	Z195	Z196	L176.2	S68	M167	Z220	Z278	M153	Final Haplogroup	DYS19	DYS389	389I	389II	DYS390	DYS391	DYS392	DYS393	DYS438	DYS439	DYS437	DYS448	DYS456	DYS458	DYS635	GATAH4
F439	G85e 114	Resident Basques	0	0	0	0	0	0	0	0	0	0	0	0	0		15	17	25	10	11	12	9	11	16	19	13	16	13	17	22	11
F440	G85e 119	Resident Basques	0	0	0	0	0	0	0	0	0	0	0	0	0		13	16	24	9	11	13	10	10	14	20	15	18	13	14	21	11
F441	G85e 12	Resident Basques	0	0	0	0	0	0	0	0	0	0	0	0	0		16	16	25	10	11	13	10	12	15	21	14	16	13	16	22	10
F442	G85e 122	Resident Basques	0	0	0	0	0	0	0	0	0	0	0	0	0		16	18	25	11	11	13	10	10	14	20	15	15	11	14	23	12
F443	G85e 139	Resident Basques	0	0	0	0	0	0	0	0	0	0	0	0	0		14	16	23	10	11	13	10	11	16	20	14	16	14	14	22	11
F444	G85e 152	Resident Basques	0	0	0	0	0	0	0	0	0	0	0	0	0		13	16	24	9	11	13	10	10	14	20	15	17	14	14	21	12
F445	G85e 176	Resident Basques	0	0	0	0	0	0	0	0	0	0	0	0	0		15	15	23	10	14	13	9	11	14	19	15	15	14	16	21	10
F446	G85e 33	Resident Basques	0	0	0	0	0	0	0	0	0	0	0	0	0		15	16	23	11	11	12	9	9	16	20	15	16	15	17	23	11
F447	G85e 41	Resident Basques	0	0	0	0	0	0	0	0	0	0	0	0	0		15	16	23	10	11	12	9	11	14	21	17	14	13	16	22	12
F448	G85e 57	Resident Basques	0	0	0	0	0	0	0	0	0	0	0	0	0		15	15	23	11	12	13	10	12	15	21	17	17	12	12	22	13
F449	G85e 58	Resident Basques	0	0	0	0	0	0	0	0	0	0	0	0	0		16	18	23	10	11	12	10	11	14	20	15	17.2	13	18	20	10
F450	G85e 70	Resident Basques	0	0	0	0	0	0	0	0	0	0	0	0	0		13	15	23	10	13	13	9	12	14	19	15	16	13	15	21	11
F451	G85e 74	Resident Basques	0	0	0	0	0	0	0	0	0	0	0	0	0		17	15	24	10	11	12	10	12	15	21	16	17	12	12	25	11
F452	G85e 83	Resident Basques	0	0	0	0	0	0	0	0	0	0	0	0	0		14	16	24	9	11	13	10	10	14	20	17	17.2	13	14	21	12
F453	G85e 89	Resident Basques	0	0	0	0	0	0	0	0	0	0	0	0	0		16	16	23	10	11	13	11	10	14	20	15	17	11	15	25	13
F454	TX 111	Resident Basques	0	0	0	0	0	0	0	0	0	0	0	0	0		15	17	23	10	13	13	9	11	14	19	15	20	14	16	21	11
F455	TX 2077	Resident Basques	0	0	0	0	0	0	0	0	0	0	0	0	0		14	17	23	10	11	12	10	13	14	21	15	17.2	13	15	21	11
F456	TX 2417	Resident Basques	0	0	0	0	0	0	0	0	0	0	0	0	0		14	17	23	10	11	12	10	11	14	19	15	17.2	13	19	20	11
F457	TX 2513	Resident Basques	0	0	0	0	0	0	0	0	0	0	0	0	0		14	17	23	10	11	13	10	13	14	19	17	16	15	15	22	11
F458	TX 306	Resident Basques	0	0	0	0	0	0	0	0	0	0	0	0	0		14	17	23	8	11	12	9	11	14	20	16	15	14	19	21	11
F459	TX 319	Resident Basques	0	0	0	0	0	0	0	0	0	0	0	0	0		13	17	23	10	11	12	10	12	14	19	15	12.2	9	19	22	13
F460	TX 331	Resident Basques	0	0	0	0	0	0	0	0	0	0	0	0	0		15	18	21	11	11	14	11	13	14	21	14	17	15	17	20	11
F461	TX 332	Resident Basques	0	0	0	0	0	0	0	0	0	0	0	0	0		14	16	22	10	11	13	10	12	16	20	14	15	13	14	21	11
F462	TX 37	Resident Basques	0	0	0	0	0	0	0	0	0	0	0	0	0		14	17	23	11	11	12	9	12	15	20	13	15	14	19	21	11
F463	TX 423	Resident Basques	0	0	0	0	0	0	0	0	0	0	0	0	0		13	16	24	9	11	13	10	10	14	20	16	17	13	14	21	12
F464	TX 520	Resident Basques	0	0	0	0	0	0	0	0	0	0	0	0	0		15	16	21	10	11	16	10	12	16	20	15	15	15	16	21	11
F465	TX 536	Resident Basques	0	0	0	0	0	0	0	0	0	0	0	0	0		17	15	23	10	11	13	10	12	15	21	14	17	12	12	22	11
F466	TX 560	Resident Basques	0	0	0	0	0	0	0	0	0	0	0	0	0		15	18	23	10	11	13	9	12	14	20	15	15	13	16	23	11
F467	TX 60	Resident Basques	0	0	0	0	0	0	0	0	0	0	0	0	0		14	17	23	10	11	12	11	11	14	20	15	19.2	13	18	20	11
F468	A 3731	Resident Basques	0	0	0	0	0	0	0	0	0	0	0	0	0		14	16	24	10	13	13	12	12	15	19	16	17	11	14	23	12

Attached Table 2. Continuation.

ID	Original code	Population	S116	DF27	L617	L881	DF17	Z195	Z196	L176.2	S68	M167	Z220	Z278	M153	Final Haplogroup	DYS19	DYS389	389I	389II	DYS390	DYS391	DYS392	DYS393	DYS438	DYS439	DYS437	DYS448	DYS456	DYS458	DYS635	GATAH4
F469	A 3884	Resident Basques	0	0	0	0	0	0	0	0	0	0	0	0	0		14	16	23	11	13	13	12	12	15	19	15	18	11	14	23	11
F470	G85e 121	Resident Basques	0	0	0	0	0	0	0	0	0	0	0	0	0		14	16	23	11	13	13	12	13	16	19	15	18	11	15	25	12
F471	TX 2242	Resident Basques	0	0	0	0	0	0	0	0	0	0	0	0	0		14	15	23	11	13	13	12	12	15	19	15	17	11	15	23	13
F472	TX 535	Resident Basques	0	0	0	0	0	0	0	0	0	0	0	0	0		14	17	25	10	13	13	11	13	15	19	15	18	11	14	23	12
F473	A 3742	Resident Basques	1	0	0	0	0	0	0	0	0	0	0	0	0		14	16	23	11	13	13	12	12	15	19	15	17	11	14	24	12
F474	A 3889	Resident Basques	1	0	0	0	0	0	0	0	0	0	0	0	0		14	16	24	11	13	13	12	12	15	19	15	18	10	14	23	12
F475	G85e 104	Resident Basques	1	0	0	0	0	0	0	0	0	0	0	0	0		13	16	23	10	11	14	10	13	14	21	15	16	17	17	21	11
F476	G85e 140	Resident Basques	1	0	0	0	0	0	0	0	0	0	0	0	0		14	16	25	11	13	13	13	12	15	19	16	18	11	14	23	12
F477	G85e 238	Resident Basques	1	0	0	0	0	0	0	0	0	0	0	0	0		14	16	24	11	13	13	11	12	14	19	16	17	13	14	23	12
F478	G85e 243	Resident Basques	1	0	0	0	0	0	0	0	0	0	0	0	0		14	17	25	11	13	13	12	12	15	19	15	17	11	14	23	12
F479	G85e 53	Resident Basques	1	0	0	0	0	0	0	0	0	0	0	0	0		14	16	24	11	13	13	12	12	15	19	16	17	11	14	23	11
F480	TX 2082	Resident Basques	1	0	0	0	0	0	0	0	0	0	0	0	0		14	16	25	11	13	13	12	12	14	18	17	17	11	14	24	12
F481	TX 417	Resident Basques	1	0	0	0	0	0	0	0	0	0	0	0	0		14	16	24	11	14	13	12	11	15	19	15	17	11	14	23	13
F482	TX 443	Resident Basques	1	0	0	0	0	0	0	0	0	0	0	0	0		16	16	24	11	13	13	12	12	15	18	16	17	10	15	23	12
F483	TX 80	Resident Basques	1	0	0	0	0	0	0	0	0	0	0	0	0		15	16	24	10	13	13	13	13	15	19	14	17	11	16	23	12
	A 3928	Resident Basques	0	0	0	0	0	0	0	0	0	0	0	0	0																	
	G85e 173	Resident Basques	0	0	0	0	0	0	0	0	0	0	0	0	0																	
	TX 2302	Resident Basques	0	0	0	0	0	0	0	0	0	0	0	0	0																	
	TX 58	Resident Basques	0	0	0	0	0	0	0	0	0	0	0	0	0																	
R260	ARA007	Aragon	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	12	28	24	11	13	13	12	11	15	20	15	16	23	12
R261	ARA042	Aragon	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	17	14	31	24	11	13	12	12	12	15	19	15	17	25	12
R262	ARA054	Aragon	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	13	29	24	12	13	13	12	12	15	19	16	17	23	11
R263	ARA072	Aragon	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	13	16	13	29	23	10	13	13	9	12	15	19	17	20	21	11
R264	ARA076	Aragon	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	14	30	25	11	13	13	12	11	14	19	15	20	23	11
R265	ARA089	Aragon	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	13	29	24	11	13	13	12	12	15	19	15	18	23	12
R266	ARA098	Aragon	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	14	30	24	11	13	13	12	11	15	18	16	17	24	12
R267	ARA114	Aragon	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	13	29	24	11	13	13	12	12	15	19	16	19	23	12
R268	ARA120	Aragon	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	14	30	25	11	13	13	12	12	15	18	14	17	23	12
R269	ARA141	Aragon	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	15	16	13	29	23	11	15	13	12	12	15	19	16	18	23	12
R270	ARA146	Aragon	1	1	0	0	0	0	0	0	0	0	0	0	0	DF27p	14	16	14	30	24	11	13	13	12	11	15	19	15	17	24	12

Attached Table 2. Continuation.

ID	Original code	Population	S116	DF27	L617	L881	DF17	Z195	Z196	L176.2	S68	M167	Z220	Z278	M153	Final Haplogroup	DYS19	DYS389	389I	389II	DYS390	DYS391	DYS392	DYS393	DYS438	DYS439	DYS437	DYS448	DYS456	DYS458	DYS635	GATAH4
R271	ARA050	Aragon	1	1	0	0	1	1	1	0	0	0	0	0	0	DF17	15	16	13	29	23	11	14	13	12	11	15	18	15	17	23	12
R272	ARA009	Aragon	1	1	0	0	0	1	1	1	0	0	0	0	0	L176.2	14	16	14	30	24	10	13	13	12	14	15	19	17	17	23	12
R273	ARA013	Aragon	1	1	0	0	0	1	1	1	0	0	0	0	0	L176.2	14	16	13	29	24	11	13	13	12	12	15	19	16	17	23	13
R274	ARA073	Aragon	1	1	0	0	0	1	1	1	0	0	0	0	0	L176.2	14	16	14	30	24	10	13	13	9	12	14	18	15	17	24	11
R275	ARA087	Aragon	1	1	0	0	0	1	1	1	0	0	0	0	0	L176.2	14	16	13	29	24	10	13	13	12	12	15	20	16	17	23	12
R276	ARA160	Aragon	1	1	0	0	0	1	1	1	0	0	0	0	0	L176.2	14	15	13	28	24	11	13	13	12	11	15	19	16	17	23	12
R277	ARA161	Aragon	1	1	0	0	0	1	1	1	1	0	0	0	0	S68	14	15	13	28	25	11	13	12	12	12	15	20	16	17	23	12
R278	ARA036	Aragon	1	1	0	0	0	1	1	1	0	1	0	0	0	M167	14	15	13	28	24	11	13	13	12	12	15	19	16	17	23	12
R279	ARA064	Aragon	1	1	0	0	0	1	1	1	0	1	0	0	0	M167	14	16	13	29	24	11	14	13	12	12	15	19	16	19	22	12
R280	ARA088	Aragon	1	1	0	0	0	1	1	1	0	1	0	0	0	M167	14	17	13	30	24	10	13	13	12	12	15	20	15	17	23	12
R281	ARA139	Aragon	1	1	0	0	0	1	1	1	0	1	0	0	0	M167	14	16	14	30	24	11	12	13	12	12	15	19	15	18	23	12
R282	ARA010	Aragon	1	1	0	0	0	1	1	0	0	0	1	1	0	Z278	14	16	14	30	24	10	13	13	12	12	14	18	16	16	23	11
R283	ARA020	Aragon	1	1	0	0	0	1	1	0	0	0	1	1	0	Z278	14	17	13	30	25	11	13	13	12	12	14	18	15	18	24	11
R284	ARA037	Aragon	1	1	0	0	0	1	1	0	0	0	1	0	0	Z220	15	16	13	29	24	11	13	13	12	13	14	18	16	17	23	11
R285	ARA132	Aragon	1	1	0	0	0	1	1	0	0	0	1	0	0	Z220	14	17	13	30	24	11	13	13	12	12	14	18	16	17	23	11
R286	ARA134	Aragon	1	1	0	0	0	1	1	0	0	0	1	0	0	Z220	14	16	14	30	24	11	13	13	12	12	14	18	17	17	24	11
R287	ARA143	Aragon	1	1	0	0	0	1	1	0	0	0	1	0	0	Z220	14	16	12	28	23	11	13	14	12	13	14	18	16	17	23	11
R288	ARA051	Aragon	1	1	0	0	0	1	1	0	0	0	1	1	1	M153	14	16	13	29	24	10	13	13	12	12	14	18	15	17	23	12
	ARA027	Aragon	1	1	0	0	0	0	0	0	0	0	0	0	0																	
	ARA115	Aragon	1	1	0	0	0	0	0	0	0	0	0	0	0																	
	ARA116	Aragon	1	1	0	0	0	0	0	0	0	0	0	0	0																	
R484	ARA001	Aragon	1	0	0	0	0	0	0	0	0	0	0	0	0		14	16	23	11	13	13	12	12	15	19	16	18	11	15	23	12
R485	ARA005	Aragon	1	0	0	0	0	0	0	0	0	0	0	0	0		15	16	24	10	13	13	12	12	14	19	16	17	11	14	23	14
R486	ARA025	Aragon	1	0	0	0	0	0	0	0	0	0	0	0	0		14	16	24	11	13	13	12	12	15	19	16	16	11	14	23	12
R487	ARA031	Aragon	1	0	0	0	0	0	0	0	0	0	0	0	0		13	16	24	10	13	13	12	12	16	19	15	18	11	11	23	12
R488	ARA035	Aragon	1	0	0	0	0	0	0	0	0	0	0	0	0		14	16	25	11	13	13	12	13	15	19	16	16	11	15	23	12
R489	ARA067	Aragon	1	0	0	0	0	0	0	0	0	0	0	0	0		14	17	23	11	13	13	12	12	15	21	16	16	11	16	23	12
R490	ARA071	Aragon	1	0	0	0	0	0	0	0	0	0	0	0	0		14	16	24	10	13	13	12	12	15	19	14	18	11	14	23	12
R491	ARA090	Aragon	1	0	0	0	0	0	0	0	0	0	0	0	0		14	16	25	11	13	13	12	12	15	19	17	17	11	14	23	13
R492	ARA092	Aragon	1	0	0	0	0	0	0	0	0	0	0	0	0		14	16	24	11	13	13	12	12	15	19	16	16	11	14	23	11

Attached Table 2. Continuation.

ID	Original code	Population	S116	DF27	L617	L881	DF17	Z195	Z196	L176.2	S68	M167	Z220	Z278	M153	Final Haplogroup	DYS19	DYS389	389I	389II	DYS390	DYS391	DYS392	DYS393	DYS438	DYS439	DYS437	DYS448	DYS456	DYS458	DYS635	GATAH4
R493	ARA102	Aragon	1	0	0	0	0	0	0	0	0	0	0	0	0		14	16	24	10	14	13	11	11	16	19	15	19	11	14	23	11
R494	ARA106	Aragon	1	0	0	0	0	0	0	0	0	0	0	0	0		14	16	23	11	13	13	12	12	15	19	15	17	11	14	23	12
R495	ARA119	Aragon	1	0	0	0	0	0	0	0	0	0	0	0	0		14	16	25	11	13	13	12	11	15	19	15	16	11	15	23	12
R496	ARA121	Aragon	1	0	0	0	0	0	0	0	0	0	0	0	0		14	16	24	11	13	13	12	12	15	19	15	15	11	14	23	12
R497	ARA124	Aragon	1	0	0	0	0	0	0	0	0	0	0	0	0		15	17	25	11	13	13	12	14	15	19	16	17	11	12	23	12
R498	ARA125	Aragon	1	0	0	0	0	0	0	0	0	0	0	0	0		14	16	26	11	13	12	12	12	15	19	16	17	11	13	23	12
R499	ARA144	Aragon	1	0	0	0	0	0	0	0	0	0	0	0	0		14	16	24	11	13	13	12	12	15	19	14	16	10	15	23	10
R500	ARA145	Aragon	1	0	0	0	0	0	0	0	0	0	0	0	0		14	16	24	11	13	13	12	12	15	19	16	16	10	14	23	12
R501	ARA147	Aragon	1	0	0	0	0	0	0	0	0	0	0	0	0		14	16	24	11	13	13	12	13	15	19	16	18	11	14	23	12
R502	ARA165	Aragon	1	0	0	0	0	0	0	0	0	0	0	0	0		14	16	24	11	13	13	12	12	15	18	16	18	11	14	23	11
R503	ARA167	Aragon	1	0	0	0	0	0	0	0	0	0	0	0	0		14	17	23	10	13	12	12	12	15	19	15	16	11	13	23	11
R504	ARA179	Aragon	1	0	0	0	0	0	0	0	0	0	0	0	0		14	16	24	11	13	13	12	11	15	19	17	17	11	14	23	12
R505	ARA189	Aragon	1	0	0	0	0	0	0	0	0	0	0	0	0		14	16	24	11	13	14	12	12	15	19	16	17	11	14	23	12
	ARA002	Aragon	0	0	0	0	0	0	0	0	0	0	0	0	0																	
	ARA003	Aragon	0	0	0	0	0	0	0	0	0	0	0	0	0																	
	ARA008	Aragon	0	0	0	0	0	0	0	0	0	0	0	0	0																	
	ARA011	Aragon	0	0	0	0	0	0	0	0	0	0	0	0	0																	
	ARA019	Aragon	0	0	0	0	0	0	0	0	0	0	0	0	0																	
	ARA021	Aragon	0	0	0	0	0	0	0	0	0	0	0	0	0																	
	ARA022	Aragon	0	0	0	0	0	0	0	0	0	0	0	0	0																	
	ARA028	Aragon	0	0	0	0	0	0	0	0	0	0	0	0	0																	
	ARA029	Aragon	0	0	0	0	0	0	0	0	0	0	0	0	0																	
	ARA032	Aragon	0	0	0	0	0	0	0	0	0	0	0	0	0																	
	ARA038	Aragon	0	0	0	0	0	0	0	0	0	0	0	0	0																	
	ARA044	Aragon	0	0	0	0	0	0	0	0	0	0	0	0	0																	
	ARA055	Aragon	0	0	0	0	0	0	0	0	0	0	0	0	0																	
	ARA066	Aragon	0	0	0	0	0	0	0	0	0	0	0	0	0																	
	ARA069	Aragon	0	0	0	0	0	0	0	0	0	0	0	0	0																	
	ARA070	Aragon	0	0	0	0	0	0	0	0	0	0	0	0	0																	
	ARA078	Aragon	0	0	0	0	0	0	0	0	0	0	0	0	0																	

Attached Table 2. Continuation.

ID	Original code	Population	S116	DF27	L617	L881	DF17	Z195	Z196	L176.2	S68	M167	Z220	Z278	M153	Final Haplogroup	DYS19	DYS389	389I	389II	DYS390	DYS391	DYS392	DYS393	DYS438	DYS439	DYS437	DYS448	DYS456	DYS458	DYS635	GATAH4
	ARA079	Aragon	0	0	0	0	0	0	0	0	0	0	0	0	0																	
	ARA080	Aragon	0	0	0	0	0	0	0	0	0	0	0	0	0																	
	ARA085	Aragon	0	0	0	0	0	0	0	0	0	0	0	0	0																	
	ARA096	Aragon	0	0	0	0	0	0	0	0	0	0	0	0	0																	
	ARA097	Aragon	0	0	0	0	0	0	0	0	0	0	0	0	0																	
	ARA100	Aragon	0	0	0	0	0	0	0	0	0	0	0	0	0																	
	ARA101	Aragon	0	0	0	0	0	0	0	0	0	0	0	0	0																	
	ARA103	Aragon	0	0	0	0	0	0	0	0	0	0	0	0	0																	
	ARA107	Aragon	0	0	0	0	0	0	0	0	0	0	0	0	0																	
	ARA117	Aragon	0	0	0	0	0	0	0	0	0	0	0	0	0																	
	ARA122	Aragon	0	0	0	0	0	0	0	0	0	0	0	0	0																	
	ARA127	Aragon	0	0	0	0	0	0	0	0	0	0	0	0	0																	
	ARA133	Aragon	0	0	0	0	0	0	0	0	0	0	0	0	0																	
	ARA135	Aragon	0	0	0	0	0	0	0	0	0	0	0	0	0																	
	ARA136	Aragon	0	0	0	0	0	0	0	0	0	0	0	0	0																	
	ARA142	Aragon	0	0	0	0	0	0	0	0	0	0	0	0	0																	
	ARA164	Aragon	0	0	0	0	0	0	0	0	0	0	0	0	0																	
	ARA166	Aragon	0	0	0	0	0	0	0	0	0	0	0	0	0																	
	ARA177	Aragon	0	0	0	0	0	0	0	0	0	0	0	0	0																	
	ARA178	Aragon	0	0	0	0	0	0	0	0	0	0	0	0	0																	
	ARA188	Aragon	0	0	0	0	0	0	0	0	0	0	0	0	0																	

Attached Table 3. Y-STR haplotypes from the studied population groups. N: number of individuals. [1] Núñez et al. 2012; [2] Valverde et al. 2015; [3] Villaescusa et al. 2017. Corresponds to Supplementary Table S3 from *Study number 5*.

Asians from Thailand (N=102)	DYS388	DYS426	DYS461	DYS485	DYS525	DYS561	Final Y-SNP typed in this study	Final Y-SNP previously typed [1,2,3]	Haplogroup
A1	12	11	10	16	9	10	M175		O
A3	12	11	10	15	9	10	M175		O
A4	16	11	10	15	10	11	P126 (xM258)		J
A5	12	11	10	16	9	10	M175		O
A6	12	11	11	15	10	12	M69		H1
A7	12	11	11	15	9	10	M175		O
A8	10	11	13	15	10	11	M175		O
A9	12	11	12	16	9	11	M175		O
A10	12	11	10	15	9	10	M175		O
A11	12	11	10	15	9	11	M175		O
A12	13	11	12	15	10	9	M175		O
A13	12	11	10	15	8	10	M175		O
A14	10	11	11	15	9	11	M175		O
A15	12	11	11	15	9	10	M175		O
A16	12	11	11	16	9	10	M175		O
A18	12	11	10	15	9	10	M175		O
A19	12	11	11	17	9	10	M175		O
A20	10	11	11	15	9	11	M175		O
A21	12	11	12	16	9	11	M175		O
A22	12	11	10	15	9	11	M175		O
A23	12	11	10	15	9	10	M175		O
A24	12	12	11	15	12	11	M207 (xM269)		R
A25	12	11	10	16	11	11	P170		E
A26	10	11	11	15	9	10	M175		O
A27	12	11	11	16	9	11	M175		O
A28	12	11	10	15	9	10	M175		O
A29	12	11	12	15	10	10	M175		O
A30	12	11	12	16	9	12	M175		O
A31	12	11	10	15	9	10	M175		O
A32	12	11	13	16	9	11	M175		O

Attached Table 3. Continuation.

Asians from Thailand (N=102)	DYS388	DYS426	DYS461	DYS485	DYS525	DYS561	Final Y-SNP typed in this study	Final Y-SNP previously typed [1,2,3]	Haplogroup
A33	12	11	10	15	10	10	M175		O
A34	12	11	11	15	9	10	M175		O
A35	12	11	11	16	10	11	M175		O
A36	12	11	11	15	10	9	M175		O
A37	10	11	11	15	9	10	M175		O
A38	12	11	9	15	9	11	M175		O
A39	12	12	10	15	10	10	M207 (xM269)		R
A40	12	11	10	15	9	10	M175		O
A41	12	11	11	15	9	11	M175		O
A42	12	11	11	15	9	10	M175		O
A43	12	11	10	15	9	10	M175		O
A44	12	11	9	16	9	10	M175		O
A45	12	11	11	16	9	11	M175		O
A46	12	11	11	16	9	11	M175		O
A47	12	11	13	14	9	11	M175		O
A48	12	11	9	15	9	10	M175		O
A49	12	11	10	15	9	10	M175		O
A50	10	11	11	15	9	10	M175		O
A51	12	11	10	16	9	10	M175		O
A52	12	11	11	15	10	11	M69		H1
A53	12	12	10	15	10	10	M207 (xM269)		R
A54	12	11	11	15	10	10	M69		H1
A56	12	11	9	15	9	10	M175		O
A58	12	12	10	16	10	10	M207 (xM269)		R
A59	12	12	10	15	10	11	M207 (xM269)		R
A60	10	11	11	15	10	11	M175		O
A61	12	11	11	16	9	11	M175		O
A62	12	11	10	15	9	10	M175		O
A69	12	11	11	16	9	11	M175		O
A70	12	11	10	15	10	10	M175		O
A71	10	11	11	15	9	10	M175		O

Attached Table 3. Continuation.

Asians from Thailand (N=102)	DYS388	DYS426	DYS461	DYS485	DYS525	DYS561	Final Y-SNP typed in this study	Final Y-SNP previously typed [1,2,3]	Haplogroup
A72	12	11	11	16	9	9	M175		O
A75	13	11	13	15	9	10	M175		O
A77	12	11	10	14	9	10	M175		O
A78	10	11	11	15	9	10	M175		O
A79	14	11	10	15	8	10	M175		O
A80	12	11	10	15	9	10	M175		O
A82	13	11	10	15	9	11	M231		N
A83	12	11	11	15	9	11	M175		O
A84	10	11	11	15	9	10	M175		O
A85	12	11	11	17	9	10	M175		O
A86	13	11	11	14	9	11	M175		O
A89	12	11	10	15	9	10	M175		O
A90	12	11	9	16	9	10	M175		O
A91	12	11	10	15	9	10	M175		O
A92	12	11	11	15	8	10	M175		O
A95	12	11	10	15	9	10	M175		O
A96	14	11	12	13	9	9	M201		G
A97	16	11	11	15	13	10	P126 (xM258)		J
A98	12	11	10	15	9	10	M175		O
A99	10	11	11	15	10	11	M175		O
A100	12	11	10	16	9	10	M175		O
A101	11	12	10	15	10	10	M207 (xM269)		R
A102	10	11	11	15	9	10	M175		O
A103	12	11	10	15	9	10	M175		O
A104	12	11	10	15	9	10	M175		O
A105	12	11	11	15	9	10	M175		O
A106	12	11	9	15	9	11	M175		O
A107	12	11	10	15	9	10	M175		O
A108	12	11	11	15	9	10	M175		O
A109	12	11	10	15	9	10	M175		O
A110	12	11	11	17	9	11	M175		O

Attached Table 3. Continuation.

Asians from Thailand (N=102)	DYS388	DYS426	DYS461	DYS485	DYS525	DYS561	Final Y-SNP typed in this study	Final Y-SNP previously typed [1,2,3]	Haplogroup
A111	12	11	11	16	9	11	M175		O
A112	13	11	10	15	9	10	M175		O
A113	12	11	11	15	9	10	M175		O
A114	12	11	10	16	9	10	M175		O
A115	12	11	12	16	9	10	M175		O
A116	10	11	11	15	9	10	M175		O
A117	12	11	12	15	9	10	M175		O
A119	12	11	11	17	10	11	M231		N
A120	12	11	10	16	9	10	M175		O
A121	12	11	10	16	9	11	M175		O

Native Americans from Guatemala (N=50)	DYS388	DYS426	DYS461	DYS485	DYS525	DYS561	Final Y-SNP typed in this study	Final Y-SNP previously typed [1,2,3]	Haplogroup
GUA047	12	12	11	14	9	10	M3		Q1a2
GUA051	12	12	11	14	9	10	M3		Q1a2
GUA052	12	12	11	12	9	10	M3		Q1a2
GUA053	12	12	11	12	9	10	M3		Q1a2
GUA056	12	12	11	16	10	10	M242 (xM3)		Q
GUA060	12	12	13	15	9	10	M242		Q
GUA069	11	12	12	14	11	10	M3		Q1a2
GUA071	12	12	11	14	9	10	M3		Q1a2
GUA076	12	12	11	16	10	10	M242 (xM3)		Q
GUA077	12	12	11	14	9	10	M3		Q1a2
GUA080	12	12	12	14	9	10	M3		Q1a2
GUA082	13	12	11	15	10	10	M3		Q1a2
GUA083	11	12	11	13	11	10	M3		Q1a2
GUA085	13	12	12	14	9	11	M242 (xM3)		Q
GUA094	12	12	11	14	9	10	M3		Q1a2
GUA102	12	12	11	13	9	11	M3		Q1a2

Attached Table 3. Continuation.

Native Americans from Guatemala (N=50)	DYS388	DYS426	DYS461	DYS485	DYS525	DYS561	Final Y-SNP typed in this study	Final Y-SNP previously typed [1,2,3]	Haplogroup
GUA103	12	12	11	14	9	11	M242 (xM3)		Q
GUA104	12	12	11	14	9	10	M3		Q1a2
GUA106	12	12	11	14	9	10	M3		Q1a2
GUA107	13	12	12	14	9	10	M242 (xM3)		Q
GUA108	12	12	11	14	9	10	M3		Q1a2
GUA118	12	12	11	14	9	10	M242 (xM3)		Q
GUA122	12	12	12	14	9	10	M3		Q1a2
GUA124	12	12	11	14	9	10	M3		Q1a2
GUA131	12	12	11	14	9	10	M3		Q1a2
GUA138	12	12	11	14	9	10	M3		Q1a2
GUA147	12	12	11	15	10	10	M269		R1b
GUA162	12	12	11	14	9	10	M3		Q1a2
GUA164	12	12	10	12	9	11	M242 (xM3)		Q
GUA165	12	12	11	14	9	11	M3		Q1a2
GUA168	12	12	11	14	9	11	M3		Q1a2
GUA169	16	11	9	15	10	10	P126 (xM258)		J
GUA170	13	12	11	15	10	10	M3		Q1a2
GUA179	12	12	11	14	9	9	M3		Q1a2
GUA199	12	12	11	15	9	10	M242 (xM3)		Q
GUA200	12	12	11	16	10	10	M3		Q1a2
GUA203	12	12	11	14	9	10	M242 (xM3)		Q
GUA204	12	12	11	14	9	10	M242 (xM3)		Q
GUA206	12	12	11	16	10	10	M242 (xM3)		Q
GUA207	13	12	11	15	10	10	M3		Q1a2
GUA208	12	11	11	15	10	10	P170		E
GUA209	12	12	11	14	9	10	M3		Q1a2
GUA210	12	12	11	14	9	10	M3		Q1a2
GUA211	12	12	12	15	9	10	M242 (xM3)		Q
GUA212	12	12	11	14	9	11	M242 (xM3)		Q
GUA213	12	12	11	14	9	11	M3		Q1a2
GUA214	12	12	11	14	9	10	M3		Q1a2

Attached Table 3. Continuation.

Native Americans from Guatemala (N=50)	DYS388	DYS426	DYS461	DYS485	DYS525	DYS561	Final Y-SNP typed in this study	Final Y-SNP previously typed [1,2,3]	Haplogroup
GUA215	12	12	11	14	9	10	M3		Q1a2
GUA216	12	12	11	16	10	10	M242 (xM3)		Q
GUA217	12	12	12	15	9	10	M242 (xM3)		Q

Hispanics from Colombia (N=60)	DYS388	DYS426	DYS461	DYS485	DYS525	DYS561	Final Y-SNP typed in this study	Final Y-SNP previously typed [1,2,3]	Haplogroup
1709P	12	12	11	16	10	10	M242		Q
1713P	13	11	10	15	10	10	M201		G
1718P	15	11	12	16	11	9	P126 (xM258)		J
1726P	12	11	12	17	10	10	P170		E
1743P	12	12	11	15	10	10	M269		R1b
1747P	13	11	11	14	10	9	M258		I
1756P	12	11	12	14	10	11	P170		E
1764P	13	11	11	16	10	11	P170		E
1787P	12	12	11	15	10	10	M269		R1b
1796P	12	11	11	15	11	11	P170		E
1805P	15	11	10	16	10	10	P126 (xM258)		J
1810P	12	12	11	15	10	10	M269		R1b
1817P	12	11	12	14	10	10	P170		E
1822P	18	11	11	15	11	10	P126 (xM258)		J
1829P	12	12	11	15	11	10	M269		R1b
1835P	12	12	11	15	10	10	M269		R1b
1839P	12	12	11	15	10	10	M269		R1b
1841P	12	12	11	15	10	10	M269		R1b
1843P	12	11	12	17	10	10	P170		E
1853P	16	11	10	15	10	10	P126 (xM258)		J
1861P	12	12	11	14	10	10	M242		Q
1865P	12	12	11	15	10	10	M269		R1b
1866P	17	11	10	15	10	10	P126 (xM258)		J

Attached Table 3. Continuation.

Hispanics from Colombia (N=60)	DYS388	DYS426	DYS461	DYS485	DYS525	DYS561	Final Y-SNP typed in this study	Final Y-SNP previously typed [1,2,3]	Haplogroup
1871P	12	12	11	14	10	10	P126 (xM258)		J
1879P	12	11	10	15	11	10	M272		T
1885P	12	11	12	14	10	10	P170		E
1889P	12	12	11	16	10	10	M269		R1b
1893P	13	11	10	14	12	9	M258		I
1906P	12	12	11	15	10	10	M269		R1b
1910P	10	12	10	15	11	12	M269		R1b
1915P	12	12	11	14	10	10	P170		E
1923P	12	12	11	15	11	9	M269		R1b
1976P	12	12	11	15	10	10	M269		R1b
1984P	12	12	11	15	10	10	M269		R1b
1995P	12	12	11	14	10	12	M242		Q
2701P	12	12	11	15	10	10	M269		R1b
2906P	12	12	12	15	10	10	M269		R1b
2922P	12	12	11	15	10	10	M269		R1b
2926P	12	12	11	14	12	10	M269		R1b
2952P	12	12	11	15	11	9	M269		R1b
2954P	12	12	11	15	10	10	M269		R1b
2994P	12	12	12	15	10	10	M269		R1b
3016P	12	11	11	17	13	9	M272		T
3024P	12	11	10	15	11	10	M272		T
3067P	12	12	11	15	10	10	M269		R1b
3085P	12	12	11	15	10	11	M269		R1b
3091P	12	12	11	15	10	10	M269		R1b
3099P	12	12	11	15	10	10	M269		R1b
3123P	12	12	11	15	10	10	M269		R1b
3137P	12	12	11	15	11	10	M269		R1b
3156P	12	12	11	14	10	10	M242 (xM3)		Q
3193P2	12	12	10	14	10	10	M269		R1b
3239P	12	12	11	15	10	9	M207 (xM269)		R
3284P	12	12	11	15	10	10	M269		R1b

Attached Table 3. Continuation.

Hispanics from Colombia (N=60)	DYS388	DYS426	DYS461	DYS485	DYS525	DYS561	Final Y-SNP typed in this study	Final Y-SNP previously typed [1,2,3]	Haplogroup
3432P	12	12	11	15	10	10	M269		R1b
3447P	12	12	10	15	10	11	M269		R1b
3483P	17	11	10	15	12	10	M258		I
3496P	12	11	10	15	11	10	M272		T
3500P	12	12	11	15	12	10	M269		R1b
3501P	12	11	11	15	10	11	M201		G

Hispanics from Nicaragua (N=66)	DYS388	DYS426	DYS461	DYS485	DYS525	DYS561	Final Y-SNP typed in this study	Final Y-SNP previously typed [1,2,3]	Haplogroup
NIC001	12	11	10	15	11	10		M9	KLT
NIC002	12	12	11	15	10	10		M343	R
NIC006	15	11	9	15	10	9		M102	J
NIC008	12	12	11	14	10	10		M343	R
NIC009	12	12	11	15	10	10		M343	R
NIC010	15	11	11	17	10	9		M410	J
NIC011	13	11	10	14	12	9		P37.2	I
NIC014	12	11	10	15	10	10		M201	G
NIC016	12	12	11	15	10	10		M343	R
NIC017	12	12	11	16	10	11		M343	R
NIC019	12	12	11	16	10	10		M343	R
NIC020	12	12	11	15	10	10		M343	R
NIC021	12	11	12	15	11	11		M35	E
NIC022	12	12	11	15	10	10		M343	R
NIC023	12	12	10	15	9	10		M3	Q1a2
NIC024	12	12	11	17	10	11		M343	R
NIC026	12	12	11	15	10	10		M343	R
NIC027	12	11	11	16	11	10		M201	G
NIC028	12	12	11	15	10	11		M343	R
NIC029	14	11	11	12	10	11		M253	I

Attached Table 3. Continuation.

Hispanics from Nicaragua (N=66)	DYS388	DYS426	DYS461	DYS485	DYS525	DYS561	Final Y-SNP typed in this study	Final Y-SNP previously typed [1,2,3]	Haplogroup
NIC030	12	11	12	16	11	11		M35	E
NIC031	12	11	13	15	11	10		M35	E
NIC033	12	12	11	14	9	10		M346	Q
NIC034	12	12	11	15	10	10		M343	R
NIC035	12	12	11	15	10	10		M343	R
NIC036	12	11	11	16	10	10		M35	E
NIC037	13	12	10	15	10	10		M343	R
NIC038	12	12	11	15	12	10		M343	R
NIC039	12	11	12	15	11	10		M35	E
NIC040	14	11	12	16	11	10		M410	J
NIC042	12	11	12	15	10	10		M35	E
NIC043	16	11	11	17	11	10		M410	J
NIC044	12	12	11	13	10	10		M343	R
NIC045	12	12	12	15	10	10		M343	R
NIC048	12	12	11	14	10	10		M343	R
NIC051	15	11	11	17	10	11		M410	J
NIC054	12	12	11	15	10	10		M343	R
NIC055	12	12	11	13	10	10		M343	R
NIC056	12	12	11	15	10	10		M343	R
NIC057	12	12	11	15	10	10		M343	R
NIC058	13	11	10	14	11	9		P37.2	I
NIC059	16	11	11	17	11	10		M410	J
NIC060	12	12	11	15	10	10		M343	R
NIC062	12	12	11	14	10	9		M343	R
NIC063	12	11	11	17	12	9		M9	KLT
NIC066	12	11	12	14	10	10		M2	E
NIC067	12	11	12	16	11	11		M35	E
NIC068	12	12	11	15	10	10		M343	R
NIC069	12	12	12	15	10	10		M343	R
NIC070	12	12	11	15	10	10		M343	R
NIC071	12	12	11	15	10	10		M343	R

Attached Table 3. Continuation.

Hispanics from Nicaragua (N=66)	DYS388	DYS426	DYS461	DYS485	DYS525	DYS561	Final Y-SNP typed in this study	Final Y-SNP previously typed [1,2,3]	Haplogroup
NIC073	12	12	11	14	9	11		M346	Q
NIC074	13	11	11	16	10	10		M2	E
NIC075	13	12	12	15	10	10		M343	R
NIC076	12	12	9	14	9	11		M3	Q1a2
NIC077	12	12	11	14	10	10		M343	R
NIC078	17	11	10	15	10	10		M267	J
NIC081	12	11	11	17	12	9		M9	KLT
NIC082	12	12	11	15	10	10		M343	R
NIC083	16	11	10	15	10	10		M267	J
NIC084	12	12	11	15	10	10		M343	R
NIC085	12	11	11	15	11	10		M35	E
NIC086	12	12	11	14	10	10		M343	R
NIC089	16	11	12	17	13	9		M410	J
NIC090	12	12	11	15	10	10		M343	R
NIC091	14	11	10	12	10	10		M253	I

Africans from Malawi (N=31)	DYS388	DYS426	DYS461	DYS485	DYS525	DYS561	Final Y-SNP typed in this study	Final Y-SNP previously typed [1,2,3]	Haplogroup
MW002	12	11	11	14	10	9	P170		E
MW004	12	11	12	14	10	11	P170		E
MW007	12	11	12	14	10	10	P170		E
MW009	12	11	11	14	10	10	P170		E
MW011	12	11	11	14	10	11	P170		E
MW015	10	11	10	16	11	10	xM168		AB(xCDEF)
MW019	12	11	13	15	10	10	P170		E
MW023	12	11	11	14	10	10	P170		E
MW024	12	11	11	14	10	10	P170		E
MW025	12	11	11	14	10	10	P170		E
MW026	12	11	11	14	10	11	P170		E

Attached Table 3. Continuation.

Africans from Malawi (N=31)	DYS388	DYS426	DYS461	DYS485	DYS525	DYS561	Final Y-SNP typed in this study	Final Y-SNP previously typed [1,2,3]	Haplogroup
MW028	12	11	12	14	10	9	P170		E
MW029	12	11	12	14	10	11	P170		E
MW034	12	11	12	14	10	11	P170		E
MW037	12	11	13	14	11	10	P170		E
MW040	12	11	12	16	11	10	P170		E
MW042	12	11	12	14	11	11	P170		E
MW047	12	11	12	14	10	9	P170		E
MW048	12	11	13	14	9	10	P170		E
MW049	13	11	12	14	10	10	P170		E
MW050	12	11	11	14	10	11	P170		E
MW053	12	11	12	15	10	11	P170		E
MW057	12	11	12	14	10	10	P170		E
MW061	12	11	10	14	10	10	P170		E
MW062	12	11	12	14	9	10	P170		E
MW063	12	11	12	18	10	9	P170		E
MW064	12	11	12	14	10	11	P170		E
MW067	12	11	12	14	11	12	P170		E
MW068	12	11	12	16	11	11	P170		E
MW069	12	11	13	14	9	10	P170		E
MW071	12	11	12	14	10	11	P170		E

European Caucasians from Spain (N=219)	DYS388	DYS426	DYS461	DYS485	DYS525	DYS561	Final Y-SNP typed in this study	Final Y-SNP previously typed [1,2,3]	Haplogroup
ALI007	15	11	9	15	10	9	P126 (xM258)		J
ALI013	12	12	11	15	10	11		M167	R1b
ALI021	12	12	12	15	10	10		M167	R1b
ALI034	12	11	11	15	10	10	P170		E
ALI036	12	12	11	15	10	10		Z220	R1b
ALI038	12	11	11	15	11	10	M201		G

Attached Table 3. Continuation.

European Caucasians from Spain (N=219)	DYS388	DYS426	DYS461	DYS485	DYS525	DYS561	Final Y-SNP typed in this study	Final Y-SNP previously typed [1,2,3]	Haplogroup
ALI050	12	12	11	15	10	10		Z220	R1b
ALI051	15	11	12	17	10	9	P126 (xM258)		J
ALI063	12	12	11	15	10	11		P312	R1b
ALI070	12	12	11	15	10	9		DF27	R1b
ALI072	12	12	11	16	9	10		P312	R1b
ALI073	12	11	11	14	11	9	M258		I
ALI074	12	12	11	15	12	10		Z220	R1b
ALI080	12	11	12	15	11	11	P170		E
ALI085	12	12	11	15	10	10		DF27	R1b
ALI109	12	12	11	14	10	10		DF27	R1b
ALI130	12	12	11	17	9	10		DF27	R1b
ALI131	12	12	11	15	10	10		DF27	R1b
ALI132	12	12	11	15	10	10		DF27	R1b
ALI169	12	12	11	15	10	10		DF27	R1b
ALI181	12	12	11	15	10	10		DF27	R1b
ALI202	12	12	11	15	10	10		DF27	R1b
ALI204	12	12	11	15	10	10		DF27	R1b
ALI205	12	12	11	15	10	10		DF27	R1b
ALI242	12	12	11	13	10	10		DF27	R1b
ALI252	12	12	11	15	10	10		DF27	R1b
ALI259	12	12	11	15	10	11		DF27	R1b
ALI266	12	12	11	15	10	11		DF27	R1b
ALI272	12	12	11	15	10	10		DF27	R1b
ALI301	12	11	11	15	10	10	M269		R1b
ALI310	14	11	12	17	11	10	P126 (xM258)		J
ALI317	15	11	13	16	10	9	P126 (xM258)		J
ALI341	12	12	10	15	11	10	M207 (xM269)		R
ALI345	12	12	11	15	10	10		DF27	R1b
ALI354	12	12	11	15	10	10		DF27	R1b
ALI355	12	12	11	15	10	10		P312	R1b
ALI357	12	12	11	15	10	10		U106	R1b

Attached Table 3. Continuation.

European Caucasians from Spain (N=219)	DYS388	DYS426	DYS461	DYS485	DYS525	DYS561	Final Y-SNP typed in this study	Final Y-SNP previously typed [1,2,3]	Haplogroup
ALI358	13	12	9	15	10	10		DF27	R1b
ALI362	12	12	11	15	10	10	M207 (xM269)		R
ALI363	12	14	11	15	10	10		U152	R1b
ALI364	15	11	13	17	11	9	P126 (xM258)		J
ALI366	12	12	10	15	10	10		M269	R1b
ALI368	12	12	11	15	11	10		M153	R1b
ALI380	12	12	11	15	10	10		Z278	R1b
ALI381	12	12	11	15	9	10		M153	R1b
ALI383	12	12	11	15	11	10		Z278	R1b
ALI384	12	12	11	15	10	11		P312	R1b
ALI386	12	11	11	15	11	11	P170		E
ALI393	12	12	11	15	10	10		DF27	R1b
ALI394	12	12	12	15	10	10		P312	R1b
BCN009	16	11	9	15	10	9	P126 (xM258)		J
BCN011	12	12	11	15	10	10		U152	R1b
BCN012	12	12	11	15	11	11		U152	R1b
BCN013	12	12	11	15	10	10		DF27	R1b
BCN014	12	12	11	15	9	10		DF27	R1b
BCN015	15	11	9	15	10	9	P126 (xM258)		J
BCN016	12	12	11	15	10	10		DF27	R1b
BCN017	12	12	11	15	10	10		DF27	R1b
BCN018	12	12	11	15	10	11		DF27	R1b
BCN019	12	12	11	15	10	9		M529	R1b
BCN020	12	11	10	15	10	10	M201		G
BCN021	12	12	11	15	10	10		DF27	R1b
BCN022	12	12	11	15	10	10		DF27	R1b
BCN023	12	12	11	15	10	10		P312	R1b
BCN024	12	12	11	15	10	11		DF27	R1b
BCN025	12	13	10	15	10	10	M269		R1b
BCN026	12	12	11	15	10	10		DF27	R1b
BCN027	12	12	11	15	11	10		DF27	R1b

Attached Table 3. Continuation.

European Caucasians from Spain (N=219)	DYS388	DYS426	DYS461	DYS485	DYS525	DYS561	Final Y-SNP typed in this study	Final Y-SNP previously typed [1,2,3]	Haplogroup
BCN029	12	12	11	15	10	10		DF27	R1b
BCN030	15	11	9	15	10	9	P126 (xM258)		J
BCN031	13	11	10	16	9	10	M201		G
BCN032	12	12	11	15	10	10		DF27	R1b
BCN033	13	11	11	14	11	9	M258		I
BCN034	12	12	9	15	10	10		DF27	R1b
BCN035	14	11	11	12	10	10	M258		I
BCN036	12	11	10	16	10	11	M201		G
BCN037	12	11	11	17	10	10	P170		E
BCN038	12	12	11	15	10	10		P312	R1b
BCN039	12	11	12	15	11	11	P170		E
BCN040	12	12	11	15	10	10		DF27	R1b
BCN052	12	11	11	15	10	10	P170		E
BCN059	12	12	11	15	10	10		DF27	R1b
BCN060	12	12	11	15	10	10		DF27	R1b
BCN061	12	12	11	15	10	10		P312	R1b
BCN062	12	12	13	14	10	10		DF27	R1b
BCN063	16	11	11	18	11	9	P126 (xM258)		J
BCN067	12	12	11	15	10	10		P312	R1b
BCN068	12	12	9	15	9	11		M269	R1b
BCN069	12	12	11	15	10	10		P312	R1b
BCN071	12	12	11	15	10	10		DF27	R1b
BCN072	12	12	11	15	11	10		DF27	R1b
BCN073	12	11	10	15	10	10	M201		G
BCN074	13	11	10	14	12	9	M258		I
BCN075	12	12	11	14	10	10		DF27	R1b
BCN076	12	12	11	14	10	10		DF27	R1b
BCN077	12	12	11	15	10	10		DF27	R1b
BCN078	14	11	11	12	10	8	M258		I
BCN080	12	11	11	15	10	11	P170		E
BCN082	12	12	11	15	10	10		DF27	R1b

Attached Table 3. Continuation.

European Caucasians from Spain (N=219)	DYS388	DYS426	DYS461	DYS485	DYS525	DYS561	Final Y-SNP typed in this study	Final Y-SNP previously typed [1,2,3]	Haplogroup
BCN083	12	11	12	17	10	10	P170		E
BCN084	12	12	11	15	10	10		DF27	R1b
BCN087	13	11	11	13	9	9	M258		I
BCN094	13	12	10	15	10	10		DF27	R1b
BCN095	12	12	11	15	10	10		DF27	R1b
MAD010	12	11	11	16	10	11	P170		E
MAD011	14	11	11	12	10	10	M258		I
MAD012	12	12	11	15	10	10		M153	R1b
MAD013	12	12	11	15	10	10		Z196	R1b
MAD014	12	12	11	15	10	10		L176.2	R1b
MAD015	12	12	11	15	10	10		DF27	R1b
MAD016	13	12	12	15	10	11		U152	R1b
MAD017	12	12	11	15	10	10		L176.2	R1b
MAD018	12	12	11	15	10	9		P312	R1b
MAD019	14	11	11	12	11	11	M258		I
MAD020	12	12	11	14	11	10		L11	R1b
MAD021	12	12	11	15	10	10		P312	R1b
MAD022	12	12	11	15	10	10		S356	R1b
MAD023	12	12	12	16	10	10		S356	R1b
MAD024	16	11	11	15	10	10	P126 (xM258)		J
MAD025	12	12	11	15	11	10		Z196	R1b
MAD026	12	12	12	15	10	10		S356	R1b
MAD028	12	11	10	15	10	11	P170		E
MAD029	12	12	10	15	10	10		M269	R1b
MAD030	12	12	11	15	10	11		DF27	R1b
MAD031	12	13	10	15	10	10		M269	R1b
MAD032	12	12	11	15	10	10		DF27	R1b
MAD033	12	12	11	15	10	10		S356	R1b
MAD034	12	12	11	14	10	10		L176.2	R1b
MAD035	15	11	11	16	10	9	M258		I
MAD036	12	12	11	15	10	10		DF27	R1b

Attached Table 3. Continuation.

European Caucasians from Spain (N=219)	DYS388	DYS426	DYS461	DYS485	DYS525	DYS561	Final Y-SNP typed in this study	Final Y-SNP previously typed [1,2,3]	Haplogroup
MAD037	12	12	11	15	10	10		S68	R1b
MAD038	12	12	11	15	10	10		U152	R1b
MAD039	12	12	11	15	10	10		Z196	R1b
MAD040	12	11	13	15	11	11	P170		E
MAD041	15	11	13	15	12	9	P126 (xM258)		J
MAD062	13	11	11	13	10	9	M258		I
MAD063	12	12	11	16	10	10		DF27	R1b
MAD064	12	12	11	15	10	10		P312	R1b
MAD065	12	11	11	15	8	10	M201		G
MAD066	12	11	10	14	10	10	M201		G
MAD067	16	11	10	15	10	10	P126 (xM258)		J
MAD068	14	11	11	12	10	11	M258		I
MAD070	14	10	11	12	10	10	M258		I
MAD071	12	12	10	15	11	10		M269	R1b
MAD072	13	12	12	15	10	10		U152	R1b
MAD073	12	12	11	15	11	10		DF27	R1b
MAD077	12	12	11	14	10	10		Z196	R1b
MAD078	12	13	10	14	10	10		M269	R1b
MAD080	12	12	11	15	10	10		U106	R1b
MAD081	12	12	11	15	10	10		S356	R1b
MAD082	12	12	12	15	10	10		P312	R1b
MAD083	12	11	10	15	12	11	M272		T
MAD085	13	11	11	14	10	9	M258		I
MAD086	12	12	12	15	10	10		S356	R1b
MAD087	12	11	11	15	10	10		DF27	R1b
MAD088	12	11	11	15	10	10	P170		E
MAD089	12	12	11	15	10	10		DF27	R1b
MAD090	13	11	11	14	10	9	M258		I
MAD091	12	11	12	15	10	10	P170		E
MAD092	12	12	11	15	10	10		DF27	R1b
MAD093	12	12	12	15	10	11		DF27	R1b

Attached Table 3. Continuation.

European Caucasians from Spain (N=219)	DYS388	DYS426	DYS461	DYS485	DYS525	DYS561	Final Y-SNP typed in this study	Final Y-SNP previously typed [1,2,3]	Haplogroup
MAD094	12	11	10	15	10	10	M201		G
MAD095	12	12	11	15	8	10		P312	R1b
MAD097	12	11	12	16	11	11	P170		E
MAD098	16	11	11	17	10	9	P126 (xM258)		J
MAD099	12	12	12	15	10	10		DF27	R1b
A3731	12	12	11	15	10	10		M269	R1b
A3736	12	12	12	15	10	10		DF27	R1b
A3742	12	12	11	15	10	10		P312	R1b
A3752	12	12	12	14	11	10		P312	R1b
A3753	12	12	11	15	10	10		Z278	R1b
A3762	12	12	11	15	10	10		DF27	R1b
A3765	12	12	11	15	10	10		DF27	R1b
A3796	12	12	11	15	10	10		Z278	R1b
A3803	12	12	11	15	10	10		DF27	R1b
A3815	12	12	11	15	11	10		M167	R1b
A3824	12	12	10	16	10	10		S356	R1b
A3827	12	12	11	15	10	9		DF27	R1b
A3871	12	11	11	15	10	10		M269	R1b
A3880	12	12	10	14	10	10		M167	R1b
A3884	12	12	11	15	12	10		U106	R1b
A3889	13	12	11	15	10	10		P312	R1b
A3912	12	12	11	15	10	10		Z278	R1b
A3918	12	12	11	15	10	10		L176.2	R1b
A3933	12	11	10	14	10	10	M201		G
A3940	12	12	11	15	10	10		DF27	R1b
A3958	12	11	10	16	11	10	M272		T
G85e079	12	12	11	15	11	10		DF27	R1b
G85e081	13	12	11	13	10	10		DF27	R1b
G85e083	12	11	13	15	11	11	P170		E
G85e089	12	12	10	15	10	10	M207 (xM269)		R
G85e094	12	12	11	15	10	10		DF27	R1b

Attached Table 3. Continuation.

European Caucasians from Spain (N=219)	DYS388	DYS426	DYS461	DYS485	DYS525	DYS561	Final Y-SNP typed in this study	Final Y-SNP previously typed [1,2,3]	Haplogroup
G85e097	13	11	12	14	11	11	M258		I
G85e098	12	12	11	15	10	10		DF27	R1b
G85e102	12	12	10	14	10	10		P312	R1b
G85e103	12	12	11	15	10	10		Z196	R1b
G85e104	12	11	11	15	13	10	P170		E
G85e109	12	12	11	15	10	10		DF27	R1b
G85e112	12	11	12	15	11	11	P170		E
G85e115	12	12	11	15	10	10		M153	R1b
G85e116	12	12	11	15	10	10		P312	R1b
G85e139	14	11	11	12	10	11	M258		I
G85e140	12	12	11	15	10	10		P312	R1b
G85e142	12	12	11	15	10	10		DF27	R1b
G85e143	12	12	11	15	11	10		DF27	R1b
G85e150	12	12	11	15	10	11		S356	R1b
G85e167	12	12	11	15	10	10		DF27	R1b
G85e168	12	12	11	15	10	10		DF27	R1b
G85e169	12	12	11	15	10	9		DF27	R1b
G85e171	12	12	11	15	10	10		DF27	R1b
G85e173	12	11	13	15	11	10	P170		E
G85e174	12	12	11	15	10	10		M529	R1b
G85e177	12	12	11	15	10	10		DF27	R1b
G85e178	12	12	12	15	10	10		Z196	R1b
G85e212	12	12	11	15	10	9		DF27	R1b
G85e214	12	12	11	16	11	10		P312	R1b
G85e222	12	12	11	15	10	10		Z196	R1b
G85e235	12	12	11	15	10	10		DF27	R1b
G85e243	12	12	11	15	10	10		P312	R1b

Attached Table 3. Continuation.

European Caucasians Autochthonous Basques (N=100)	DYS388	DYS426	DYS461	DYS485	DYS525	DYS561	Final Y-SNP typed in this study	Final Y-SNP previously typed [1,2,3]	Haplogroup
1ftvAPa	12	12	11	15	10	10		DF27	R1b
3vf2sPa	12	12	11	15	10	10		DF27	R1b
5s7zmPa	12	12	10	13	10	10		DF27	R1b
6b45yPa_4A	12	12	11	15	10	10		DF27	R1b
6qY33Pa	12	12	11	15	10	10		P312	R1b
6vDj6Pa	12	12	11	15	10	10		DF27	R1b
7e560Pa	12	12	11	15	10	10		DF27	R1b
7eo6bPa	12	12	11	15	10	10		U152	R1b
8TI0QPa	12	12	11	15	10	10		Z278	R1b
94783Pa	12	12	11	15	10	10		DF27	R1b
99psYPa	12	12	10	15	11	10		DF27	R1b
ART003	12	12	11	15	10	11		DF27	R1b
ART005	13	11	10	15	9	10	M201		G
ART006	12	12	12	15	10	10		DF27	R1b
ART008	12	12	11	15	10	10		DF27	R1b
ART011	12	12	10	14	10	10		P312	R1b
ART013	12	12	11	15	10	10		P312	R1b
ART017	12	12	11	15	10	11		DF27	R1b
ART018	12	12	12	15	10	10		DF27	R1b
ART020	12	12	11	15	10	10		DF27	R1b
ART022	12	12	11	15	10	10		P312	R1b
ART024	12	12	11	15	10	10		DF27	R1b
ART025	12	12	11	15	10	10		DF27	R1b
DaMWyPa	12	12	11	15	10	10		Z278	R1b
DDARG062	12	12	11	15	10	9	M269		R1b
DDARG154	12	12	11	15	10	10	M207 (xM269)		R
DDARG162	12	14	11	15	10	10	M207 (xM269)		R
DDARG171	12	12	11	15	10	10	M207 (xM269)		R
DDARG263	12	12	11	15	10	10	M207 (xM269)		R
DDARG265	12	12	11	15	10	10	M207 (xM269)		R
eLr02Pa	12	12	10	15	10	10		DF27	R1b

Attached Table 3. Continuation.

European Caucasians Autochthonous Basques (N=100)	DYS388	DYS426	DYS461	DYS485	DYS525	DYS561	Final Y-SNP typed in this study	Final Y-SNP previously typed [1,2,3]	Haplogroup
ERE001	12	12	10	14	10	10		P312	R1b
ERE003	12	12	11	14	10	10		DF27	R1b
ERE005	12	12	12	15	10	10		M153	R1b
ERE006	12	12	11	15	10	10		Z278	R1b
ERE008	12	12	11	15	10	10		Z278	R1b
ERE011	12	12	10	14	10	10		P312	R1b
ERE015	12	12	11	15	10	10		DF27	R1b
ERE016	12	12	11	15	10	10		M167	R1b
ERE018	12	12	11	15	10	10		DF17	R1b
ERE020	12	12	11	15	10	11		Z220	R1b
ERE026	12	12	11	16	10	11		Z220	R1b
ERE027	12	12	11	15	10	10		P312	R1b
ERE030	14	12	10	15	10	10		DF27	R1b
ERE033	12	12	10	15	10	10		DF27	R1b
ERE034	12	12	10	14	10	10		P312	R1b
F60C5Pa	12	12	11	15	10	10		P312	R1b
g0u13Pa	12	12	11	15	10	9		DF27	R1b
GhXOGPa	12	12	11	15	10	10		DF27	R1b
gK6XgPa	12	12	11	14	10	10		DF27	R1b
HSCEPPa	12	12	11	15	10	10		Z198	R1b
jUegPa	12	12	11	15	10	10		L176.2	R1b
JvehsPa	12	12	11	15	10	9		DF27	R1b
KOR001	16	11	11	15	10	10	P126 (xM258)		J
KOR002	12	12	11	15	10	10		M153	R1b
KOR004	12	12	10	15	10	10		DF27	R1b
KOR005	12	12	11	15	10	10		P312	R1b
KOR014	12	12	11	15	10	10		M167	R1b
KOR015	12	12	11	16	10	11		Z220	R1b
KOR018	12	12	12	15	10	10		Z278	R1b
KOR019	12	12	12	13	10	10		DF27	R1b
KOR025	12	12	11	15	10	10		M167	R1b

Attached Table 3. Continuation.

European Caucasians Autochthonous Basques (N=100)	DYS388	DYS426	DYS461	DYS485	DYS525	DYS561	Final Y-SNP typed in this study	Final Y-SNP previously typed [1,2,3]	Haplogroup
KOR026	12	12	11	15	10	9		DF27	R1b
KOR035	12	12	11	15	10	10		M167	R1b
KOR037	12	12	11	12	11	10		DF27	R1b
KOR040	12	12	11	15	10	10		P312	R1b
KOR044	12	12	11	15	10	10		Z278	R1b
KOR045	14	11	12	12	10	10	M269		R1b
KOR049	12	12	12	15	10	10		Z278	R1b
KOR053	12	12	12	15	10	10		Z220	R1b
KOR056	12	12	11	15	10	10		Z278	R1b
NAB002	12	12	11	13	10	10		DF27	R1b
NAB005	12	12	11	15	10	10		DF27	R1b
NAB006	12	12	11	15	10	10		DF27	R1b
NAB010	12	12	10	15	10	10		Z195	R1b
NAB014	12	12	11	13	10	10		DF27	R1b
NAB017	12	12	11	15	10	10		Z220	R1b
NAB020	12	12	11	15	10	10		M153	R1b
NAB022	12	12	11	15	10	11		DF27	R1b
NAB023	12	12	11	15	10	10		M167	R1b
NAB024	12	12	11	15	10	10		Z220	R1b
p5z34Pa	12	12	10	15	10	10		P312	R1b
P7Uz5Pa	11	12	11	15	10	10		DF27	R1b
q6hKQPa	12	12	11	15	10	10		DF27	R1b
qBInYPa	12	12	11	15	10	9		DF27	R1b
S1NQDPa	12	12	11	15	10	10		Z278	R1b
VG016	12	12	10	15	10	10	M207 (xM269)		R
VG021	12	11	12	17	10	10	P170		E
VG030	12	12	11	15	10	10	M207 (xM269)		R
VG031	12	12	11	15	10	10		M269	R1b
VG032	12	12	11	14	10	10		M269	R1b
VG038	12	12	11	13	10	10		M269	R1b
VG040	12	12	11	15	10	10		M269	R1b

Attached Table 3. Continuation.

European Caucasians Autochthonous Basques (N=100)	DYS388	DYS426	DYS461	DYS485	DYS525	DYS561	Final Y-SNP typed in this study	Final Y-SNP previously typed [1,2,3]	Haplogroup
VG054	12	12	10	15	9	9		M269	R1b
WAc27	13	11	11	14	11	9	M258		I
Y3TVGPa	12	12	11	15	10	10		DF27	R1b
yaPSYPa	12	12	11	15	10	10		P312	R1b
yKtvgPa	12	12	11	15	10	10		U152	R1b
yOpszPa	12	12	11	15	10	10		P312	R1b
Z2MvcPa	12	12	11	15	10	10		DF27	R1b