Amama
Aita, Ama, Ekaitz
eta Angelari

(c)2015 URTZI LACONCHA SANTAMARIA

ESKER ONAK

Norbaitek inoiz esan zidan bidaiak bi motatakoak bakarrik izan zitezkeela: luzeak eta aspergarriak, edo laburrak eta biziak. Eta, azken finean, zer da ba Doktorego-tesi bat, bizitzako bidaia/etapa bat baino? Nire kasuan horrela izan da behintzat. Baina inolaz ere ez aipatutako multzo bietako batean sailkatzeko bidaia; luzea, bai noski, baina guztiz bizia, gorabehera izugarriekin. Pertsonalki eta profesionalki hazten eta hezten lagundu didan bidaia aberasgarria izan da, ezbairik gabe.

Bidaia luze honetan modu batera edo bestera bidelagun izan dudan zenbaiten zordun naiz. Hasteko gidari izan zaituztedan Andone eta Haritz. Genetikaren arloan, Andone, zuk erakutsi dizkidazu jakintza eta, askotan beharrezkoagoak izan diren, planteamendu molde garrantzitsuenak. Haritz, zuk aldiz, hain arrotzak zitzaizkidan tunidoen ekologia eta arrantza-kudeaketa arloetako ezagutza eta gorabeherak aplikatzen lagundu didazu. Gainera, biek ala biek erakutsi dizkidazue zientzia mundu honen aurpegiak, atseginenak eta baita gordinenak ere. Zuen zuzendaritza, gomendioak eta baikortasuna, zientzia munduko nire lehen pausuentzat bultzada izan dira. Mila esker.

Mikel eta Carmen, eskerrik asko zuei ere. Zuen laguntza, ikuspuntu, eta ideien iraulketa zentzudunik gabe, nekez ikusiko lukete argia artikuluek, eta are gutxiago tesi honek. Genetikako taldera heldu nintzenean “Abereak” zuen izena, eta zuen artean gustora sentiarazi ninduzuen hasieratik. Eskerrik asko taldekide guztiei behar izan zaituztedanean hor egoteagatik: Antropoko garaietako lan & komeriengatik Iratxe, Igor eta Ir (haiek garaiak!), Peitx maisua (animo kamarada), Otsanda eta Aitor aholkuengatik eta Txusa (encurtido invertebrado), filogenien inguruko jarraibideengatik. Eskerrik asko ere zerbitzuko Irati eta Ferri, Taqman teknologian “aditu” bilakatzen laguntzeagatik (jeje). Ezin ahaztu Mikel, Jorge, David eta Alba, taldeko kimu berriak, batez ere informatika alorrean beti laguntzeko prest egoteagatik. Eskerrik asko baita ere Jokin, Unai (hire aitari ere eskerrak luzatu), Haritz, Dani, Carolina, Urko eta Asier, zuen egonaldietan asko lagundu didazue eta. Nahiz eta taldekoa ez izan, eskerrik asko Montse! zure animo eta kafeak momentu aldatsuenak garaitzeko premiazkoak izan baitira.

Mila esker Genetika saileko lagunei, orotariko zalantza edo beharrei erantzuna bilatzen laguntzeagatik, baita ere Inmari, azken txanpako galderak ebazten laguntzeagatik. Eskerrik asko Maialen (zazpi bizitza oparitzeagatik), Iraia, Ane, Olatz^2, Naiara, Jone, One, Koldo, Ainhoa, Eguz, Dorkaitz, Amaia eta Apraizi, bazkalorduak eta kafeak barre artean eta gustura asko pasarazteagatik.

AZTIko Nicolas, Igaratza, Iñigo, Igor, Maite eta Borjari eskerrik asko, Ziburun, Hondarribian edo Cambrils-en laginketa-lan mardulean laguntza emateagatik. Eskerrak baita Jon, Katerin, Josu, Naiara, Iñaki, Alaitz eta Sarairi, askotariko laguntza behar izan dudanean beti

baiezkoa erantzuteagatik. Eskerrik asko Cambrils-en ezagutu zintuztedan Toni, Silvia, Maria eta Joan-i, arrantza-egun ederrengatik, eta baita ardoagatik ere! Neure buruari ere eskerrak, goizaldeko 03`00retako Arrigorriaga-Bermeo, Arrigorriaga-Hondarribi etab.-eko hamaika bidaietan emandako bakarrizketa eta animoengatik.

Ezin ahaztu laborategira gatza eta piperra ekarri zenituen andaluziar-txokoari: Electroni-Miguel, milesker munduaren aurrean zaren modukoa agertzeagatik beti, eta Giltzero, ez zinen adiskidetasun bila etorri baina nik behintzat, lagun handia topatu dut zugan. Zurekin batera Clara ekarri zenuen, mundua ulertzeko ikuspegi ezin alaiagoa erakusten duen pertsona handi hori. Eskerrik asko bioi guztiagatik! Arrantza-taldeari eskerrak, nahiz eta azkenaldian baztertuta zaituztedan, deskonektatzeko paradak eskaintzeagatik, Asturiaseko lupinak zain ditugu!!Eskerrik asko baita ere Goiri, Aitzpea (nahiz eta ez sinetsi, asko lagundu didazu) eta Amalur, batez ere “aitatxo pistatxo” moduko momentuak oparitzeagatik!

Eskerrik asko aita eta ama. Eskerrik asko beti nigan sinetsi izan duzuelako eta behar izan dudanean animatu egin nauzuelako, eta horretan jarraitzen duzuelako. Momentu txarrenetan nirekin egon zarete eta hemendik aurrera, poztasunak eta momentu ederrak ere zuekin konpartitzen jarraitu nahi dut. Mila esker txikitatik bizitzan hain garrantzitsuak diren printzipio, aholku eta onberatasunaren garrantzia helarazteagatik eta haiekin hezteagatik. Eskerrik asko Ekaitz, bizitzako alderdi askotan gidari ezin hobea izan zaitudalako, eta askotan diozun bezala, zuri tokatu izan zaizulako beti bidea irekitzea. Eskerrik asko ere laginketa kontuetan eta behar izan dudanean laguntzeko prest egoteagatik. Hiruroi eskerrak baita ere kobratu gabeko 20 hilabeteetan, eskatu gabe ere, ekonomikoki laguntzeko beti prest egoteagatik.

Bukatzeko, eskerrik asko Angela. Zure laguntza, animo eta sakrifizio infinituengatik. Hasiera-hasieratik hor egon zara, eta bidaia honen bukaeran ere hemen zara ahal zenuen guztian laguntzen, beti bezala nire ondoan. Maiz falta izan zaidan oxigenoa zuk eman didazu, makalguneak zure bultzadek ordezkatu dituzte, munduaren garraztasunei zure gozoki paregabeekin aurre egin diegu, eta garrantzitsuena, kanpoko ankertasunak zure hitzek uxatu dituzte.

Ziur asko eskerrak jaso gabe jende asko utzi dudala, hala bada, bidaide izan zaituztedan guzti-guztioi eskerrik asko eta laster arte!

Índice

Resumen 3

Abstract 4

Capítulo I: Introducción 5

1. El género *Thunnus* 7
    1.1. Sistemática del género *Thunnus* 7
    1.2. Situación poblacional del atún blanco y del atún rojo del Atlántico 9
        1.2.1. Atún blanco 9
        1.2.2. Atún rojo del Atlántico 10
        1.2.3. Importancia socioeconómica de las pesquerías 11
2. Genética marina 12
    2.1. Identificación de la estructuración poblacional 13
    2.2. Conservación y sostenibilidad genética 14
    2.3. Identificación de especies y trazabilidad 16

Hipótesis y objetivos 19

Capítulo II: 25

- Estudio I: SNP discovery in albacore and Atlantic bluefin tuna provides insights into world-wide population structure 27

Capítulo III: 57

- Estudio II: New nuclear SNP markers unravel the genetic structure and effective population size of albacore tuna (*Thunnus alalunga*) 59

Capítulo IV: 85

- Estudio III: Phylogeographic patterns in albacore and phylogenetic relationships between *Thunnus* species inferred from direct mtDNA analysis 87

Capítulo V: 109

- Estudio IV: A genetic traceability tool for albacore stock authentication based on SNP markers 111

Capítulo VI: Discusión general 123

1. Descubrimiento de marcadores SNP en el Atún blanco 128

2. Estructura genética del atún blanco 130
2.1. Mar Mediterráneo 131
2.2. Océano Atlántico y océano Pacífico 133
2.3. Océano Índico 133
3. Aplicación en la gestión del atún blanco 134
3.1. Definición y delimitación de stocks de atún blanco 134
3.2. Estimación del tamaño efectivo del atún blanco del Atlántico Norte 135
4. Relaciones filogenéticas y evolutivas del género *Thunnus* 137
4.1. Estructura poblacional del atún blanco basado en el ADNmt 137
4.2. Evolución del género *Thunnus* 139
5. Desarrollo de una herramienta genética para la autentificación del origen del atún blanco 141
Conclusiones 143
Bibliografía 147

Resumen

RESUMEN

La óptima gestión del atún blanco (*Thunnus alalunga* Bonn., 1978) requiere de una mejor comprensión de la estructura genética-poblacional a nivel global. Con el objetivo de analizar la variabilidad genética, en esta Tesis Doctoral se presentan diferentes aproximaciones para el descubrimiento de polimorfismos de un solo nucleótido (SNPs) en el atún blanco, organismo no-modelo, y la aplicación de estos marcadores en un número de muestras que representan todo el área de distribución de la especie, además de varias generaciones. Dos paneles compuestos por 53 y 58 SNPs nucleares obtuvieron valores de F_{ST} que oscilan entre 0.0 y 0.051 entre puntos de muestreo tras genotipar 1331 individuos de atún blanco procedentes de 26 puntos de muestreo con una distribución mundial. De este modo, se definieron cuatro poblaciones homogéneas: el mar Mediterráneo, el océano Atlántico, el océano Índico y el océano Pacifico. Además, se analizaron las secuencias de cinco regiones mitocondriales (CR, 12S, 16S, Cyt *b* y COI) y nueve SNPs mitocondriales en muestras de los stocks definidos actualmente en el atún blanco. Los análisis filogeográficos mostraron la presencia de dos haplogrupos correlacionados con las distancias geográficas, excepto en el Mediterráneo, el cual muestra un nivel de aislamiento inusual. Finalmente, se estimó el tamaño poblacional efectivo, contemporáneo e histórico, en el atún blanco del Atlántico Norte. Los resultados a este respecto indicaron que no ha existido una reducción de la diversidad poblacional histórica en esta población.

La filogenia del género *Thunnus* basada en las secuencias de las regiones mitocondriales anteriormente mencionadas y las secuencias del genoma mitocondrial completos, no apoya la actual subdivisión de las ocho especies en los subgéneros *Neothunnus* y *Thunnus*. Los análisis de datación evidencian que el atún blanco es la especie más antigua del género, seguida del *T. obesus*. Tras estos eventos de especiación, durante un periodo corto de tiempo, ocurrieron el resto de especiaciones del género. Asimismo, en este trabajo se propone el origen reciente del atún rojo del Pacífico, como resultado de un efecto fundador ocurrido cuando el ancestro común de los atunes rojos del Atlántico y del Pacífico, que vivía en el Atlántico, pobló el océano Pacífico. Este estudio también evidenció la existencia de hibridaciones antiguas y recientes entre especies del género *Thunnus*.

Finalmente, se ha diseñado una herramienta de trazabilidad basado en SNPs nucleares y mitocondriales para que puedan ser utilizados en la prevención de la pesca ilegal, no declarada y no regulada, mediante la certificación de los productos derivados. En esta Tesis Doctoral mostramos que los SNPs son suficientemente informativos a la hora de asignar el origen geográfico individual. De hecho, al utilizar 32 SNPs informativos se mostró que el 68% de los individuos fueron asignados al origen Mediterráneo, Atlántico o Indo-Pacífico con el 100% de fiabilidad.

ABSTRACT

The optimal management of the over-exploited, pelagic tuna, albacore (*Thunnus alalunga* Bonn., 1978) requires a better understanding of population structure than has been provided by previous molecular methods. Despite numerous studies the population structure remains controversial. This PhD thesis reports the different approaches for single nucleotide polymorphisms (SNPs) development in albacore and the application of these SNPs to survey genetic variability across the geographic range of the species. Two panels of 53 and 58 nuclear SNPs yielded values of F_{ST} ranging from 0.0 to 0.051 between samples after genotyping 1,331 albacore individuals collected throughout the worldwide distribution of this species. Results indicated the existence of four genetically homogeneous populations delimited within the Mediterranean Sea, the Atlantic Ocean, the Indian Ocean and the Pacific Ocean. For the North Atlantic Ocean albacore, *short*- and *long-term* effective population sizes were estimated, and results showed no historical decline for this population.

In addition, sequences from five mitochondrial regions (CR, 12S, 16S, Cyt *b* and COI), plus nine mitochondrial SNPs were analyzed in samples from all actually defined stocks of albacore. The phylogeographic analyses revealed the presence of two haplogroups, which frequencies were correlated with geographic distances except for the Mediterranean, which shows an unusual level of isolation also with SNP based analysis.

Regarding the 8 species included in *Thunnus* genus, the five mitochondrial regions above and mitochondrial genome wide sequences don’t support actual subdivision in two subgenera *Neothunnus* and *Thunnus*. Results also evidence hybridization events between species of the genus *Thunnus* either, in this regard, in this PhD thesis has been proposed an ancient introgression of albacore mtDNA in the common ancestor of Pacific and Atlantic bluefin, which then inhabited the Pacific Ocean and originated the Pacific bluefin tuna.

Finally a genetic traceability tool based on both nuclear and mitochondrial SNPs has been designed to be implemented to hinder Illegal, Unreported and Unregulated fishing certifying the labelling of derived products. In this PhD thesis, we show that SNPs are highly informative for assigning individuals to their geographic origin. In fact, applying 32 informative SNPs we found that 68% of individuals were assigned to their Mediterranean, Atlantic or Indo-Pacific origins, with 100% reliability.

Capítulo I

Introducción

INTRODUCCIÓN

1. El género *Thunnus*

1.1 Sistemática del género *Thunnus*

El género *Thunnus* se compone de ocho especies con un registro fósil que se extiende hasta el Eoceno Medio, unos 40 Millones de años atrás (Carrol, 1988; Benton, 1993). Collette (1978) clasificó a los túnidos en dos subgéneros: por un lado el subgénero templado *Thunnus* (grupo de los atunes rojos) y el subgénero tropical *Neothunnus* (grupo del rabil) (Tabla 1.1) basado en la presencia, en el primer grupo, o ausencia, en el segundo, del sistema de intercambio de calor interno. Otras características anatómicas y del esqueleto axial también se tienen en cuenta para la clasificación de las especies en los dos grupos mencionados (Nakamura 1965; Gibbs y Collette 1967).

Tabla 1.1. Clasificación de la especies del género *Thunnus* según Collette (1978), nombre científico, nombre común en español e inglés y el código FAO para cada especie.

	Nombre científico	Nombre común español	Nombre común inglés	Código FAO
Subgénero *Thunnus*				
	Thunnus alalunga	Atún blanco	Albacore	ALB
	Thunnus thynnus	Atún rojo del Atlántico	Atlantic bluefin tuna	BFT
	Thunnus orientalis	Atún rojo del Pacífico	Pacific bluefin tuna	PBF
	Thunnus maccoyii	Atún rojo del sur	Southern bluefin tuna	SBF
	Thunnus obesus	Patudo	Bigeye tuna	BET
Subgénero *Neothunnus*				
	Thunnus albacares	Rabil	Yellowfin tuna	YFT
	Thunnus tonggol	Atún tongol	Longtail tuna	LOT
	Thunnus atlanticus	Atún aleta negra	Blackfin tuna	BLF

El patudo (*Thunnus obesus*) comparte el mismo número de características con cada uno de los subgéneros, sin embargo está clasificado dentro del subgénero *Thunnus* debido a que las

características compartidas con este subgénero se consideran derivadas y de adaptación a aguas más templadas (Collette 1978). La distribución alopátrica y la ausencia de diferencias morfológicas entre el atún rojo del Atlántico (*Thunnus thynnus*) y del atún rojo del Pacífico (*Thunnus orientalis*) han hecho que los taxonomistas considerasen subespecies a estas dos especies (Jones y Silas 1960; Iwai y Nakamura 1964; Iwai *et al.* 1965; Gibbs y Collette 1967). Aunque las secuencias nucleares del espaciador interno transcrito 1 (ITS1) resultaron idénticos en estas dos especies, varios estudios moleculares basados en genes mitocondriales (Sharp y Pirages, 1978; Chow y Inoue, 1993; Chow y Kishino, 1995; Alvarado Bremer *et al.* 1997) han evidenciado que las diferencias genéticas entre dichas especies son similares a las que existen entre las demás especies del género *Thunnus*.

El origen monofilético del género *Thunnus* está apoyado tanto por datos anatómicos (Collette 1978) como por los estudios moleculares. Sin embargo, la mayoría de estudios moleculares no apoyan la división del género en los subgéneros *Thunnus* y *Neothunnus*. Sharp y Pirages (1978) y Chow y Kishino (1995) indicaron la gran divergencia del atún blanco respecto al resto de especies del género *Thunnus*, sugiriendo que el subgénero *Thunnus* no constituye un grupo monofilético. Chow y Inoue (1993) mediante marcadores de Polimorfismos de Longitud de Fragmentos de Restricción (RFLP) y Elliott y Ward (1995) mediante alozimas situaban filogenéticamente al atún rojo del Atlántico y al atún rojo del Sur (*Thunnus maccoyii*) más próximos a las especies del subgénero *Neothunnus* que al atún blanco (*Thunnus alalunga*) y al atún rojo del Pacífico. Chow *et al.* (2006) mediante la comparación de secuencias mitocondriales por un lado y nucleares por otro lado contradecía la división basada en los subgéneros mencionados. Tan solo Alvarado Bremer *et al.* (1997) obtuvo resultados que indicarían un orígen monofilético de ambos subgéneros mediante la secuenciación de la Región Control (CR), aun así no consiguió clasificar al Patudo en ninguno de los dos subgéneros. Los estudios moleculares que combinan el análisis de secuencias mitocondriales y nucleares (Chow *et al.* 2006) apuntan hacia eventos intermitentes de especiación dentro del género *Thunnus*, rechazando así la existencia de los subgéneros mencionados.

En resumen, la existencia de los subgéneros *Thunnus* y *Neothunnus* basada en datos moleculares y las relaciones filogenéticas entre las especies del género sigue siendo un tema inconcluso (Tseng *et al.* 2012). En este sentido la recopilación del máximo número de secuencias mitocondriales disponibles en la base de datos GenBank serviría para definir la historia evolutiva del género *Thunnus*. La mayoría de trabajos se centran en análisis de uno o dos genes para llegar a determinar las relaciones filogenéticas entre las 8 especies del género *Thunnus*, no obstante el enfoque desde múltiples tamaños de secuencias (genoma mitocondrial completo, varios genes concatenados y un solo gen) y la comparación entre los resultados derivados de la mismas podrían arrojar luz sobre la filogenia del género, y referenciar temporalmente los eventos de especiación del género *Thunnus*.

1.2. Situación poblacional del atún blanco y del atún rojo del Atlántico

1.2.1. Atún blanco

El atún blanco, *Thunnus alalunga* (Bonnaterre 1788) es una especie oceánica, epi y mesopelágica, presente en aguas tropicales, subtropicales y templadas (Collette y Nauen, 1983).

Entre las especies del género *Thunnus* es la que presenta la distribución más amplia, ya que es la única que abarca los océanos Pacífico, Índico y Atlántico incluyendo el mar Mediterráneo. Los límites latitudinales de distribución se extienden desde el paralelo 50-55ºN al paralelo 40-45ºS (Arrizabalaga *et al.* 2014). Esta distribución cosmopolita en aguas tropicales y templadas refleja la adaptación a diferentes condiciones oceanográficas, con variaciones muy amplias en temperatura. Esta especie se encuentra entre los 13,5ºC y 25,2ºC, tolerando temperaturas de hasta 9,5ºC durante periodos cortos. (Arrizabalaga 2003).

Es el túnido de menor tamaño dentro del género *Thunnus*. La talla máxima para el atún blanco fue establecida por Collete y Nauen (1983) en 127 cm, y Le Gall (1974) la estimó en 130 cm como una probable talla máxima en el Atlántico. Aunque ningún atún marcado ha permanecido en libertad por más de diez años (Arrizabalaga 2003; Santiago y Arrizabalaga 2005) la longevidad teórica calculada es de 15 años (Le Gall, 1974). Se asume que el 50% de los peces han alcanzado la madurez sexual a los 90 cm (edad de 5 años) en el Atlántico (Bard, 1981) y a los 62 cm en el Mediterráneo (Arena *et al.* 1980). Comparado con el ciclo de vida de otras especies de interés comercial (la anchoa europea *Engraulis encrasicolus*; Lenguado *Xystreurys rasile*) el del atún blanco es relativamente largo. Este parámetro hace que la recuperación tras el efecto de la sobrepesca pueda ser más lenta (Collette *et al.* 2011).

Aunque la reproducción del atún blanco es muy poco conocida respecto a otras especies de túnidos (Juan-Jordá *et al.* 2013), la estrategia sexual consiste en un amplio período de puesta con varias liberaciones de huevos en diferentes episodios en aguas con temperaturas altas. Temperaturas por encima de 24ºC y una termoclina profunda parecen estimular la actividad reproductora en los túnidos. Las zonas de desove del atún blanco en el Atlántico se encuentran en áreas occidentales subtropicales de ambos hemisferios: en las aguas frente a Venezuela, mar de los Sargazos (Le Gall 1974, Nishikawa *et al.* 1985), el Golfo de México (Richards 1984) y frente a la costa este de Brasil (Beardsley 1969, Koto 1969). Dentro del mar Mediterráneo se ha descrito la presencia de larvas en diversas áreas (Ueyanagi 1971, Dicenta 1975, Lalami *et al.* 1973, Piccinetti y Manfrin 1993, Marano 1999, García *et al.* 2002).

Cuatro Organizaciones Regionales de Ordenación Pesquera (RFMO) gestionan las pesquerías de túnidos a nivel mundial (Figura 1.1):

- **ICCAT/CICAA:** Comisión Internacional para la Conservación del Atún Atlántico (encargada de los túnidos del Atlántico y Mediterráneo).
- **IOTC:** Comisión del Atún para el Océano Índico (encargada de los túnidos del Índico).
- **WCPFC:** Comisión de Pesca del Pacífico Occidental y Central (encargada de la zona occidental y central del Pacífico).
- **IATTC/CIAT:** Comisión Interamericana del Atún Tropical (encargada de la zona oriental del Pacífico).

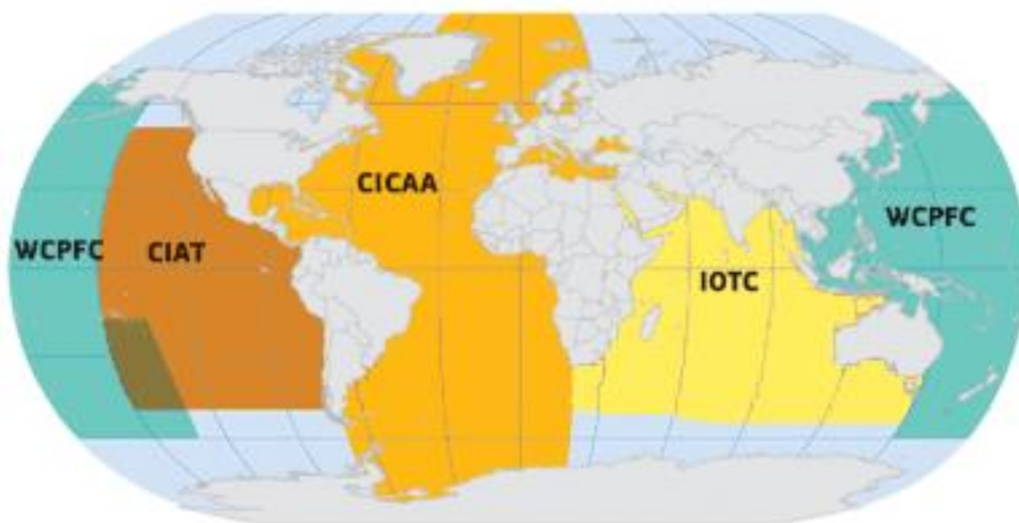


Figura 1.1.: Organizaciones Regionales de Ordenación Pesquera (RFMO) que se dedican a gestionar los stocks del atún blanco (Modificado de "Comunicación de la comisión al parlamento europeo, al consejo, al comité económico y social europeo y al comité de las regiones sobre la dimensión exterior de la Política Pesquera Común" Bruselas, 13.7.2011 COM(2011) 424 final).

Dichas RFMO gestionan el atún blanco considerando seis stocks: (i) Mar Mediterraneo, (ii) Atlántico Norte, (iii) Atlántico Sur, (iv) Índico (v) Pacífico Norte y (vi) Pacífico Sur.

1.2.2. Atún rojo del Atlántico

El atún rojo del Atlántico (*Thunnus thynnus* Linnaeus 1758), también conocido como Cimarrón, vive sobre todo en aguas templadas del Atlántico Norte y el Mediterráneo (Nakamura 1969; Collette y Nauen 1983; Fromentin y Fonteneau 2001; Riccioni *et al.* 2011). Ocupa preferentemente las aguas superficiales y sub-superficiales aunque puede sumergirse hasta profundidades de más de 1.000 m (Block *et al.* 2001) siendo capaces de mantener la temperatura corporal.

A los 20 años, un atún rojo alcanza los 270 cm y 400 kg, es la especie que mayor tamaño alcanza entre las especies del género *Thunnus*. Es una especie longeva, con un ciclo vital de aproximadamente 40 años (Cort 2004). Un ciclo de vida extenso, una madurez sexual tardía (alrededor de 4-5 años para el Atlántico este), zonas restringidas de puesta y épocas reproductivas cortas (1 a 2 meses) son características que la convierten en una especie susceptible de colapso bajo una presión pesquera excesiva y continua (Collette *et al.* 2011).

La Comisión Internacional para la Conservación del Atún Atlántico (ICCAT) es la organización pesquera intergubernamental responsable de la conservación del atún rojo del Atlántico. Desde 1980 se definen dos stocks basándose en la distribución discontinua de las capturas históricas en el Atlántico (Fromentin y Powers, 2005). Las dos unidades de gestión se separan por el meridiano 45ºW: (i) Atlántico Oeste y (ii) Atlántico Este (incluyendo el Mediterráneo Actualmente existe un plan de recuperación del atún rojo (ICCAT Recomendación [14-04]) que incluye la implementación de una Captura Total Permitida (TAC) anual, así como otras medidas suplementarias para incentivar la recuperación de los stocks.

1.2.3. Importancia socioeconómica de las pesquerías

Actualmente hay importantes pesquerías de túnidos comerciales y deportivas en todo el mundo debido al alto valor comercial que alcanzan en las ventas tanto en fresco como en conserva. Del género *Thunnus*, el atún blanco es una de las la especies de mayor valor comercial para conserva y la cuarta en cuanto a capturas totales (FAO 2011). Las capturas mundiales de atún blanco han aumentado considerablemente desde los años 50, fluctuando en torno a las 200 000 toneladas anuales en las últimas cuatro décadas (Figura 1.2).

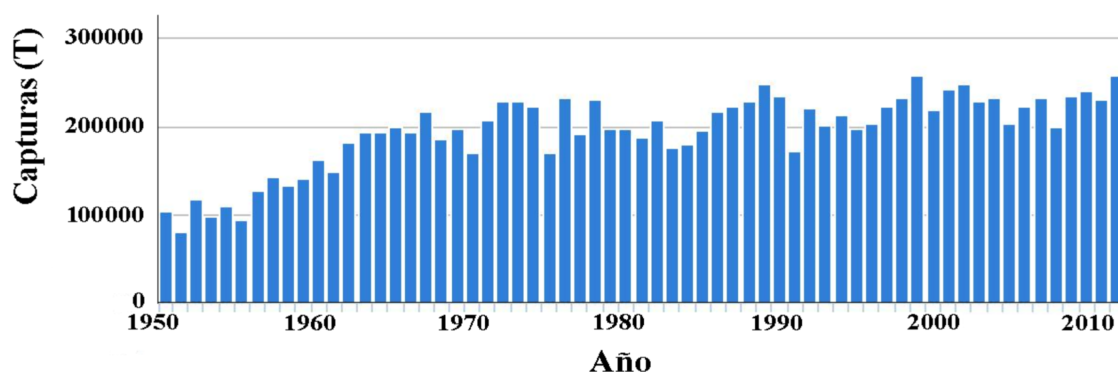


Figura 1.2. Evolución de las capturas globales en toneladas de atún blanco desde 1950 hasta el 2012 (modificado de FishStatJ- FAO 2014).

El valor económico del atún rojo del Atlántico es mucho más alto debido al reciente desarrollo del marcado de sushi y sashimi en los años 80, alcanzando precios desorbitados en los

mercados japoneses (Fromentin y Powers, 2005). Por ejemplo, un solo ejemplar de 222 kg fue vendido por 1,38 millones de euros en 2013 en el mercado de Tsukiji de Tokyo. Las capturas de la especies alcanzaron su máximo a mediados de los 90 y en los últimos años se han reducido debido a la aplicación de los TACs anuales (Figura 1.3).

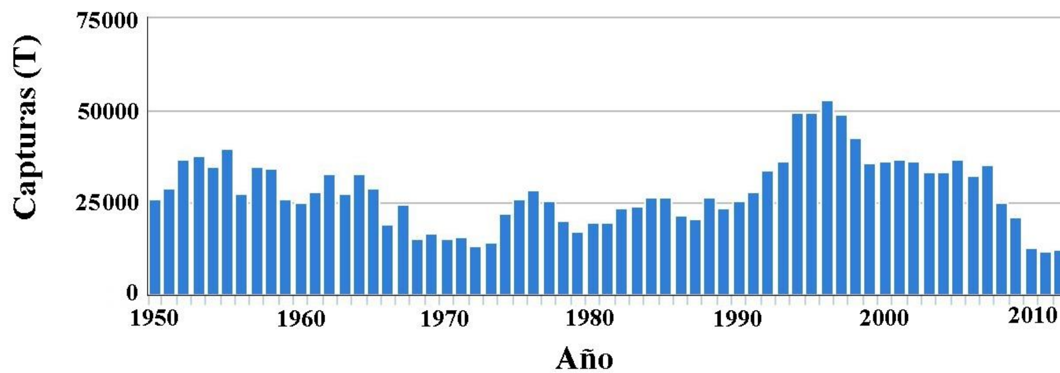


Figura 1.3. Evolución de las capturas globales en toneladas de atún rojo del Atlántico desde 1950 hasta el 2012 (modificado de FishStatJ- FAO 2014).

Las características propias de ambas especies (ciclo de vida extenso, zonas de puestas poco conocidas y épocas reproductivas cortas) sumado al alto valor económico hacen que estas especies sean especialmente vulnerables al efecto de la sobrepesca (Collette *et al.* 2011). Está vulnerabilidad hace que las medidas de gestión que se tomen entorno a las dos especies tengan que ser efectivas a largo plazo. Para ello, la información multidisciplinar que las diferentes metodologías científicas aporten al aumento del conocimiento sobre los stocks explotados (estructura poblacional, migraciones, variabilidad genética, etc.) debe ser fiable, actualizada y útil para los gestores que vayan a tomar las decisiones oportunas.

2. Genética marina

Frente a la problemática de la sobreexplotación pesquera y el desconocimiento sobre la estructura poblacional de muchas especies marinas la genética marina y la biotecnología ayudan al mejor entendimiento de los modelos y procesos que influyen en la biodiversidad de los océanos. Desde el desarrollo de la primera metodología de secuenciación de Frederick Sanger en los años 70 la investigación genética ha experimentado una rápida y considerable expansión. Así, la aplicación de la genética en especies marinas explotadas ha mejorado el entendimiento de las pautas de abundancia, distribución y estructura poblacional de diferentes especies y ha supuesto una revolución conceptual de las teorías ecológicas y evolutivas (Hauser y Seeb, 2008). Además, la reciente secuenciación de un amplio rango de genomas de peces está arrojando una gran cantidad de

indicios sobre la evolución de los genomas de los vertebrados en general (Roest y Weissenbach, 2005) y de los peces en concreto.

2.1. Identificación de la estructuración poblacional

El concepto de stock o población es esencial para el estudio de las especies explotadas, ya que es la unidad básica que se utiliza para la definición y aplicación de las políticas de gestión pesquera. De este modo la definición de stocks (o unidades de gestión) y su delimitación son conceptos claves para llegar a obtener un manejo sostenible de las especies explotadas. Para conseguir este fin, la utilización de múltiples aproximaciones en el estudio de la estructura poblacional es un enfoque que ayuda a obtener una información completa, ya que cada técnica por separado sólo aporta conocimientos relativos a aspectos específicos de la población (Ihssen *et al.* 1981; Ruzzante *et al.* 1999; Waldman *et al*. 1997). Desde el punto de vista biológico y genético resulta más útil una definición de stock con integridad genética (Jamieson 1973; Ovenden 1990) que permita la conservación de la diversidad genética de los recursos a largo plazo, siendo esta última la visión considerada en el presente trabajo.

La gestión incorrecta de las pesquerías, es decir, alejada de la definición de unidades de gestión acordes a la estructura poblacional real de la especie, puede acarrear una reducción poblacional y en consecuencia, una reducción de la productividad (Hauser y Carvalho, 2008) e incluso llevar al colapso a alguna de las pesquerías (Fu y Fanning 2004). Además las poblaciones sobreexplotadas pueden sufrir una pérdida de la variabilidad genética (Hauser *et al.* 2002) que se puede traducir en una pérdida de la capacidad de respuesta adaptativa de la especie frente a un medio en constante cambio.

Diferentes metodologías son aplicadas en el análisis de la estructura poblacional de túnidos: marcado-recaptura (Arrizabalaga *et al.* 2004), elementos traza en otolitos (Macdonald *et al.* 2013) y datos morfométricos (Bard 1981). Entre los marcadores moleculares, los análisis de grupos sanguíneos (Arrizabalaga *et al.* 2004), secuencias nucleares y mitocondriales (Nakadate *et al.* 2005) y microsatélites (Davies *et al.* 2011; Montes *et al.* 2012) no han conseguido resolver de modo consensuado la estructura poblacional del atún blanco a nivel mundial.

Actualmente los Polimorfismos de Nucleótido Único (SNP) son los marcadores moleculares candidatos en los estudios de estructuración poblacional frente a otros marcadores como microsatélites. Los SNPs son más abundantes en el genoma, y al ser más simples (no se basan en el número de repeticiones como los microsatélites) su genotipado también es más sencillo. Sin embargo, para obtener el mismo grado de discriminación que los microsatélites se necesita un mayor número de SNPs (Morin *et al.* 2004). No obstante, las tecnologías de secuenciación masiva actuales permiten el descubrimiento y validación de un gran número de SNPs distribuidos por todo el genoma. Estas técnicas son particularmente útiles para las especies no-modelo, que no cuentan

con un genoma de referencia (Metzker, 2010). En el presente trabajo, se ha acoplado la secuenciación masiva con la amplificación de marcadores entre diferentes especies (amplificación cruzada entre especies - *cross species amplification*). SNPs descubiertos en el atún rojo del Atlántico por secuenciación masiva, han sido testados por amplificación en otra especie filogenéticamente cercana, el atún blanco, lo que nos ha permitido validar en el atún blanco un gran número de marcadores descubiertos previamente en el atún rojo del Atlántico. La misma validación también se ha llevado a cabo en el sentido contrario; es decir, SNPs descubiertos en el atún blanco se validaron en el atún rojo del Atlántico. Existen varios trabajos para microsatélites (Takagi *et al.* 2001; Montes *et al.* 2012) y también para SNPs (Helyar *et al.* 2012) que sirven de ejemplo.

Por otro lado, las secuencias mitocondriales permiten analizar procesos que guían la distribución geográfica de linajes de genes dentro de una especie determinada (análisis filogeográficos) (Lourie y Vincent, 2004; Viñas *et al.* 2004; Kunal *et al*. 2014) así como el análisis de las relaciones genéticas y de la historia evolutiva entre especies cercanas (análisis filogenéticos) (Chow y Kishino 1995; Alvarado Bremer *et al.* 1997).

En el presente trabajo, centrado principalmente en el atún blanco, se ha realizado el análisis de su estructura poblacional a nivel mundial mediante la aplicación de dos paneles diferentes de SNPs nucleares, y un estudio filogeográfico mediante la secuenciación de 5 genes mitocondriales. Además, también se analizó la estructura poblacional del atún rojo del Atlántico, mediante amplificación cruzada de SNPs descubiertos en el atún blanco. Por último, y aprovechando la disponibilidad de secuencias en la base de datos GenBank, se han estudiado las relaciones filogenéticas entre las ocho especies que componen el género *Thunnus*.

2.2. Conservación y sostenibilidad genética

La evaluación del grado de variabilidad genética dentro de las poblaciones y/o entre las poblaciones, o incluso entre diferentes generaciones de una misma población, es un concepto estrechamente ligado a la conservación genética. Constituye una medida que aporta información relevante sobre la "salud genética" de una población explotada (o susceptible de ser explotada), ayudando a evaluar la probabilidad de recuperación ante diferentes grados de explotación. El peso que adquiere este tipo de evaluaciones en los programas de conservación es cada vez mayor, ya que la capacidad de resistencia (supervivencia) a la sobreexplotación, fragmentación y otras perturbaciones antrópicas o de cambios del medio en el que habitan reside en la diversidad genética de las poblaciones.

La mayoría de las pesquerías europeas están sobreexplotadas, y sufren una reducción del tamaño censal alarmante. La cuestión está en si el descenso drástico en el número de individuos ha dado lugar a una reducción de la diversidad genética del atún blanco. Para dar respuesta, estimamos el tamaño efectivo de la población (N_e), un concepto que incide en la viabilidad poblacional. N_e se

define como el número de reproductores en una población ideal que muestra la misma tasa de cambio en las frecuencias alélicas debido a la deriva genética que la población real (Wright, 1931) y está estrechamente ligado a la diversidad genética que presenta esa población en el estado actual, y que predice su evolución. Así, N_e informa sobre el riesgo de extinción derivado de una disminución en la diversidad genética, o sobre el impacto de la pesca sobre la viabilidad de un stock de interés especial, o sobre la adecuación de las medidas de regulación pesquera implementadas a la sostenibilidad de la especie.

Existen tres rangos temporales en las que el tamaño poblacional efectivo se puede definir: i) **contemporáneo** (incluye una o unas pocas generaciones atrás) ii) **histórico** (incluye las últimas decenas-miles de generaciones) y iii) **antiguo** (incluye las últimas miles-millones de generaciones) (Wang 2005). En esta Tesis Doctoral se hace referencia a los dos primeros rangos temporales. Las diferentes estimas de N_e difícilmente se pueden comparar ya que hacen referencia a diferentes escalas espaciales y diferentes rangos temporales (Schwartz *et al.* 1998).

La mayoría de los trabajos que estiman los tamaños efectivos de especies marinas están basados en marcadores microsatélites, como es el caso del corvinón ocelado (*Sciaenops ocellatus*) (Turner *et al.* 2002), la solla (*Pleuronectes platessa*) (Hoarau *et al.* 2005) o el bacalao del Atlántico (*Gadus morhua*) (Poulsen *et al.* 2006). En menor medida existen estudios que utilizan marcadores como RFLPs en la lubina estriada (*Morone saxatilis*) (Diaz *et al.* 2000), alozimas en el sargo común (*Diplodus sargus*) (Lenfant y Planes, 2002) y secuencias mitocondriales en el pez espada (*Xiphias gladius*) (Alvarado Bremer *et al*. 1995). En cuanto a los túnidos existen varios trabajos que aportan estimaciones del N_e, todos ellos basados en marcadores microsatélites o secuencias mitocondriales (Riccioni *et al.* 2010; Qiu y Miyamoto 2011, Qiu *et al.* 2013; Bravington et al. 2014 y Nomura *et al.* 2014). De todos ellos tan solo uno presenta estimaciones de N_e del atún blanco (Qiu *et al.* 2013) y ninguno utiliza marcadores SNP.

En el presente trabajo se ha estimado el tamaño efectivo (N_e) del atún blanco del Atlántico norte. Para ello, se realizó un muestreo temporal con una separación de hasta 24 años. Se obtuvieron dos estimas de N_e: 1) N_e contemporáneo, estimado a partir de la fluctuación de las frecuencias alélicas entre cohortes y 2) N_e histórico basado en la teoría de coalescencia. Las comparaciones entre las N_e obtenidas respecto a la estimas del tamaño censal (N_c) del atún blanco del Atlántico norte son indicadores de la viabilidad genética del stock estudiado.

2.3. Identificación de especies y trazabilidad

En la misma línea de la sobrepesca, la pesca ilegal y la pesca no declarada perjudican la recuperación de especies que han visto reducidas sus poblaciones (Beddington *et al.* 2007). Estas prácticas provocan la pérdida de alrededor de 580 millones de dólares anuales, tan sólo teniendo en cuenta la pesca de túnidos. En el caso del atún rojo del Atlántico, se cree que la mitad de las capturas de atún rojo del Mediterráneo fueron no declaradas a finales de los años 90 (ICCAT 2014). Además del perjuicio económico, las prácticas ilegales provocan daños importantes en el medio marino debido a que suelen llevar a cabo prácticas pesqueras insostenibles (Agnew *et al.* 2009). Por ello, para hacer frente al fraude en la declaración de especies pescadas u origen de las mismas hace falta desarrollar herramientas cada vez más específicas. Existe un gran número de trabajos que describen diferentes métodos moleculares que permiten discriminar entre las especies de túnidos (Bartlett *et al.* 1991; Takeyama et al. 2001; Pardo *et al.* 2004; Michelini *et al.* 2007; Lowenstein *et al.* 2009; Viñas y Tudela, 2009; Santaclara *et al.* 2014). Estas herramientas son muy útiles para identificar casos de fraude en productos enlatados, congelados y en fresco y sirven también para identificar irregularidades en los desembarcos de capturas.

Existen referencias de trabajos de asignación de origen poblacional para el bacalao (*Gadus morhua*), la merluza europea (*Merluccius merluccius*), el arenque del Atlántico (*Clupea harengus*), la anchoa europea (*Engraulis encrasicolus*) y el lenguado común (*Solea solea*). Sin embargo, tan solo existe una referencia en túnidos (Patente ES 2 392 293 B1) diseñada para determinar el origen del atún blanco. La asignación del origen geográfico de un individuo permitiría un mayor control de la pesca ilegal, el fraude en la comercialización y en consecuencia, maximizaría los esfuerzos que se realizan actualmente en la conservación de un stock en concreto. La posible pesca no declarada que sea identificada sirve para que las estimas de mortalidad pesquera sean más rigurosas, permitiendo una gestión más eficiente de los stocks. Por otro lado el consumidor de pescado es cada vez más consciente de la importancia que debe tener la sostenibilidad en la pesca, por lo que un etiquetado que sea garantía de sostenibilidad es cada vez más apreciado por el consumidor. Para garantizar la protección de los stocks y los derechos del consumidor todos los productos pesqueros consumidos en la Unión Europea deben tener identificado su origen geográfico (McCluskey y Lewison, 2008).

Es posible obtener el origen geográfico o poblacional individual si se aplican metodologías de alta resolución. En este sentido los marcadores SNPs son una herramienta muy útil por su abundancia en el genoma y por la facilidad de descubrirlos mediante secuenciación comparativa y, cada vez más, mediante Secuenciación de Nueva Generación (NGS) (Metzker, 2010). Hoy en día los marcadores SNPs suponen la herramienta molecular más rentable en el campo de la trazabilidad de peces marinos (Nielsen *et al.* 2012).

En cualquier caso, además de utilizar el tipo de marcador o marcadores suficientemente informativo para cada objetivo (caracterización poblacional, cálculo de tamaño efectivo, trazabilidad…) otra cuestión no menos importante es la cobertura muestral que se abarca y el número de individuos representativo de cada muestra. Algunos estudios dan como inconcluyentes sus resultados o recomiendan mejorar los análisis debido a que el número de muestras no es suficiente (Viñas *et al.* 1999) o porque no se ha abarcado toda el área de distribución de la especie (Davies *et al.* 2011).

En este trabajo se ha utilizado un total de 137 SNPs para definir un panel que posibilite la asignación de cualquier individuo de atún blanco a su lugar de origen. En definitiva, una herramienta fiable que permita la trazabilidad de los atunes desde el momento de pesca hasta el procesado de los mismos. Para ello se ha contado con un número muestral suficiente tanto en el diseño del panel de referencia como para testar el mismo.

Hipótesis y Objetivos

HIPÓTESIS Y OBJETIVOS

El alto valor comercial del atún blanco (*Thunnus alalunga*) implica que aumente la probabilidad de que se explote por encima del rendimiento máximo sostenible (MSY).Por ello, los gestores de las pesquerías deben aplicar las acciones regulatorias sobre unidades de gestión acordes a la estructura poblacional real de la especie. Las herramientas moleculares, y entre ellas los marcadores SNPs, ayudan en la correcta definición de las unidades de gestión (stocks) y son cada vez más utilizadas en este campo de la investigación. Además de la estructura poblacional, estos marcadores sirven como indicadores de la posible pérdida de variabilidad genética de los stocks debida al efecto de la sobrepesca. Asimismo, las secuencias y los SNP mitocondriales son útiles para clarificar filogenias y filogeografía. Y los diferentes tipos de SNPs sirven para mejorar la trazabilidad aplicada en convatir la pesca ilegal y el fraude en los productos pesqueros.

Desde una perspectiva genética, son varios los aspectos del atún blanco que resulta interesante resolver:

i) No existe un consenso en cuanto a la estructura poblacional mundial de la especie.

ii) No se conoce el impacto que ha podido tener la sobrepesca en el tamaño efectivo poblacional, y por consiguiente en la diversidad genética de las poblaciones de atún blanco.

La resolución de ambas cuestiones permitirá delimitar el número de stocks mínimo a considerar para la gestión sostenible de la especie, y también valorar la salud genética de la especie. Estas respuestas suponen una información esencial a la hora de adoptar medidas de conservación para la especie y evaluar el funcionamiento de estas medidas.

iii) En el caso del género *Thunnus,* las relaciones filogenéticas entre las 8 especies que lo componen sigue siendo un tema no resuelto. Tampoco se conocen con exactitud las dataciones de los eventos de especiación y cladogénesis del género.

HIPÓTESIS

Las hipótesis de esta Tesis Doctoral que engloba el atún blanco y el género *Thunnus* son:

Respecto a la estructuración poblacional del atún blanco:

-H0: A nivel mundial el atún blanco se compone de seis poblaciones que se corresponden a los seis stocks definidos por las Organizaciones Regionales de Ordenación Pesquera.

-H1: A nivel mundial el atún blanco se compone de un número diferente de poblaciones que no se corresponde exactamente con los seis stocks definidos por las Organizaciones Regionales de Ordenación Pesquera.

Respecto a la evaluación del N_e en el océano Atlántico:ç

-H0: La sobrepesca ejercida en el stock del Atlantico norte ha producido una disminución en el número censal pero no en tal magnitud que haya afectado al tamaño efectivo poblacional, y por tanto tampoco ha reducido la diversidad genética de una manera significativa.

-H1: La sobrepesca ejercida en el stock noratlántico ha producido una disminución en el número censal que ha afectado al tamaño efectivo poblacional, y por tanto la diversidad genética se ha visto disminuida.

Respecto al estudio filogenético del género *Thunnus*:

-H0: Los datos genéticos apoyan la división actual del género *Thunnus* en dos subgéneros basada en datos anatómicos y fisiológicos.

-H1: Los datos genéticos, cuestionan la división actual del género *Thunnus* en dos subgéneros basada en datos anatómicos y fisiológicos.

OBJETIVOS

Los objetivos generales de esta Tesis Doctoral son, por un lado, la caracterización genética del atún blanco (*Thunnus alalunga*) con el objeto de contribuir a su gestión sostenible, y por otro, aportar nuevo conocimiento sobre su historia evolutiva enmarcada dentro del género *Thunnus*. Para ello se han definido cuatro objetivos específicos:

i)- Resolver la estructura genética del atún blanco a nivel mundial y clarificar la estructuración temporal del atún blanco en el Golfo de Bizkaia, mediante el descubrimiento de SNPs y una recolección de muestras con gran cobertura espaciotemporal. La resolución de la estructuración mundial y temporal supondrá una ayuda en el asesoramiento de la toma de medidas regulatorias por parte de las RFMO.

ii)- Evaluar la posible pérdida de variabilidad genética del atún blanco en el Atlántico norte, mediante la estima del N_e contemporáneo e histórico basada en el muestreo de cohortes que abarque el mayor número de generaciones posible. Estas evaluaciones permitirán conocer si la sobreexplotación ejercida en el atún blanco del Atlántico norte ha supuesto un cuello de botella para el stock.

iii)- Comprender las relaciones filogenéticas entre las especies que componen el género *Thunnus* y datar los eventos de especiación mediante múltiples secuencias mitocondriales. Esto supondría el establecimiento de pautas evolutivas dentro del género y solventar la controversia existente hoy en día entre los datos anatómicos y moleculares.

iv)- Mejorar la herramienta de trazabilidad existente para diferenciar el origen geográfico del atún blanco, mediante la incorporación de SNPs nuevos a los SNPs utilizados en la herramienta

molecular anterior. Si se consiguiese una eficacia del 100%, esta nueva herramienta de trazabilidad resultaría un instrumento útil para combatir el fraude comercial y las prácticas pesqueras ilegales.

Capítulo II

SNP discovery in albacore and Atlantic bluefin tuna provides insights into world-wide population structure

ESTUDIO I

SNP discovery in albacore and Atlantic bluefin tuna provides insights into world-wide population structure

A. Albaina, M. Iriondo, I. Velado, U. Laconcha, I. Zarraonaindia, H. Arrizabalaga, M.A. Pardo, M. Lutcavage, W. S. Grant, A. Estonba

Abstract

The optimal management of the commercially important, but over-exploited, pelagic tunas, albacore (*Thunnus alalunga* Bonn., 1978) and Atlantic bluefin tuna (*T. thynnus* L., 1758), requires a better understanding of population structure than has been provided by previous molecular methods. Despite numerous studies of both species, their population structures remain controversial. This study reports the development of single nucleotide polymorphisms (SNPs) in albacore and Atlantic bluefin tuna (BFT) and the application of these SNPs to survey genetic variability across the geographic ranges of these tunas. A total of 616 SNPs were discovered in 35 albacore tuna by comparing sequences of 54 nuclear DNA fragments. A panel of 53 SNPs yielded values of F_{ST} ranging from 0.0 to 0.050 between samples after genotyping 460 albacore collected throughout the distribution of this species. No significant heterogeneity was detected within oceans, but between-ocean comparisons (Atlantic, Pacific and Indian oceans along with Mediterranean Sea) were significant. Additionally, a 17 SNPs panel was developed in Atlantic BFT by cross-species amplification in 107 fish. This limited number of SNPs discriminated between samples from the two major spawning areas of Atlantic BFT (F_{ST} = 0.116). The SNP markers developed in this study can be used to genotype large numbers of fish without the need for standardizing alleles among laboratories.

Keywords *Thunnus alalunga*, *Thunnus thynnus*, Single Nucleotide Polymorphism (SNP), SNP discovery, population genetics, fisheries management.

1. Introduction

Molecular genetics has led to considerable progress in understanding the ecologies of marine species by providing new insights into the demographic and evolutionary dynamics of wild populations (Hauser & Carvalho 2008). Genetic markers are widely used to identify stocks, to estimate mixed stocks in a fishery, to monitor genetic diversity within populations and to measure connectivity between populations, among many other applications (e.g. Nielsen *et al.* 2009;

Waples & Naish 2009). These studies have overturned the classic notion that large marine populations are genetically homogeneous with limited local adaptation by showing extensive genetic population structure in many marine species (reviewed in Hauser & Carvalho 2008). Moreover, genetic structuring has been reported even across small spatial scales (e.g. Knutsen *et al.* 2003; Jørgensen *et al.* 2005; Knutsen *et al.* 2007; Knutsen *et al.* 2011). However, data are still rare for the vast majority of highly exploited species, even though a large number of studies on genetic population structure of marine fish have been published in the past decades. Genetic studies are needed to improve the management of species for which stock structure and migration patterns are still unclear. An improper management of these fishery resources can lead to the extirpation of small independent stocks or to the under-utilization of large stocks.

This study focuses on two widely distributed pelagic tunas, albacore (*Thunnus alalunga* Bonn., 1978) and Atlantic bluefin tuna (*T. thynnus* L., 1758). Albacore is one of the smallest tunas and Atlantic bluefin tuna (BFT) one of the largest in the family Scombridae. While albacore is a widely distributed species, inhabiting both temperate and tropical pelagic waters of all oceans, the distribution of Atlantic BFT is limited to the North Atlantic Ocean and Mediterranean Sea (e.g. Nakamura 1969; Collette & Nauen 1983; Fromentin & Fonteneau 2001). Both species coexist in the Mediterranean Sea. Harvests of these species are large and of high economical value, especially Atlantic BFT, which is sold for high prices in Japanese fish markets (Magnuson *et al.* 1994). An important Atlantic BFT aquaculture industry, based on the fattening of locally collected fish in floating cages, has developed in the Mediterranean. Moreover, these tunas' life history traits make them susceptible to collapse under continued excessive fishing pressure, as their population growth rates are low (De Roos & Persson 2002). Critical life-history traits include long life spans, large body sizes, late sexual maturity (around 4-5 years but up to 8 years for the Western Atlantic BFT), geographically restricted spawning sites, and relatively short spawning periods of 1 or 2 months (Fromentin & Fonteneau 2001; Fromentin & Powers 2005; Rooker *et al.* 2007; Fromentin 2009; Juan-Jordá *et al.* 2011).

Since stocks of albacore and Atlantic BFT are currently overexploited, an urgent need exists to improve conservation and management efforts, including the development of alternative methods of population assessment (Juan-Jordá *et al.* 2011; Collette *et al.* 2011). The management of these tunas has to be at the population level, because the extent and dynamics of population structuring underlies the resilience and sustainability of harvested populations. Previous studies of population structure have used four classes of molecular markers: allozymes, blood groups, mitochondrial RFLPs and microsatellite loci (Chow & Ushiama 1995; Yeh *et al.* 2007; Takagi *et al.* 1999; Takagi *et al.* 2001; Ely *et al.* 2002; Pujolar *et al.* 2003; Arrizabalaga *et al.* 2004; Viñas *et al.* 2004; Nakadate *et al.* 2005; Arrizabalaga *et al.* 2007; Carlsson *et al.* 2007; Boustany *et al.* 2008; Riccioni *et al.* 2010; Davies *et al.* 2011; Viñas *et al.* 2011). Despite these efforts to describe stock structure in albacore and Atlantic BFT, the population structures of these species remain controversial (Arrizabalaga *et al.* 2004; Walli *et al.* 2009; Galuardi *et al.* 2010). Presently, albacore populations are divided into six management units (namely the Mediterranean, North

Atlantic, South Atlantic, Indian Ocean, North Pacific and South Pacific stocks) and Atlantic BFT into two units (namely the Western Atlantic stock and the Eastern Atlantic and Mediterranean stock). However, the results of population surveys based on microsatellite variability illustrate that these management units might not be consistent with the genetic structures of both species (Riccioni *et al.* 2010; Davies *et al.* 2011; Viñas *et al.* 2011).

Two factors explain the current lack of consensus on genetic structure. First, only one of the previous genetic studies (Montes *et al.* 2012) included samples over the entire distributional areas of these tunas. Second, none of the previous studies used large numbers of molecular markers, such as multiple single nucleotide polymorphisms (SNPs), which can be assayed rapidly in large numbers of fish to yield high statistical power to test population-level genetic hypotheses (e.g. Ogden 2011; Helyar *et al.* 2011).

The goals of the present study were to develop SNP markers in albacore and Atlantic BFT and to use these markers to make a preliminary survey genetic variability among populations over the geographic ranges of these tunas (Table 2.1., Fig. 2.1.). The development of SNP markers will aid in the ability to conduct collaborative studies among laboratories without the need for standardizing alleles. Further, the development of a large number of markers will increase the power of genetic analysis to detect the small differences among populations that are expected in high gene-flow species, such as the tunas.

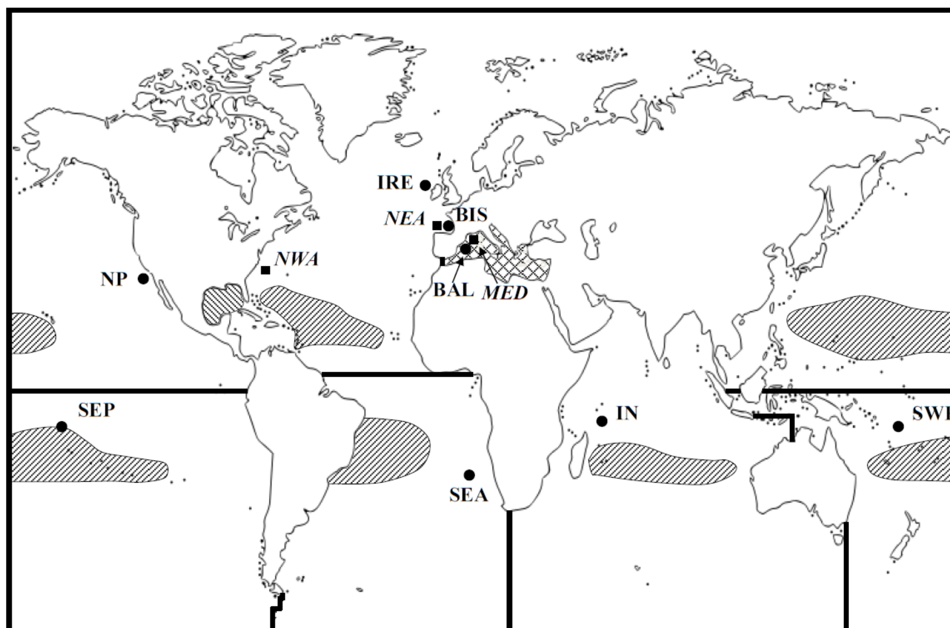


Figure 2.1. Sampling locations and approximate locations of spawning areas in *Thunnus alalunga* and *T. thynnus*. Respectively, black circles and right-oriented hatching for *T. alalunga*, and black squares and left-oriented hatching for *T. thynnus*; both species spawn in Mediterranean waters.

2. Material and methods

2.1. Sample collection

Samples of muscle, fin or heart tissue from 460 albacore were collected at 8 locations (representing samples from feeding grounds and including a mixture of juveniles and adults) over the distribution of the species (Table 2.1; Fig. 2.1.). Additional tissue samples from 107 Atlantic bluefin tuna (BFT) were collected from 3 locations: Western Atlantic, Bay of Biscay and Mediterranean Sea (Table 2.1.). While the Bay of Biscay sample included a mixture of juveniles and adults from a feeding ground, the samples of Western Atlantic and Mediterranean Sea were composed of young-of-the-year (YOY) individuals incapability of trans-oceanic migration and, thus, represented reference samples for both spawning areas (Rooker *et al.* 2008). Samples were either frozen and stored at -20º C, or were preserved in 96% ethanol at 4º C. DNA was extracted from tissues using the DNeasy 96 Blood & Tissue Kit (Qiagen, Hilden, Germany) and quantified using a NanoDrop 1000 spectrophotometer (Thermo Fisher Scientific Inc., Waltham, MA) prior to storage at -20ºC for further analysis.

Table 2.1. Sampling details. Sample code, number of individuals per sample (*N*), sample location, current management stock, FAO major fishing area and geographical coordinates along with year of capture.

Sample	Abbreviation	*N*	Location	Latitude	Longitude	Year	Current stock	FAO
Albacore tuna								
1	BAL	50	Balearic Sea	40.00	1.58	2005	Mediterranean	37
2	BIS	52	Bay of Biscay	45.10	-4.35	2009	North Atlantic	27
3	IRE	57	Ireland	54.17	-12.89	2008	North Atlantic	27
4	SEA	91	South Africa	-24.25	4.42	2009	South Atlantic	47
5	IN	24	Seychelles	-7.11	54.65	2008–2009	Indian	51
6	NP	101	California	43.50	-127.00	2008	North Pacific	77
7	SWP	30	New Caledonia	-18.53	165.97	2003–2008	South Pacific	71
8	SEP	55	French Polynesia	-19.01	-152.84	2003–2008	South Pacific	71
Atlantic Bluefin tuna								
9	NEA	46	Bay of Biscay	45.10	-4.35	2009	East Atlantic	27
10	MED	46	Balearic Sea	40.58	1.21	2009	East Atlantic	37
11	NWA	15	Northwest Atlantic	36.24	74.49	2008	West Atlantic	21

2.2. SNP discovery in albacore (*Thunnus alalunga*)

Single nucleotide polymorphism (SNP) discovery via comparative sequencing of nuclear DNA fragments was performed on 35 albacore from widely separated areas. Specifically, five fish were selected from each of the six currently hypothesized stocks, Mediterranean, Indian Ocean and Northern and Southern parts of Atlantic and Pacific Oceans, except for the North Atlantic, where 10 fish were used. SNPs were mined from 54 nuclear DNA fragments (Table S2.1.) and were

amplified with primers designed with Primer3 (Rozen & Skaletsky 2000). In approach I, EPIC primers (Exon-Priming, Intron-Crossing primers; Slate *et al.* 2009) for 19 DNA fragments (average length 318 bp) were obtained from the literature (references in Table S2.1.). The primers for the remaining 35 DNA fragments were designed from the alignment of sequences from publically available databases (GenBank and Ensembl). In approach II, 17 pairs of degenerate primers were designed from several teleost sequences (average length 420 bp). In approach III, 18 pairs of primers were designed from *Thunnus* DNA sequences (average length 487 bp).

Tables S2.1. and S2.2. show fragment amplification specifications. Briefly, we used conventional polymerase chain reaction (PCR) to amplify 30 of 54 fragments and used touchdown (TD) methodology to amplify the remaining fragments. Reactions were carried out in a thermo-cycler, GeneAmpPCR System 2700, GeneAmpPCR System 9700 or Veriti 96 well Thermal Cycler (Applied Biosystems, Foster City, CA), and iCycler (Biorad Laboratories, Hercules, CA). The purified PCR products were sequenced in one direction with either the forward or the reverse PCR primer on an Applied Biosystems (ABI) 3130X capillary electrophoresis Analyzer, with ABI BigDye Terminator version 3.1 Chemistry (Applied Biosystems). Base-calling from chromatograms was performed using SeqScape v2.5 (Applied Biosystems). The BLASTN algorithm was used to verify that the target locus had been amplified.

Nucleotide differences at a site in aligned sequences were considered to be a SNP, but only when flanking sequences had high quality and the alternative nucleotide was present in at least two individuals (out of 35; see above). After filtering for SNPs matching the technical requirements of the assays, we gave priority to selecting at least one SNP per fragment for a total of 128 SNPs for genotyping the 460 albacore with TaqMan OpenArray technology (Applied Biosystems). Moreover, a SNP was chosen if it was not located near the ends of the sequence and if it was more than two bases away from any other SNP (*Custom TaqMan Genomic Assays Protocol Submission Guidelines*).

2.3. Cross-species amplifications in Atlantic bluefin tuna (*Thunnus thynnus*)

SNP discovery in Atlantic BFT was performed by cross-species amplification of the 128 SNPs selected for genotyping in albacore. The 128 SNPs were genotyped with TaqMan OpenArray technology in 107 Atlantic BFT samples. The same criteria used to validate SNPs in albacore were used to validate individual SNPs in Atlantic BFT.

2.4. Population analysis

Figure 2.2. outlines the procedures used in this study, starting from SNP discovery in albacore, to the selection of a subset of SNPs (panel) with origin assignment (including loci under selection) or demographic analysis. Genotyping call rate and minor allele frequencies (MAF) were obtained for each SNP locus using AutoCaller 1.1 (Applied Biosystems). A SNP was validated if the polymorphism remained in the genotyping results and could be reliably scored. SNPs with unclear genotypes and those with a call rate below 70% were discarded. Deviation from Hardy-Weinberg expectations (HWE) was evaluated for each locus and each sampling location (Fisher's exact test in GENEPOP 4.0; Rousset 2008). The exact test for linkage disequilibrium (LD), as implemented in GENEPOP, was used to detect disequilibria between SNPs on the same DNA fragment and between SNPs on different fragments; $P < 0.001$ was used as critical probability for LD tests. SNP loci exhibiting significant LD were phased into haplotype blocks using the Bayesian statistical method implemented in PHASE 2.1 for each sample independently (Stephens *et al*. 2001).

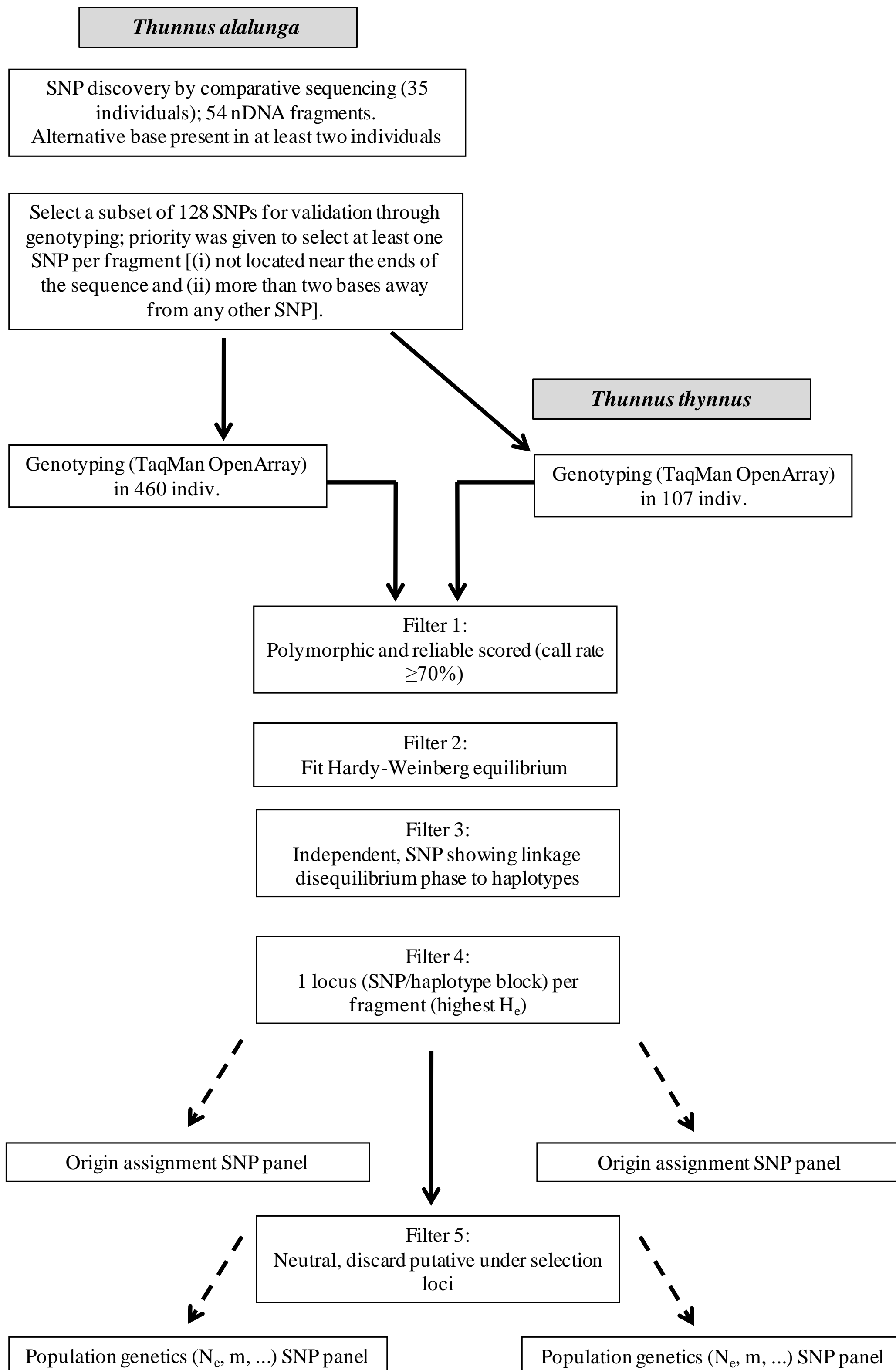


Figure 2.2: SNP selection design. Design of the filtering steps used to select SNP panels for individual assignment and for genetic population surveys in *T. alalunga* and *T. thynnus* (see Methods for further information)

Expected heterozygosity (H_e), F_{IS} and F_{ST} were estimated with FSTAT 2.9.3 (Goudet 2001). SNPs exhibiting significant departures from HWE ($P < 0.001$) in one or more samples were deemed unsuitable for estimating population structure and were discarded. When two or more loci occurred on the same DNA fragment, including both haplotype blocks and individual SNP loci, the locus with the largest H_e was selected to ensure independence among the markers. Loci with large heterozygosities provide more statistical power for population structure analysis than loci with small heterozygosities (Haasl & Payseur 2011; Morin *et al.* 2004; Rosenberg *et al.* 2003; Ryman *et al.* 2006). The SNP panel consisted of these filtered SNPs.

We searched for candidate loci under selection (outlier loci) among the remaining loci using the Bayesian likelihood method, as implemented in BAYESCAN 2.0 (Foll & Gaggiotti 2008) and LOSITAN (Beaumont & Nichols 1996; Antao *et al.* 2008). Loci identified by BAYESCAN and LOSITAN as outliers were removed from the SNP panel (Richter-Boix 2011). Briefly, two alternative models were defined in BAYESCAN (including and excluding the effect of selection) and their respective posterior odds (PO), the ratio of the posterior probabilities of the two models, were calculated using 20 pilot runs of 5000 iterations and an additional burn-in of 50 000 iterations, for a total of 100 000 iterations (sample size of 5000 and thinning interval of 10). While only loci with positive $\log_{10}$PO values were considered, strong support for selection was limited to loci with $\log_{10}\text{PO} > 1$ ($P < 0.09$; BAYESCAN 2.0 user manual). On the other hand, LOSITAN identifies outlier loci from a plot of heterozygosity versus F_{ST}. To avoid an upward bias in quantiles, LOSITAN was first run for all loci to estimate the mean neutral F_{ST}. Loci that were outside the 0.99 confidence interval were removed using only the putative neutral loci to compute again the mean neutral F_{ST}. A second run was then conducted with all loci using the last computed mean.

POWSIM 4.0 (Ryman *et al.* 2006) was used to estimate the statistical power required to detect various levels of differentiation with the SNPs developed for albacore and Atlantic BFT in this study. Burn-in consisted of 1000 steps followed by 100 batches of 1000 steps. X^2 and Fisher's probabilities were used to test the significance of an F_{ST} value for each replicate run. The number of significant F_{ST} values in 1000 replicate simulations provided an estimate of statistical power for a given level of divergence, which was controlled by allowing frequencies to drift for a given number of generations. Simulated effective populations sizes (N_e) equalled 2000 fish. For albacore, all 41 SNPs + blocks, 9 blocks only and 32 SNPs only were examined separately for statistical power for divergences ranging from $F_{ST} = 0.0005$ to 0.016.

Population structure was estimated with F_{ST} between samples and with Bayesian individual assignments implemented in STRUCTURE 2.3.3 (Pritchard *et al.* 2000). Pairwise F_{ST} (Weir & Cockerham 1984) values were estimated with FSTAT 2.9.3 (Goudet 2001) and with globally corrected *p*-values. FSTAT combines individual locus *p*-values weighting them according to their polymorphism level (Petit *et al*. 2001). Population groups were defined by non-significant values of mean F_{ST} between samples and by significant values of F_{ST} with other populations (e.g.

Waples & Gaggiotti 2006). STRUCTURE uses a Bayesian method to identify the number of clusters (K) of related individuals using HWE and gametic disequilibria among multilocus genotypes. We used the admixture model, independent allele frequencies between populations and the LOCPRIOR option. We compared log-likelihood ratios in 10 STRUCTURE runs for values of K = 1 to 10 (Pritchard *et al.* 2000). Each run consisted of 10 000 iterations with a burn-in of 10 000. CLUMPP 1.1.2 (Jakobsson & Rosenberg 2007) was used to determine optimal assignments of individual to clusters by maximizing the similarity between pairs of genotypes in different replicates. These groupings were visualized with DISTRUCT 1.1 (Rosenberg 2004). Outlier loci that were not used to estimate F_{ST}, were added for the STRUCTURE analyses as the latter does not require neutral markers, unlike the F_{ST} analysis.

3. Results

3.1. SNP discovery and validation in albacore tuna

Fifty-four fragments of nuclear DNA were sequenced in 35 albacore tunas (Table S2.1.). Thirty-five of the 54 fragments showed a high degree of homology with the orthologous sequences in several fishes (BLASTN; E-value < 10^{-5}), including 14 best hits corresponding to other species of *Thunnus* (Table S2.3.). A total of 616 SNPs were discovered, in which an alternative allele was present in at least 2 individuals, with a mean of 11.4 (SD ±10) SNPs per fragment and a ratio of 1/36 bp. At least 1 SNP was present in each DNA fragment, except for a fragment coding for metallothionein (*MT*). A total of 195 SNPs were present in fragments amplified with EPIC primers (approach I); 182 SNPs were found in fragments amplified with degenerate teleost primers (approach II); and 239 SNPs were present in DNA fragments amplified with *Thunnus* spp. primers (approach III). In addition to SNPs, 19 small indels, 1–5 nucleotides in length, were found in 14 fragments with a majority corresponding to mono- or bi-nucleotide indels (84.2%).

A total of 128 candidate SNPs were selected to genotype the albacore population samples and included 32, 47 and 49 SNPs selected from fragments obtained with approaches I, II and III, respectively. A total of 23 (18%) SNPs failed to amplify for routine genotyping, and 2 SNP loci failed to exceed call rates above 70%. Another 24 loci among the remaining 103 SNPs could not be reliably scored. The remaining 79 validated SNPs (Table S2.4) showed a mean call rate of 91 ±5% and an average minor allele frequency (MAF) of 0.17 ±0.14 (range 0.001–0.489). Validation success was 72%, 66% and 51% for approaches I, II and III, respectively.

3.2. SNP panel for population genetic studies in albacore

Twelve of the 79 validated SNPs departed significantly from HWE in one or more sampling locations and were discarded. The remaining 67 SNPs were tested for linkage disequilibrium (LD). No SNPs were found in LD between DNA fragments; however, 21 SNPs were in LD within fragments and were phased into 9 haplotype blocks (Table S2.5). After selection of only one independent locus per DNA fragment (see Methods), the final panel of markers included 41 independent markers: 32 individual SNPs, plus 9 haplotype blocks (53 SNPs in total; Table 2.2). Analysis of these SNPs with BAYESCAN showed no candidate loci influenced by selection. However, LOSITAN detected 3 SNP loci, *HIF1A4, MTF1* and *MYC*, with significantly larger genetic divergences than expected from neutrality. These loci were assumed to be embedded in candidate genes under divergent selection or in DNA fragments linked to genes influenced by selection. In contrast, the low level of divergence for *PRDX2* suggested that balancing selection may be influencing this locus, or a nearby gene. Interestingly, *MTF1* and *PRDX2* sequences showed the highest homology with orthologs of the teleosts *Fugu rubripes* and *Thunnus maccoyii*, respectively (Table S2.3), but no homologies were found for *HIF1A4* and *MYC* in the teleosts tested. Overall, the average expected heterozygosity over loci was H_e = 0.278 ±0.201 and the average inbreeding coefficient was F_{IS} = 0.032 ±0.085 for the 41 SNP loci in the final panel (Table 2.2).

Table 2.2. Selected set of SNPs for population genetic studies in *T. alalunga*. SNP code (in bold), the nuclear DNA fragment were it was discovered and the discovery approach (I, II and III; see Methods) are shown, for the panel of 53 SNPs, representing 41 loci including 32 individual SNPs and 9 haplotype blocks (shaded), along with the number of alleles or haplotypes and mean values, across samples, for both expected heterozygosity (H_e) and inbreeding coefficient (F_{IS}).

Fragment	Method	SNP	no. alleles/ haplotypes	Mean H_e (±SE)	Mean F_{IS} (±SE)
ADRB2	I	***ADRB2-97***	2	0.025±0.016	-0.006±0.006
ALDOB1	I	***ALDOB1-47*** ***ALDOB1-95***	4	0.373±0.058	0.150±0.104
CALM4	I	***CALM4-124***	2	0.432±0.040	0.247±0.165
GNRH3-1	I	***GNRH3-1-124***	2	0.365±0.040	0.030±0.047
GNRH3-3	I	***GNRH3-3-219***	2	0.077±0.035	-0.033±0.018
LDHB	I	***LDHB-287***	2	0.362±0.061	0.176±0.205
LYZ	I	***LYZ-128*** ***LYZ-138*** ***LYZ-340***	8	0.579±0.079	-0.005±0.155
MLL25a	I	***MLL25a-144*** ***MLL25a-183***	4	0.526±0.057	0.092±0.076
MYC	I	***MYC-91***	2	0.043±0.082	-0.024±0.050

MYL3	I	***MYL3-97***	2	0.024±0.025	-0.009±0.009
RHO	I	***RHO-111***	2	0.033±0.039	-0.014±0.017
RPS7-2	I	***RPS7-2-69***	2	0.270±0.081	0.180±0.138
APOE	II	***APOE-148***	2	0.269±0.039	0.014±0.116
CYCS	II	***CYCS-132*** ***CYCS-218***	4	0.468±0.062	-0.050±0.096
DAD1	II	***DAD1-444***	2	0.088±0.021	-0.039±0.011
FGB	II	***FGB-257***	2	0.053±0.036	-0.024±0.020
FOS	II	***FOS-107***	2	0.154±0.084	-0.089±0.060
HGF	II	***HGF-375***	2	0.039±0.022	-0.015±0.011
HMOX1	II	***HMOX1-416***	2	0.316±0.102	0.181±0.214
MMP9	II	***MMP9-68*** ***MMP9-111***	4	0.220±0.107	0.051±0.074
MTF1	II	***MTF1-263***	2	0.316±0.110	0.077±0.104
PSM	II	***PSM-33*** ***PSM-117*** ***PSM-138***	8	0.737±0.037	-0.045±0.102
PTGS2	II	***PTGS2-56*** ***PTGS2-317***	4	0.708±0.022	-0.017±0.154
RASA3	II	***RASA3-188***	2	0.063±0.048	-0.028±0.022
RHOC	II	***RHOC-55***	2	0.054±0.056	-0.021±0.027
RPL12	II	***RPL12-213*** ***RPL12-423***	4	0.444±0.169	0.042±0.042
c-mos	III	***c-mos-242***	2	0.313±0.059	-0.055±0.139
CS1	III	***CS1-197*** ***CS1-442*** ***CS1-512***	6	0.533±0.042	-0.030±0.124
CS3	III	***CS3-394***	2	0.449±0.039	0.006±0.220
CS5	III	***CS5-44***	2	0.494±0.013	0.209±0.161
ELOVL2	III	***ELOVL2-519***	2	0.253±0.052	0.095±0.239
ELOVL3	III	***ELOVL3-365***	2	0.129±0.081	-0.071±0.057
FGG	III	***FGG-242***	2	0.371±0.102	-0.009±0.148
GPX-458	III	***GPX-458***	2	0.495±0.009	0.107±0.115
HIF1A2-3	III	***HIF1A2-3-350***	2	0.449±0.027	0.073±0.083
HIF1A4	III	***HIF1A4-219***	2	0.145±0.122	-0.022±0.043
MB	III	***MB-188***	2	0.174±0.068	0.095±0.225
OPC02	III	***OPC02-249***	2	0.167±0.046	-0.032±0.111
PRDX2	III	***PRDX2-452***	2	0.013±0.023	-0.004±0.009
RAG2	III	***RAG2-114***	2	0.312±0.088	0.147±0.195
Tmo-4C4	III	***Tmo-4C4-188***	2	0.063±0.046	-0.026±0.027

3.3 Genetic structure in albacore tuna

POWSIM simulations indicated that the 41 independent markers (32 SNPs and 9 haplotype blocks) together were able to detect significant differences among samples with F_{ST} = 0.002 in about 84% of the tests and with $F_{ST} \geq 0.004$ in 100% of the tests (Table 2.3). The nine multi-allelic haplotype blocks alone provided about the same amount of statistical power for detecting differences among populations as did the 32 individual SNPs. Each set of markers was able to detect values of $F_{ST} \geq$ 0.004 in at least 97% of the tests. Fisher's method for detecting significant differences among samples provided less statistical power than did the X^2 tests.

Table 2.3 Probability of detecting a particular level of differentiation (F_{ST}) among populations of albacore. Power analysis conducted with POWSIM (1000 replicates) with mean frequencies and sample sizes used for albacore tuna.

	41 SNPs and blocks		**9 Blocks**		**32 SNPs**	
F_{ST}	P X^2	P **Fisher**	P X^2	P **Fisher**	P X^2	P **Fisher**
0.0005	0.197	0.186	0.125	0.127	0.156	0.150
0.001	0.439	0.404	0.276	0.278	0.286	0.289
0.002	0.886	0.844	0.662	0.638	0.628	0.625
0.004	1.0	0.999	0.981	0.970	0.980	0.976
0.008	1.0	1.0	1.0	1.0	1.0	1.0
0.016	1.0	1.0	1.0	1.0	1.0	1.0

Analyses using the 32 SNPs and nine haplotype blocks together revealed an overall F_{ST} = 0.017 ±0.003 ($P < 0.05$) among the eight albacore sampling locations. Levels of divergence were not significant between sampling locations within oceans, but were significant between oceans (Table 2.4). Samples from the NE Atlantic (IRE and BIS) were not significantly different from each other or from a sample from the SE Atlantic (SEA). Likewise, no divergence was detected between the three samples from the Pacific (NP, SEP, SWP). However, all comparisons between oceans were significant, yielding four differentiated genetic entities: 1) Atlantic Ocean, 2) Mediterranean Sea, 3) Pacific Ocean and 4) Indian Ocean. Fish from the western Mediterranean (BAL) showed the highest divergence from the other locations with an average F_{ST} = 0.034 (range: 0.021–0.050). Fish from the Indian Ocean (IN) were most divergent from the Atlantic and Mediterranean sample locations (mean: F_{ST} = 0.030), but less divergent from Pacific Ocean samples (F_{ST} = 0.010). The individual Bayesian clustering (STRUCTURE) indicated the largest likelihood of population structure was K = 3, placing samples into three groups: Mediterranean Sea, Atlantic Ocean, and Indo-Pacific (Fig. 2.3). Analysis with K = 4 showed that Indian Ocean albacore were differentiated to a small degree from Pacific Ocean albacore, as reflected in the distribution of F_{ST} values between these locations.

Table 2.4. Pairwise F_{ST} values between samples of albacore tuna (*Thunnus alalunga*). F_{ST} values appear below the diagonal and standard errors above the diagonal. Sample abbreviations as in Table 2.1, F_{ST} values significantly larger than 0.0 are in bold ($^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$)

	BAL	BIS	IRE	SEA	IN	NP	SEP	SWP
BAL		0.008	0.008	0.008	0.013	0.010	0.011	0.008
BIS	**0.026*****		0.002	0.005	0.008	0.006	0.007	0.008
IRE	**0.030*****	0.000		0.002	0.007	0.004	0.005	0.008
SEA	**0.033*****	0.004	-0.001		0.008	0.005	0.004	0.007
IN	**0.050*****	**0.020*****	**0.017*****	**0.020*****		0.004	0.005	0.006
NP	**0.033*****	**0.014*****	**0.019*****	**0.016*****	**0.008***		0.002	0.002
SEP	**0.043*****	**0.017*****	**0.016*****	**0.012*****	**0.007***	-0.000		0.003
SWP	**0.021*****	**0.014****	**0.011****	**0.012****	**0.011***	0.001	0.001	

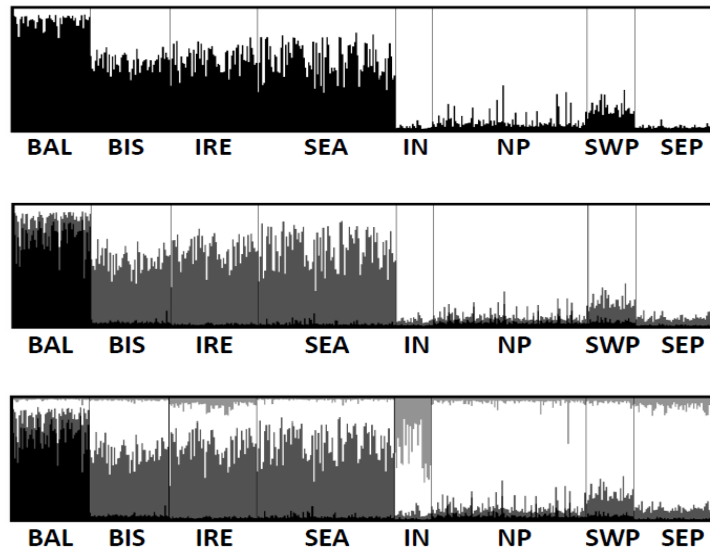


Figure 2.3. *Thunnus alalunga* STRUCTURE results. Individual clustering analysis with STRUCTURE (respectively, K = 2, K = 3 and K = 4) of 460 *T. alalunga* individuals for 53 SNPs located in 41 independent fragments. Each vertical bar represents an individual, and sampling locations are separated by vertical black lines. The colour proportions of each bar correspond to the individual's estimated membership fraction to each of the clusters (cluster membership coefficient).

3.4. SNP panel for Atlantic bluefin tuna

Primers for the 128 albacore SNPs were used in cross-species reactions to develop SNPs in BFT. Although 32 SNPs successfully amplified, 9 SNPs had low call rates (below 70%), had unclear genotypes, or were not polymorphic, and hence were discarded. This yielded 23 validated SNPs (18%) for BFT. Out of these 23, 18 had been validated also when genotyping albacore while the other five SNPs (*GNRH3-1-107*, *LDHB-129*, *CYCS-161*, *CS3-118*, and *OPC02*-45; Table S2.6) were reported as reliably scored and polymorphic only in BFT. The validation success rates in BFT for SNP that had been discovered in albacore by approaches I, II and III were 28% (9 of 32 SNPs), 15% (7 of 47 SNPs) and 14% (7 of 49 SNPs), respectively. Additionally, one SNP was discarded from the final panel due to a significant deviation from HWE in at least one sample (Table S2.6). Tests for LD between the remaining 22 BFT SNPs detected two cases of two linked SNP loci in a single fragment (Table S2.5). Linked loci were phased into haplotypes. Significant LD was not detected among SNPs on different DNA fragments. A final set of 15 independent markers, 13 individual SNPs and 2 haplotype blocks, were suitable for surveys of BFT populations (17 SNPs in total; Table 2.5).

Table 2.5. Selected set of SNPs for population genetic studies in *T. thynnus*. SNP code (in bold), the identity of the nuclear DNA fragment where it was discovered in *T. alalunga* along with the applied SNP discovery approach (see Methods) are shown for the panel of 17 SNPs, representing 15 independent loci including 13 SNPs and 2 haplotype blocks (shaded), along with the number of alleles or haplotypes and mean values, pooling all sampling locations, for both expected heterozygosity (H_e) and inbreeding coefficient (F_{IS}).

Fragment	Method	SNP	no. alleles/ haplotypes	Mean H_e (±SE)	Mean F_{IS} (±SE)
ADRB2	I	***ADRB2-97***	2	0.046±0.040	-0.016±0.014
GNRH3-1	I	***GNRH3-1-107*** ***GNRH3-1-124***	4	0.676±0.023	-0.056±0.178
LDHB	I	***LDHB-129***	2	0.468±0.068	0.057±0.268
LYZ	I	***LYZ-128***	2	0.062±0.054	0.439±0.380
RPS7-2-313	I	***RPS7-2-313***	2	0.084±0.079	0.083±0.176
TPMA	I	***TPMA-53***	2	0.355±0.100	0.000±0.351
CYCS	II	***CYCS-161***	2	0.175±0.195	0.216±0.240
HGF	II	***HGF-375***	2	0.211±0.090	0.106±0.140
MTF1-263	II	***MTF1-263***	2	0.273±0.055	0.010±0.162
RPL12	II	***RPL12-423***	2	0.486±0.012	0.006±0.157
CS3	III	***CS3-118***	2	0.291±0.253	0.023±0.070
CS5	III	***CS5-395*** ***CS5-425***	4	0.243±0.211	0.137±0.183

FGG	III	***FGG-242***	2	0.401±0.108	0.215±0.198
HIF1A2-3	III	***HIF1A2-3-417***	2	0.145±0.041	0.216±0.500
OPC02	III	***OPC02-45***	2	0.167±0.289	0.000±0.000

Average expected heterozygosity among the 15 loci was H_e = 0.272 ±0.178, and the average inbreeding coefficient was F_{IS} = 0.096 ±0.133. The low number of markers tested in BFT precluded the use of outlier detection software, as a larger number of SNPs are required to obtain a reliable estimate of the neutral expectation from which the outliers are detected. Therefore, all 15 loci were used to estimate population structure with F_{ST} and STRUCTURE.

3.5. Genetic structure in Atlantic bluefin tuna

The POWSIM simulations and the average frequencies for the 13 SNPs and 2 blocks indicated that statistical power increased from only 0.064 for F_{ST} = 0.001 (Fisher's method of determining significance), to 0.579 for F_{ST} = 0.010 and to 1.0 for F_{ST} = 0.080 (Table 2.6). The 15 BFT SNPs showed significant overall differentiation among populations in the NW Atlantic, NE Atlantic and Mediterranean (F_{ST} = 0.029 ±0.024, $P < 0.05$). Populations in the Bay of Biscay and the Mediterranean differed from the populations in the NW Atlantic (NEA–NWA, F_{ST} = 0.120 ±0.091, $P < 0.01$; MED–NWA 0.116 ±0.078, $P < 0.01$). However, the Bay of Biscay and the Mediterranean did not differ significantly from each other (F_{ST} = 0.004 ±0.007). STRUCTURE indicated that the three samples most likely represented two populations (K = 2), 1) NW Atlantic and 2) Bay of Biscay and Mediterranean (Fig. 2.4.).

Table 2.6. Probability of detecting a particular level of differentiation (F_{ST}) among populations of Atlantic bluefin tuna. Power analysis conducted with POWSIM (1000 replicates) with mean frequencies and sample sizes used for Atlantic bluefin tuna.

	13 SNPs and 2 blocks	
F_{ST}	***P* X²**	***P* Fisher**
0.001	0.091	0.064
0.0025	0.159	0.104
0.005	0.314	0.240
0.010	0.654	0.579
0.020	0.961	0.931
0.040	1.000	0.998
0.080	1.000	1.000

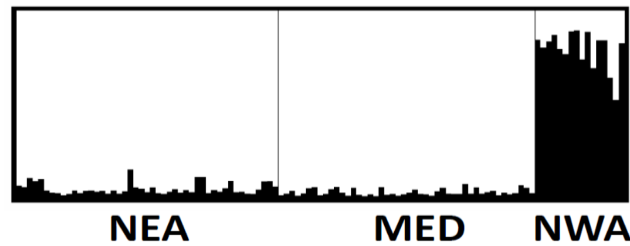


Figure 2.4. ***Thunnus thynnus*** **STRUCTURE results.** Individual clustering analysis with STRUCTURE (K = 2) analysis of 107 *T. thynnus* individuals for 17 SNPs on 15 independent DNA fragments. Each vertical bar represents an individual, and sampling locations are separated by vertical black lines. The colour proportions of each bar correspond to the individual's estimated membership fractions to each of the clusters (cluster membership coefficient).

4. Discussion

Our study outlines the development and validation of SNPs in the genomes of albacore and Atlantic bluefin tuna and provides a basis for defining discrete stocks to aid in the commercial harvests of these species. We developed *de novo* 53 SNPs for albacore and 17 cross-species SNPs for Atlantic bluefin tuna. While the focus of the study was on the development of SNP markers, the distributions of our samples allow a preliminary analysis of large-scale population structure. Several variables influence the ability of a set of molecular markers to detect genetic differences between populations. In addition to the well known effects of sample size on power, the geographical extent of a set of samples is crucial to describing population structure.

4.1. SNP discovery in Albacore tuna

Our search for variable nucleotide sites yielded 128 SNPs. Of these, 79 (62%) could be validated and were selected for routine genotyping. From these, we selected a final panel of 53 SNPs distributed over 41 loci. The 62% validation success rate is similar to SNP validation success rates in other studies of fishes, including *Gadus morhua* (54%) (Moen *et al.* 2008), *Oncorhynchus nerka* (39%), *O. keta* (54%) and *O. tshawytscha* (64%) (Smith *et al.* 2005) and *Engraulis encrasicolus* (59%) (Molecular Ecology Resources Primer Development Consortium *et al.* 2012).

Our results allow an assessment of levels of polymorphism in the albacore genome. A total of 616 SNPs were discovered with an overall ratio of 1 SNP for each 36 base pairs. This value indicates higher levels of polymorphism in albacore than has been reported, for example, in the salmonids, *Oncorhynchus keta* (1/175 bp), *O. nerka* (1/242 bp) and *O. tshawytscha* (1/301 bp) (Smith *et al.* 2005), *Salmo salar* (1/586 bp overall, Hayes *et al.* 2004) or in intronic (1/405 bp) and exonic (1/1448 bp) regions (Ryynänen & Primmer 2006). The large value for albacore, however, is similar to that for European anchovy (*Engraulis encrasicolus*) (1/54 bp, Zarraonaindia 2011).

4.2. SNP development in Atlantic bluefin tuna

We used primers for the 128 SNPs detected in albacore to search for SNPs in Atlantic BFT and achieved an overall validation success rate of 18% (23 SNPs). This success rate indicates that the regions flanking these SNPs contain highly conserved sequences so that the SNP primers developed for albacore also work in Atlantic BFT. When successful, cross-species amplifications are a cost-effective method of SNP discovery (Malhi *et al.* 2011). Successful cross-species amplifications are thought to be free of ascertainment bias, especially for SNPs embedded in conserved sequences near or within coding regions (Malhi *et al*. 2011). Cross-species amplifications have generally not been used to develop SNP markers in fish, because of generally low success rates in other vertebrates (e.g. Seeb *et al.* 2011; Miller *et al.* 2011). For example, only about 1% of the nearly 50 000 SNP loci developed for domestic sheep were polymorphic in two related ungulates (Miller *et al.* 2011). In a panel of a similarly large number of SNPs designed for cattle, only about 2.5 and 3% of the cross-species amplifications were successful in two lines of European bison and two species of antelopes, respectively (Kaminski *et al.* 2012; Ogden *et al.* 2012). However, the species used in these cross-amplification attempts were distantly related to one another, unlike albacore and Atlantic BFT, which are phylogenetically closely related (Chow & Kishino 1995; Chow *et al*. 2006) and are known to hybridize (Viñas & Tudela 2009). Our results are similar to those of Mahli *et al.* (2011), who reported up to 30% cross amplification successes between related species of Old World monkeys.

4.3. Stock structure

The statistical power of the SNPs is influenced by several factors. First, the power of the markers to detect population structure increases substantially when locus heterozygosities are larger than $H_e \geq 0.2$ (Haasl & Payseur 2011). In our study, about 60% of the SNP loci in both species showed $H_e > 0.2$. Second, while individual SNPs show less power than multi-allelic microsatellite loci (Ryman *et al*. 2006; Haasl & Payseur 2011), combining physically linked SNPs into haplotype blocks increases statistical power (Gattepaille & Jakobsson 2012). For example, Mesnick *et al.* (2011) retrieved the same population structure for sperm whales (*Physeter macrocephalus*) with 6 microsatellites, or with 36 SNPs representing 24 independent markers, including 8 haplotype blocks. In our study, POWSIM simulations showed that 9 haplotype blocks of linked SNPs in albacore had about the same statistical power as 32 individual SNPs. In any case, haplotype blocks and SNPs together yielded a type II error rate (failure to detect a real difference) for both species of 0% for divergences of $F_{ST} = 0.004$ or greater.

4.4. Stock structure in Atlantic bluefin tuna

The International Commission for the Conservation of Atlantic Tuna (ICCAT) currently manages Atlantic BFT as two stock units that are divided by the mid-ocean longitude at 45°W (Fromentin & Powers 2005). This geographically defined stock concept is supported by two spatially separated spawning areas, one in the Mediterranean and the other in the Gulf of Mexico (Rooker *et al.* 2007). However, tagging and microchemical studies suggest a more complex stock distribution pattern (Rooker *et al.* 2008; Walli *et al.* 2009; Galuardi *et al.* 2010). Foraging aggregations in the NE Atlantic and elsewhere may potentially originate from both main spawning areas, because trans-Atlantic migrations in both directions have been documented (Magnuson *et al*. 1994; Mather *et al*. 1995; Lutcavage *et al*. 1999; Block *et al*. 2001; Block *et al*. 2005; Rooker *et al*. 2006). However, homing to natal spawning areas (Boustany *et al*. 2008; Block *et al*. 2005; Teo *et al*. 2007; Dickhut *et al*. 2009) appear to isolate the two major groups. Recent studies with mtDNA and microsatellite markers show that populations in the Mediterranean Sea may also be structured into partially isolated subpopulations (Riccioni *et al*. 2010; Viñas *et al*. 2011).

The Atlantic BFT SNP panel developed in the present study, although limited in number due to the SNP discovery approach followed, is the beginning of a valuable tool to improve the management of this overexploited species. The advent of Next Generation Sequencing (NGS) technologies such as Roche´s 454 or Illumina´s HiSeq platforms (see Garvin *et al*. 2010 and Seeb *et al*. 2011 for a review) is making possible to discover hundreds-thousands of SNP type markers in non-model organisms and, further application of these techniques on BFT (and also in albacore) will improve the relevance of the present reported tool. The preliminary analysis in the present study confirms the genetic distinction between the two major spawning areas on the western and eastern margins of the Atlantic (both F_{ST} and STRUCTURE results) and is largely consistent with the ICCAT management plan. While trans oceanic migrations of adults have been documented with tags, finer-scale population structure may also exist. Our SNP analysis showed that the foraging area sample from the Bay of Biscay clustered with the Mediterranean samples indicating the Mediterranean origins of these fish. To our knowledge, this is the first time a mixed age classes of Atlantic BFT in the NE Atlantic have been assigned to a spawning area based on DNA data. Previous insights have come from relatively expensive and laborious tagging experiments (Rooker *et al.* 2007) and from an allozyme study comparing samples from Azores (Mid-Atlantic) and the Mediterranean (Pujolar *et al.* 2003).

The larger levels of differentiation than expected under neutrality may indicate that some SNP loci are embedded in genes under selection. Although the small number of SNPs for Atlantic BFT precluded tests of neutrality, the remarkably large value of $F_{ST} = 0.116$ indicating a high level of differentiation between populations of Atlantic BFT may result from directional selection on some SNPs. Markers influenced by directional selection often show higher F_{ST} values among populations of fishes than do neutral markers (e.g. André *et al*. 2011; Ackerman *et al*. 2011;

Poulsen *et al*. 2011). Additional SNPs for Atlantic BFT are needed to test hypotheses of selection, to estimate population structure and to identify mixed-stock components in fishery areas.

4.5 Stock structure in Albacore tuna

ICCAT, the Indian Ocean Tuna Commission (IOTC), the Western and Central Pacific Fisheries Commission (WCPFC) and the Inter-American Tropical Tuna Commission (IATTC) manage albacore with a six-stock model, which includes 1) Mediterranean Sea, 2) North Atlantic, 3) South Atlantic, 4) Indian, 5) North Pacific and 6) South Pacific. These stocks are based on a limited understanding of spawning areas, the geographical distribution of fisheries, life-history variables and the results of tagging studies (Arrizabalaga *et al*. 2004).

Our analyses resolved the genetic relationships among oceanic populations. Pairwise F_{ST} values distinguish four albacore groups, 1) Mediterranean Sea, 2) Atlantic Ocean, 3) Pacific Ocean and 4) Indian Ocean, with no within-ocean heterogeneity (Table1, Fig. 2.1.). The tests for heterogeneity did not detect differences between northern and southern populations within the Pacific or within the Atlantic. The Mediterranean group appears to be most differentiated from other global populations, and this is in agreement with previous results for microsatellites (Davies *et al.* 2011). Both SNPs and microsatellites show that Mediterranean and North Atlantic populations are partially isolated from each other, and this genetic separation is consistent with tag-recapture analysis showing limited movement between the North Atlantic and Mediterranean (Arrizabalaga *et al.* 2004). The SNP data also showed that the Indian Ocean populations were genetically closer to Pacific populations than to Atlantic populations. In contrast, a closer relationship between Indian Ocean and Atlantic fish was reported from blood-group frequencies (Arrizabalaga *et al.* 2004).

Overall, the genetic results together support the existence of at least 4 genetic entities, Mediterranean Sea, Atlantic, Pacific and Indian oceans that are isolated from one another to some degree. These results support management plans in which oceanic populations, including the Mediterranean, are managed separately, but differ from the present use of six management units (stocks). However, due to our relatively limited sampling of locations and individuals within ocean basins, the current six-stock management model represents a conservative approach that reduces the risk of inadvertently overfishing some populations.

5. Conclusions

We developed 128 SNP markers *de novo* in albacore tuna and used a final panel 53 SNPs (41 SNPs, including 32 individual SNPs and 9 haplotype blocks) to genotyped over 400 individuals collected over the distributional range of the species. Although the coarse scale of sampling limits our inferences about population structure, the results for albacore are largely consistent with

previous molecular studies in indicating the existence of at least four albacore populations: Mediterranean Sea, Atlantic, Pacific and Indian oceans. We then used SNP assays developed for albacore to develop 17 validated SNPs in Atlantic bluefin tuna (15 SNPs, including 13 individual SNPs and 2 haplotype blocks). This small number of SNPs discriminated the two major spawning areas of Atlantic BFT in the Gulf of Mexico and Mediterranean Sea, and identified the Mediterranean origin of juveniles foraging in the Bay of Biscay. The additional development of new SNPs will increase the statistical power needed to resolve the population structures of these two overfished tunas.

SNPs have two advantages over other markers, such as allozymes and microsatellites, for the international management of far ranging tunas. First, a large number of SNPs, individually or in haplotype blocks, provides a large amount of statistical power to detect biologically meaningful genetic differences between stocks of highly mobile tunas. High mobility is expected to produce only small differences between stocks. Second, and most importantly, SNP assays can be used in any laboratory without the need for standardizing allelic variants among research or management groups. International cooperation in establishing a universal database is essential for the conservation and management of these vulnerable species with stocks that straddle international boundaries. The SNPs developed for both species can provide valuable tools for population management and can additionally be used as markers to trace fishery products.

Acknowledgements

The authors thank I. Miguel and F. Rendo (Sequencing and Genotyping Service of the UPV/EHU, SGIKer) for their technical assistance as well as C. Manzano (UPV/EHU), M. A. Pardo (AZTI-Tecnalia), I. Cancio (UPV/EHU) and G. Heinisch (University of New Hampshire) for their collaboration. We are grateful to J. Areso, D. Fuller, K. Schaefer, V. Allain, S.Y. Yeh, I. Fraile and I. Arregi for providing samples. This work was supported by ATM2010Hegaluze (351BI20090047) ATM2009Hegalabur (351BI20090034) and ATM2008Bonorte (ACM2008BONORTE) projects funded by the Basque Government, and the ACEITUNA (CTM2011-27505) project funded by the Spanish *Ministerio de Economía y Competitividad*. Urtzi Laconcha´s work was supported by a PhD grant by the *Fundación Centros Tecnologicos Iñaki Goenaga*. Sequencing was performed at the General Research Services Unit, SGIKer- University of the Basque Country (UPV/EHU).

Supplementary material

Table S2.1. Information on the 54 nuclear DNA fragments selected for SNP discovery in *Thunnus alalunga* (see Methods): fragment name, gene region, SNP discovery approach (1, 2 and 3; see Methods), source, forward and reverse primer sequences, PCR annealing temperatures, sequenced fragment lengths and recorded number of SNPs and indels. EPIC primers (Exon-Priming, Intron-Crossing primers; SNP discovery approach I) were obtained from the literature while primers for the remaining DNA fragments were designed from the alignment of sequences from publically available databases (GenBank and Ensembl), respectively, 17 pairs of degenerate primers from several teleost species sequences (approach II) and 18 pairs of primers from genus *Thunnus* DNA sequences (approach III).

Fragment name	Gene region	Method	Source	Forward primer (5'->3')	Reverse primer (5'->3')	Touchdown temperature	Fixed temperature	Fragment lenght	SNP #	Indels
ADRB2	*adrenergic receptor beta 2* Intron	1	Lyons *et al.* 1997	F: TTTAACTTCCAGAGCCTGCTGACA	R: TCAGCCAGCCATTTACCAACA	52ºC	40ºC	146	2	-
ALDOB1	*aldolase B* Intron 1	1	Hassan *et al.* 2002	F: GCTCCAGGAAAGGGAATCCTGGC	R: CTCGTGGAAGAAGATGATCCCGCC	72ºC	62ºC	216	5	-
ALDOB4	*aldolase B* Intron 4	1	Hassan et al. 2002	F: GCCAGATATGCCAGCATCTGCC	R: GGGTTCCATCAGGCAGGATCTCTGGC	62ºC	50ºC	123	1	-
ALDOC1	*aldolase C* Intron 1	1	Hassan et al. 2002	F: CCTGGCTGCGGACGAGTCTGTGGG	R: GGCGGTACTGTCTGCGGTTCTCC	80ºC	70ºC	187	11	-
AMY2B3	*amylase alpha 2 B* Intron 3	1	Hassan et al. 2002	F: TGGAACCGAAACATTGTGAAC	R: CCCATCCAGTCATTCTGATCC	60ºC	48ºC	113	6	-
CALM4	*calmodulin* Intron 4	1	Chow 1998	F: CTGACCATGATGGCCAGAAA	R: GTTAGCTTCTCCCCCAGGTT	62ºC	52ºC	423	8	-
G6PD4	*glucose-6-phosphate dehydrogenase* Intron 4	1	Chow & Nakadate 2004	F: GAGCAGACGTATTTTGTGGG	R: GCCAGGTAGAAGAGGCGGTT	-	62ºC	308	7	1 (1bp)
GNRH3-1	*gonadotropin-releasing hormone 3* Intron 1	1	Hassan et al. 2002	F: AATGCACCACATGCTAACAAGGC	R: CGCACCATCACTCTGCTGTTCGC	62ºC	52ºC	197	3	-
GNRH3-3	*gonadotropin-releasing hormone 3* Intron 3	1	Hassan et al. 2002	F: GCCCAAACCCAAGAGAGACTTAGACC	R: TTCGGTTGAAATGACTGGAATCATC	58ºC	46ºC	327	2	-
LDHB	*lactate dehydrogenase B* Intron	1	Friesen et al. 1999	F: TCAAACGTAAAGGGGAGATGATGGA	R: TTCCTCTGGACCAGGTTGACCCTGCTCTC	-	62ºC	368	20	-
MLL25a	*mixed lineage leukemia* Intron 25a	1	Venkatesh *et al.* 1999	F: GCNCGNTCNAAYATGTTYTTYGG	R: ATRTTNCCRCARTCRTCRCTRTT	58ºC	46ºC	552	14	-
LYZ	*lysozyme* Intron 2	1	Mitra *et al.* 2003	F: CCCGGGGGCTAAGAACG	R: ATGCCTGAAAATATGAAGAGTGG	58ºC	48ºC	591	20	-
MT	*metallothionein*	1	Rolland *et al.* 2007	F: ATGACCCTTGCGAATGCTC	R: GCAGGAGCCTCCGCAGTTGC	58ºC	46ºC	112	0	-
MYC	*C-Myc* 3'UTR	1	Panno & McKeown 1995	F: CCGGAGGTGGCTAACAATG	R: TGCCTGAACTTCCTGACAAT	58ºC	46ºC	168	11	2 (2bp&1bp)
MYL3	*myosin light chain* Intron 3	1	Touriya *et al.* 2003	F: AGTAATGACGTCGCAGATGTTCT	R: CGACAGGTTCACTCTCGAGGAG	55ºC	42ºC	135	2	-
RHO	*rhodopsin*	1	Venkatesh et al. 1999	F: CCNTAYGAYTAYCCNCARTAYTA	R: TTNCCRCARCAYAANGTNGT	58ºC	46ºC	525	2	-
RPS7-1	*S7 ribosomal protein* Intron 1	1	Chow & Hazama 1998	F: TGGCCTCTTCCTTGGCCGTC	R: AACTCGTCTGGCTTTTCGCC	58ºC	46ºC	613	42	-
RPS7-2	*S7 ribosomal protein* Intron 2	1	Chow & Hazama 1998	F: AGCGCCAAAATAGTGAAGCC	R: GCCTTCAGGTCAGAGTTCAT	-	60ºC	605	34	-

TPMA	*alpha tropomyosin* Intron	1	Friesen *et al.* 1999	F: GAGTTGGATCGCGCTCAGGAGCG	R: TCAGCCTCCTCAGCGATGTGCTT	64ºC	52ºC	341	5	-
A2M	*alpha-2-macroglobulin*	2	Present work designed primers	F: CRGGGAACACTTGGYTRACWGC	R: AGAGCCTGAAGAGCCACCACYGTGTCCTG	-	62ºC	128	6	2 (5bp&1bp)
APOE	*apolipoprotein E*	2	Present work designed primers	F: AAGCTGAAGAARCGCCTBAACAAGGAC	R: AGCTCRTCSATCTTGCCYTCCAGGGAGGT	-	66ºC	326	3	-
CYCS	*cytochrome c*, somatic, nuclear gene encoding mitochondrial protein	2	Present work designed primers	F: TTYGTCCAGAAGTGTGCCCAGTG	R: TTCTTYTTGATGCCRGCGAAGATCAT	-	58ºC	264	9	-
DAD1	*defender against cell death 1*	2	Present work designed primers	F: AARGTGGTGGACGCNTATYTGCTGTACATC	R: GGACGGTGTGAGCGAAYAGRAAGTC	-	64ºC	549	21	-
F2	*prothrombin (coagulation factor II)*	2	Present work designed primers	F: CCATGGCAGGTGATGYTSTACAA	R: GTCMCGGTTCAGRTTYTCCTTCCAG	-	56ºC	268	11	-
FGB	*fibrinogen beta*	2	Present work designed primers	F: GGAGGNTGGSTBCTCATCCAGA	R: TCCAGTCCTSCATCTCVATG	-	58ºC	312	6	-
FOS	*fos*	2	Present work designed primers	F: ARCCCATCTGCAARATCCC	R: AGBGAYTGGTCGTTGCTGBTGCT	-	60ºC	442	9	-
GPX	*glutathione peroxidase*	2	Present work designed primers	F: CCCTGCAAYCAGTTYGGMCATCAGGAGAAC	R: TGGTVAGGAAMYTYCTGCTGTAVCGCTT	-	60ºC	572	6	1 (1bp)
HGF	*hepatocyte growth factor like*	2	Present work designed primers	F: GARTCYCAVCTGGTCATGCTDCA	R: GTGCACATCTCATTCTCRCG	-	64ºC	554	4	-
HMOX1	*heme-oxygenase 1*	2	Present work designed primers	F: CAGGGAYYTGTCNGARCARATCAA	R: GACAGRTCNCCBAGGTAVCGGGTGTAAGC	-	56ºC	499	11	-
MMP9	*matrix metallopeptidase 9*	2	Present work designed primers	F: TTTGADGGAGACCTCAAATGGGA	R: TGTACATDGGGTACATGAGHGCMTCTC	-	58ºC	521	5	2 (1bp)
MTF1	*metal-regulatory transcription factor 1*	2	Present work designed primers	F: TACAGCACRGCRGGAAACCTGCG	R: GTACTTKGTGCATCCCTCKGATTCACAG	-	66ºC	636	10	-
PSM	*proteasome subunit N3*	2	Present work designed primers	F: AACATCTCYCGCCTCATGAA	R: GCTGCGGCGGTTGTACATKAC	-	60ºC	389	8	1 (1bp)
PTGS2	*prostaglandin-endoperoxide synthase 2 (prostaglandin G/H synthase and cyclooxygenase)*	2	Present work designed primers	F: CACCAGTTCTTCAARTCHGA	R: TGATGTAYGCYACYATYTGGCT	-	56ºC	626	23	-
RASA3	*RAS protein activator 3*	2	Present work designed primers	F: CTGYTGGACATYCTGGACACKGCAGG	R: AGDCCCTGTCTGGTTTTGGCWGA	-	62ºC	265	5	-
RHOC	*ras homolog family member C*	2	Present work designed primers	F: AAGCTGGTGATHGTBGGVGAT	R: GAGAAGCACATGAGGATGACRTC	-	62ºC	255	15	-
RPL12	*ribosomal protein L12*	2	Present work designed primers	F: GACTGGAAGGGYCTGAGGATCAC	R: CTGTYGATGTCMTCGATGAYRTC	-	56ºC	541	30	1 (4bp)
c-mos	*oocyte maturation factor (c-mos) gene*	3	DQ874861.1	F: AACATCGTTCGCGTCATC	R: CAGACACCCCTTCTCCTTTC	60ºC	48ºC	340	6	-
CS1	*citrate synthase* Intron1	3	AY461849.1	F: ATGGCAAAACCAACATAGGAC	R: CCTTCATCCCCCTCATACC	-	60ºC	653	11	-
CS2	*citrate synthase* Intron2	3	AY461849.1	F: AGGTATGAGGGGGATGAAGG	R: GTGACCAGCAGCCAGAAGAG	-	58ºC	326	24	-
CS3	*citrate synthase* Intron3	3	AY461849.1	F: ATCCGTTTCCGGGGCTAC	R: CAAAGCTGCTCTCGCTGTT	-	58ºC	645	32	2 (3bp&1bp)
CS5	*citrate synthase* Intron5	3	AY461849.1	F: CATGGACTTGATTGCCAAG	R: CGCTGAAGGACAGATAGGG	-	60ºC	551	27	1 (1bp)
ELOVL2	*ELOVL fatty acid elongase* Intron 2	3	FJ156735.1	F: TCTACTGCCAGGACACTCACA	R: CGAACCACCAGATATTCAGCA	62ºC	52ºC	621	7	-
ELOVL3	*ELOVL fatty acid elongase* Intron 3	3	FJ156735.1	F: CATTTCTTCACATCTACCACCAC	R: CCTATTTGGAAGTACAGCCATCC	64ºC	54ºC	611	9	1 (1bp)
FGG	*fibrinogen gamma chain* Intron	3	CA352559, BX510913.9	F: AGACGGACTATCTGTTGGATTTG	R: TCCTTTCCTTTTTCACCATTTG	-	50ºC	307	4	-
GPX	*glutathione peroxidase*	3	IEF452498.3	F: GGCAAAACCCCAGTAAACTAC	R: GGAAGATCCTTCTCCACCAC	64ºC	52ºC	637	23	1 (1bp)

HBB	*hemoglobin B Chain*	3	AB093568.1	F: AGCATCATCGCTGGCATC	R: TTTCATTGAGGAGACAAACCAC	60ºC	50ºC	478	4	-
HIF1A2-3	*hypoxia inducible factor 1, alpha* subunit Intron 2 & 3	3	EU300942.1	F: GCGTGGAAAGGAGTCGGAG	R: GGTCGCAGGGATGTATGAAG	-	62ºC	573	21	2 (2bp&1bp)
HIF1A4	*hypoxia inducible factor 1, alpha* subunit Intron 4	3	EU300942.1	F: GGAGATGCTGGTCCACAAAAC	R: TTCATTCGCAGGAAGAAGCTG	-	62ºC	627	5	-
MB	*myoglobin*	3	AF291832, AF291838, AB104433.1, AF291836	F: ATTGGAGGCCTGGTTCTGAC	R: TCTCTGATATACACCGATCTCACC	-	62ºC	455	8	-
Mhc1	*MHC class II* Intron 1	3	AF134968	F: ACTCCCGCCTGGAGTACAcGC	R: CAGGAGATCTTCTCTCCAGAC	58ºC	46ºC	172	1	-
OPC02	RAPD OPC-02 Marker	3	AF243431	F: CCCTGGAACACAGAAAATG	R: GCAGAAACTTGAGCAGGAAG	-	54ºC	306	12	-
PRDX2	*peroxiredoxin 2* Intron	3	EU093980.1	F: TCTCCAGAGACTACGGTGTACTG	R: GACGTCAGGGATGATGGTG	62ºC	48ºC	563	33	1 (1bp)
RAG2	*recombination activating gene 2*	3	DQ874771.1	F: CGCTGCAAAGAGAAAGAACTG	R: GTCCACCTGACAACCAAGG	56ºC	44ºC	640	11	1 (1bp)
Tmo-4C4	*titin (TTN)* like protein	3	DQ388108.1	F: GTTTTACGCCGAGGTGTTTG	R: GAGGGGATGGAGAACGAGAG	-	60ºC	253	1	-

-Chow S. (1998) Universal PCR primer for calmodulin gene intron in fish. Fisheries Science 64, 999-1000.
-Chow S. & Hazama K. (1998) Universal primers for S7 ribosomal protein gene introns in fish. Molecular Ecology 7, 1255-6.
-Chow S. & Nakadate M. (2004) PCR primers for fish G6PD gene intron and characterization of intron length variation in the albacore Thunnus alalunga. Molecular Ecology Notes 4, 391-3.
-Friesen V.L., Congdon B.C., Kidd M.G. & Birt T.P. (1999) Polymerase chain reaction (PCR) primers for the amplification of five nuclear introns in vertebrates. Molecular Ecology 8, 2147-9.
-Hassan M., Lemaire C., Fauvelot C. & Bonhomme F. (2002) Seventeen new exon-primed intron-crossing polymerase chain reaction amplifiable introns in fish. Molecular Ecology Notes 2, 334-40.
-Lyons L.A., Laughlin T.F., Copeland N.G., Jenkins N.A., Womack J.E. & O'Brien S.J. (1997) Comparative anchor tagged sequences (CATS) for integrative mapping of mammalian genomes. Nature Genetics 15, 47-56.
-Mitra A., Foster-Frey J., Rexroad C.E. III, Wells K.D. & Wall R.J. (2003) Molecular characterization of lysozyme type II gene in rainbow trout (Oncorhynchus mykiss): evidence of gene duplication. Animal Biotechnology 14, 7-12.
-Panno J.P. & McKeown B.A. (1995) Cloning and expression of a myc family member from the pituitary gland of the Rainbow trout, Oncorhynchus mykiss. Biochimica et Biophysica Acta 1264, 7-11.
-Rolland J.-L., Bonhomme F., Lagardere F., Hassan M. & Guinand B. (2007) Population structure of the common sole (Solea solea) in the Northeastern Atlantic and the Mediterranean Sea: revisiting the divide with EPIC markers. Marine Biology 151, 327-41.
-Touriya A., Rami M., Cattaneo-Berrebi G., Ibanez C., Augros S., Boissin E., Dakkak A. & Berrebi P. (2003) New primers for EPIC amplification of intron sequences for fish and other vertebrate population genetic studies. BioTechniques 35, 676-82.
-Venkatesh B., Ning Y. & Brenner S. (1999) Late changes in spliceosomal introns define clades in vertebrate evolution. Proceedings of the National Academy of Sciences of the United States of America 96, 10267-71.

Table S2.2. Information for the polymerase chain reactions (PCR) and cycling conditions used to amplify the 54 nuclear DNA fragments selected for SNP discovery in *T. alalunga*. While 30 fragments were amplified with a conventional PCR, touchdown (TD) methodology, involving a decreasing annealing Tº, was applied for the remaining fragments. (1) Annealing temperature started in the first cycle with that registered in Table S1 ("Touchdown temperature" column) for each fragment. Then, annealing temperature decreased 0.5ºC in each subsequent cycle. (2) Fixed annealing temperature for each fragment. See "Fixed temperature" column in Table S1.

PCR reaction		**Touchdown PCR**	**Conventional PCR**
5X GoTaq Flexi Buffer (Promega) (µl)		5.0	-
10X AmpliTaq Gold Buffer II (Applied Biosystems) (µl)		-	2.5
$MgCl_2$ (Promega) [mM]		2.0	-
$MgCl_2$ (Applied Biosystems) [mM]		-	2.5
GeneAmp dNTP Mix (Applied Biosystems) [mM]		0.8	0.8
Primers (Thermo Fisher Scientific) [µM each]		0.7	0.7
Albumin (Roche Diagnostics) (mg/ml)		0.32	-
GoTaq DNAPolymerase (Promega) (U)		1.5	-
AmpliTaq Gold DNA Polymerase (Applied Biosystems) (U)		-	1.5
DNA (ng)		40-80	40-80
Final Volume (µl)		25	25
Cycling conditions			
Initial denaturation		95°C - 10 min	95°C-10'
Cycles		20	40
	Denaturation	95°C - 30 s	95°C - 30 s
	Anneling	decreasing T(1) - 45 s	fixed T(2) - 45 s
	Extension	72°C - 30 s	72°C - 2 min
Cycles		20	
	Denaturation	95°C - 30 s	
	Anneling	fixed T(2) - 30 s	
	Extension	72°C - 30 s	
Final Extension		72°C - 10 min	72°C - 10 min

Table S2.3. BLASTN homology search results for the *T. alalunga* sequenced fragments. Only *best hits* showing identity ≥ 70% and E-value < 10e^{-5} were considered. Out of 54 fragments of nuclear DNA that were sequenced in *T. alalunga*, 35 fulfilled the required conditions.

Fragment	Method	Score	Identity	E Value	Best alignment
ALDOB1	1	220	94%	2 E-54	emb|FN257466.1| Thunnus albacares mRNA for aldolase A (aldA gene), isolated from muscle
ALDOB4	1	89.8	98%	3 E-15	gb|FJ826521.1| Perca flavescens aldolase B mRNA, complete cds
CALM4	1	309	90%	4 E-81	dbj|AB291547.1| Thunnus thynnus CAM gene for calmodulin, partial cds
G6PD4	1	342	94%	4,00E-91	dbj|AB159515.1| Thunnus alalunga G6PDH gene for glucose-6-phosphate 1-dehydrogenase, partial cds, allele:A
GNRH3-1	1	149	76%	6,00E-33	gb|EU180153.1| Sciaenops ocellatus growth hormone 3 gene, partial sequence
GNRH3-3	1	352	81%	5,00E-94	gb|EU180149.1| Cynoscion arenarius growth hormone 3 gene, partial sequence
LDHB	1	102	84%	2,00E-18	emb|BX649476.3| Zebrafish DNA sequence from clone DKEY-13I19 in linkage group 4
RPS7-1	1	159	70%	9,00E-36	gb|AY517759.1| Ambloplites cavifrons isolate AcavA S7 ribosomal protein gene, intron 1
RPS7-2	1	538	87%	7,00E-150	dbj|AB291549.1| Thunnus thynnus S7RP gene for S7 ribosomal protein, partial cds
TPMA	1	283	93%	2 E-73	dbj|AB300376.1| Thunnus thynnus TROP gene for tropomyosin, partial cds
A2M	2	73.4	91%	2,00E-10	gb|AY358020.1| Sparus aurata alpha-2-macroglobulin mRNA, partial cds
APOE	2	287	87%	3 E-74	gb|EU812517.1| Oplegnathus fasciatus apolipoprotein E (apoE) mRNA, complete cds
CYCS	2	117	91%	3 E-23	gb|BT083124.1| Anoplopoma fimbria clone afim-evh-521-351 Cytochrome c-b putative mRNA, complete cds
DAD1	2	96.9	84%	1 E-16	gb|BT083362.1| Anoplopoma fimbria clone afim-evh-529-023 Dolichyl-diphosphooligosaccharide-protein glycosyltransferase subunit DAD1 putative mRNA, complete cds
F2	2	100	82%	4,00E-18	ref|NM_213390.1| Danio rerio coagulation factor II (thrombin) (f2), mRNA
FGB	2	96.9	88%	6,00E-17	gb|EF581895.1| Paralichthys olivaceus fibrinogen beta chain precursor, mRNA, complete cds
FOS	2	462	87%	6,00E-127	gb|DQ838581.1| Dicentrarchus labrax proto-oncogene protein c-Fos (c-Fos) mRNA, complete cds
GPX	2	396	95%	4,00E-107	gb|EF452497.1| Thunnus maccoyii glutathione peroxidase mRNA, complete cds
HGF	2	154	91%	5,00E-34	gb|GU117701.1| Paralichthys olivaceus macrophage stimulating-1 protein mRNA, complete cds
MMP9	2	210	85%	7,00E-51	gb|BT026665.1| Gasterosteus aculeatus clone CEC36-D03 mRNA sequence
MTF1	2	147	83%	8,00E-32	emb|AJ131394.1| Fugu rubripes MTF-1 gene and partial INPP5P gene
PSM	2	59	94%	9,00E-06	gb|BT028445.1| Gasterosteus aculeatus clone CNB218-A02 mRNA sequence

PTGS2	2	60.8	70%	8 E-06	emb|AM232702.1| Dicentrarchus labrax amh gene for anti-Mullerian hormone, promoter region
RASA3	2	134	93%	3,00E-28	gb|FJ529376.1| Dicentrarchus labrax Ki-ras-1 mRNA, complete cds
RHOC	2	71.6	88%	2,00E-09	gb|EU411810.1| TSA: Hippoglossus hippoglossus all_halibut.762.C1 mRNA sequence
RPL12	2	135	97%	2,00E-28	gb|AY190964.1| Pagrus major ribosomal protein L12 mRNA, partial cds
c-mos	3	529	93%	4 E-147	gb|DQ874861.1| Thunnus albacares isolate JJ7:18 oocyte maturation factor (c-mos) gene, partial cds
CS3	3	102	93%	2 E-18	gb|AY461849.1| Thunnus obesus mitochondrial citrate synthase precursor (CS) mRNA, complete cds; nuclear gene for mitochondrial product
GPX	3	116	100%	1,00E-22	gb|EF452498.3| Thunnus maccoyii phospholipid hydroperoxide glutathione peroxidase mRNA, complete cds
HBB	3	378	96%	2,00E-101	dbj|AB093568.1| Thunnus thynnus mRNA for hemoglobin beta chain, complete cds
HIF1A2-3	3	273	87%	6,00E-70	gb|EF139130.1| Dicentrarchus labrax heme oxygenase 1 mRNA, complete cds
MB	3	407	99%	3,00E-110	gb|AF291832.1|AF291832 Thunnus alalunga myoglobin mRNA, complete cds
Mhc1	3	221	87%	9,00E-55	gb|AF134968.1|AF134968 Thunnus sp. MHC class II antigen gene, partial cds
PRDX2	3	226	95%	9,00E-56	gb|EU093980.1| Thunnus maccoyii 2-Cys peroxiredoxin (PRDX) mRNA, complete cds
Tmo-4C4	3	449	98%	3,00E-123	gb|DQ874835.1| Thunnus albacares isolate JJ7:18 Tmo4c4 (Tmo4c4) gene, partial cds

Table S2.4. Characteristics of the 79 validated SNPs, meaning reliably scored and polymorphic, in *Thunnus alalunga*. Marker name, SNP discovery approach (1, 2 and 3; see Methods), call rate (%), alleles and minor allele frequency (MAF) values are shown along with the global and per location deviation (*P*) from the Hardy-Weinberg equilibrium (HWE). "SIGN." means that at least 1 location was not in HWE for those markers (N.S. corresponds to not significant); as multiple comparisons were involved, a probability of $P < 0.001$ was considered significant. Last column shows NCBI Assay ID (rs#) for every individual SNP (NCBI´s dbSNP database).

SNP	Method	Call rate	Alleles	MAF	Global HWE	HWE per sample	NCBI Assay ID (rs#)
ADRB2-97	1	97,8	C:T	0,014	1.0000±0.0000	N.S.	rs193631113
ALDOB1-47	1	94,9	C:T	0,193	0.0004±0.0000	N.S.	rs193631114
ALDOB1-95	1	92,1	G:A	0,136	0.0577±0.0004	N.S.	rs193631115
ALDOB1-125	1	91,0	C:T	0,027	1.0000±0.0000	N.S.	rs193631116
CALM4-124	1	85,9	C:A	0,311	<0.0000±0.0000	N.S.	rs193631119
GNRH3-1-124	1	88,3	G:A	0,254	0.8947±0.0003	N.S.	rs193631144
GNRH3-3-219	1	98,5	T:C	0,034	1.0000±0.0000	N.S.	rs193631145
LDHB-96	1	84,5	A:C	0,101	<0.0000±0.0000	N.S.	rs193631153
LDHB-287	1	87,9	G:C	0,222	0.0036±0.0001	N.S.	rs193631154
LYZ-128	1	80,3	T:G	0,332	<0.0000±0.0000	N.S.	rs193631155
LYZ-138	1	89,8	C:T	0,288	0.5314±0.0009	N.S.	rs193631156
LYZ-320	1	81,9	A:G	0,160	0.2323±0.0007	SIGN.	rs193631157
LYZ-340	1	88,3	G:C	0,194	0.3027±0.0009	N.S.	rs193631158
MLL25a-144	1	91,2	T:A	0,468	0.0187±0.0004	N.S.	rs193631160
MLL25a-183	1	90,9	T:A	0,032	1.0000±0.0000	N.S.	rs193631161
MYC-91	1	95,2	A:G	0,031	1.0000±0.0000	N.S.	rs193631171
MYL3-73	1	96,5	G:A	0,007	1.0000±0.0000	N.S.	rs193631162
MYL3-97	1	95,6	A:G	0,015	1.0000±0.0000	N.S.	rs193631163
RHO-111	1	94,9	T:G	0,022	1.0000±0.0000	N.S.	rs193631181
RPS7-2-69	1	85,4	A:G	0,145	0.0051±0.0001	N.S.	rs193631185
RPS7-2-278	1	77,7	T:A	0,095	<0.0000±0.0000	SIGN.	rs193631186
RPS7-2-313	1	87,2	T:A	0,121	<0.0000±0.0000	SIGN.	rs193631187
TPMA-53	1	92,5	G:T	0,489	<0.0000±0.0000	SIGN.	rs193631118
APOE-148	2	94,5	C:T	0,161	0.5838±0.0006	N.S.	rs193631117
CYCS-32	2	94,3	C:T	0,057	0.2384±0.0003	N.S.	rs193631133
CYCS-56	2	73,3	C:T	0,001	0.0001±0.0000	SIGN.	rs193631134
CYCS-132	2	94,9	C:T	0,262	0.8860±0.0003	N.S.	rs193631135
CYCS-218	2	89,9	T:C	0,076	0.0313±0.0002	N.S.	rs193631136
DAD1-444	2	93,8	C:T	0,049	1.0000±0.0000	N.S.	rs193631137
DAD1-483	2	97,8	G:T	0,029	1.0000±0.0000	N.S.	rs193631138
FGB-257	2	96,0	C:T	0,030	1.0000±0.0000	N.S.	rs193631141
FOS-107	2	96,3	C:T	0,069	0.0965±0.0003	N.S.	rs193631143
HGF-375	2	95,4	C:T	0,022	1.0000±0.0000	N.S.	rs193631146
HMOX1-269	2	90,7	T:A	0,104	0.0636±0.0003	N.S.	rs193631151

HMOX1-416	2	86,3	C:A	0,168	0.0975±0.0005	N.S.	rs193631152
MMP9-68	2	96,2	G:C	0,102	0.0301±0.0002	N.S.	rs193631164
MMP9-111	2	94,1	G:A	0,025	1.0000±0.0000	N.S.	rs193631165
MMP9-306	2	96,3	G:A	0,033	1.0000±0.0000	N.S.	rs193631166
MMP9-385	2	92,9	C:T	0,010	1.0000±0.0000	N.S.	rs193631167
MTF1-152	2	94,1	G:T	0,060	0.1745±0.0004	N.S.	rs193631168
MTF1-176	2	94,3	C:T	0,158	0.0089±0.0002	N.S.	rs193631169
MTF1-263	2	91,4	C:A	0,262	0.0076±0.0002	N.S.	rs193631170
PSM-33	2	90,5	G:A	0,413	0.6851±0.0009	N.S.	rs193631176
PSM-117	2	88,3	A:C	0,473	0.0081±0.0002	N.S.	rs193631177
PSM-138	2	84,6	A:C	0,265	0.6382±0.0009	N.S.	rs193631178
PTGS2-56	2	91,2	T:C	0,452	0.2164±0.0011	N.S.	rs193631189
PTGS2-115	2	85,0	A:G	0,337	<0.0000±0.0000	SIGN.	rs193631130
PTGS2-154	2	80,1	G:A	0,260	<0.0000±0.0000	SIGN.	rs193631131
PTGS2-317	2	87,6	A:G	0,422	0.6905±0.0009	N.S.	rs193631132
RASA3-188	2	96,2	A:G	0,028	1.0000±0.0000	N.S.	rs193631179
RHOC-55	2	95,2	G:A	0,022	1.0000±0.0000	N.S.	rs193631180
RPL12-213	2	86,3	G:T	0,208	<0.0000±0.0000	N.S.	rs193631182
RPL12-364	2	90,5	G:T	0,147	<0.0000±0.0000	SIGN.	rs193631183
RPL12-423	2	90,5	A:G	0,275	0.7945±0.0005	N.S.	rs193631184
c-mos-242	3	94,1	C:T	0,201	0.2308±0.0008	N.S.	rs193631128
c-mos-278	3	93,4	G:C	0,099	0.8246±0.0003	N.S.	rs193631129
CS1-197	3	95,6	A:G	0,342	0.7523±0.0007	N.S.	rs193631120
CS1-442	3	94,5	C:G	0,055	<0.0000±0.0000	N.S.	rs193631121
CS1-512	3	96,3	T:A	0,049	0.6139±0.0002	N.S.	rs193631122
CS3-137	3	90,3	A:T	0,130	0.0133±0.0002	SIGN.	rs193631123
CS3-394	3	86,1	C:T	0,379	0.2706±0.0011	N.S.	rs193631124
CS5-44	3	91,0	T:C	0,414	<0.0000±0.0000	N.S.	rs193631125
CS5-395	3	78,2	T:A	0,185	<0.0000±0.0000	SIGN.	rs193631126
CS5-425	3	86,3	G:T	0,151	<0.0000±0.0000	SIGN.	rs193631127
ELOVL2-519	3	92,0	C:G	0,167	0.0530±0.0004	N.S.	rs193631190
ELOVL3-205	3	87,0	G:A	0,003	1.0000±0.0000	N.S.	rs193631139
ELOVL3-365	3	95,1	C:T	0,064	0.3848±0.0003	N.S.	rs193631140
FGG-242	3	92,7	C:T	0,268	0.2180±0.0008	N.S.	rs193631142
GPX-273	3	95,6	T:C	0,075	0.1383±0.0005	N.S.	rs193631173
GPX-458	3	85,9	A:T	0,486	0.0201±0.0004	N.S.	rs193631174
HIF1A2-3-350	3	91,4	A:G	0,352	0.3324±0.0012	N.S.	rs193631147
HIF1A2-3-409	3	85,2	T:C	0,198	0.0106±0.0002	N.S.	rs193631148
HIF1A2-3-417	3	89,2	C:T	0,295	1.0000±0.0000	SIGN.	rs193631149
HIF1A4-219	3	94,7	C:T	0,101	0.5148±0.0004	N.S.	rs193631150
MB-188	3	91,2	G:A	0,089	0.0040±0.0001	N.S.	rs193631159
OPC02-249	3	94,5	C:T	0,086	0.7560±0.0003	N.S.	rs193631172
PRDX2-452	3	91,6	G:A	0,006	1.0000±0.0000	N.S.	rs193631175
RAG2-114	3	87,8	C:T	0,171	0.0011±0.0000	N.S.	rs193631191
Tmo-4C4-188	3	96,7	G:A	0,036	1.0000±0.0000	N.S.	rs193631188

Table S2.5. Results from tests of Linkage disequilibrium (LD) between all possible SNP pairs (validated SNPs only) within each DNA fragment, for both *T. alalunga* and *T. thynnus* (*$P < 0.001$). No SNPs were found in LD when comparing among DNA fragments. Linked SNPs were phased into haplotypes using the Bayesian statistical method implemented in PHASE 2.1. Haplotypes were reconstructed by location to avoid biases from population structuring. Missing genotypes were classified as null genotypes to avoid haplotype reconstruction errors.

SNP#1	**SNP#2**	**P-Value±S.E.**	
	Thunnus alalunga		
ALDOB1-47	*ALDOB1-95*	<0.0000±0.0000	*
ALDOB1-47	*ALDOB1-125*	0.1366±0.0018	
ALDOB1-95	*ALDOB1-125*	0.2523±0.0023	
LDHB-96	*LDHB-287*	0.0201±0.0009	
LYZ-128	*LYZ-138*	<0.0000±0.0000	*
LYZ-128	*LYZ-340*	<0.0000±0.0000	*
LYZ-138	*LYZ-340*	0.0004±0.0002	*
MLL25a-144	*MLL25a-183*	0.0001±0.0000	*
MYL3-73	*MYL3-97*	1.0000±0.0000	
CYCS-32	*CYCS-132*	0.0039±0.0004	
CYCS-32	*CYCS-218*	0.8505±0.0019	
CYCS-132	*CYCS-218*	0.0003±0.0001	*
DAD1-444	*DAD1-483*	0.3434±0.0013	
GPX-273	*GPX-458*	0.0141±0.0009	
HMOX1-269	*HMOX1-416*	0.6627±0.0029	
MMP9-68	*MMP9-111*	<0.0000±0.0000	*
MMP9-68	*MMP9-306*	0.1361±0.0019	
MMP9-68	*MMP9-385*	0.2668±0.0021	
MMP9-111	*MMP9-306*	0.7182±0.0007	
MMP9-111	*MMP9-385*	0.6325±0.0007	
MMP9-306	*MMP9-385*	0.4751±0.0010	
MTF1-152	*MTF1-176*	0.0159±0.0007	
MTF1-152	*MTF1-263*	0.1229±0.0020	
MTF1-176	*MTF1-263*	0.0213±0.0010	
PSM-33	*PSM-117*	<0.0000±0.0000	*
PSM-33	*PSM-138*	<0.0000±0.0000	*
PSM-117	*PSM-138*	<0.0000±0.0000	*
PTGS2-56	*PTGS2-317*	<0.0000±0.0000	*
RPL12-213	*RPL12-423*	<0.0000±0.0000	*
c-mos-242	*c-mos-278*	0.0033±0.0001	
CS1-197	*CS1-442*	0.0005±0.0001	*
CS1-197	*CS1-512*	0.0009±0.0001	*
CS1-442	*CS1-512*	<0.0000±0.0000	*
ELOVL3-205	*ELOVL3-365*	0.6814±0.0008	
HIF1A2-3-350	*HIF1A2-3-409*	0.0115±0.0008	

Thunnus thynnus			
GNRH3-1-107	*GNRH3-1-124*	0.0001±0.0000	*
LDHB-129	*LDHB-287*	0.5005±0.0026	
RPS7-2-278	*RPS7-2-313*	0.0276±0.0001	
CYCS-132	*CYCS-161*	0.5807±0.0013	
MTF1-176	*MTF1-263*	0.6764±0.0013	
CS3-118	*CS3-394*	1.0000±0.0000	
CS5-395	*CS5-425*	<0.0000±0.0000	*

Table S2.6. Characteristics of the 23 validated SNPs in *Thunnus thynnus*. Marker name, SNP discovery approach (1,2 and 3), call rate (%), alleles and minor allele frequency (MAF) values are shown along with the global and per location deviation (*P*) from the Hardy-Weinberg equilibrium (HWE). "SIGN." means that at least 1 location was not in HWE for those markers (N.S. corresponds to not significant); as multiple comparisons were involved, a probability of $P < 0.001$ was considered significant. Out of the 23 SNPs validated in BFT by cross-species amplification of 128 SNPs discovered in albacore (see Methods and Fig. 2.2.), 18 had been validated also when genotyping albacore while the other five SNPs (*GNRH3-1-107, LDHB-129, CYCS-161, CS3-118,* and *OPC02*-45) were reported as reliably scored and polymorphic only in BFT (last column shows their individual NCBI Assay IDs (ss#) in the NCBI´s dbSNP database).

SNP	Method	Call rate	Alleles	MAF	Global HWE	HWE per sample	NCBI Assay ID
ADRB2-97	1	94,90	G:C	0,040	1,0000	N.S.	*see Table S2.4*
GNRH3-1-107	1	97,50	T:A	0,305	0,1040	N.S.	ss550227148
GNRH3-1-124	1	92,40	T:A	0,384	0,6299	N.S.	*see Table S2.4.*
LDHB-129	1	71,00	A:C	0,336	1,0000	N.S.	ss550227246
LDHB-287	1	74,70	T:G	0,068	0,2215	N.S.	*see Table S2.4.*
LYZ-128	1	84,80	T:C	0,045	0,0020	N.S.	*see Table S2.4.*
RPS7-2-278	1	74,70	C:A	0,051	0,2253	N.S.	*see Table S2.4.*
RPS7-2-313	1	77,20	A:C	0,236	0,0022	N.S.	*see Table S2.4.*
TPMA-53	1	79,70	C:T	0,206	0,3046	N.S.	*see Table S2.4.*
CYCS-132	2	89,90	C:G	0,035	1,0000	N.S.	*see Table S2.4.*
CYCS-161	2	72,10	T:A	0,139	0,0382	N.S.	ss550227317
HGF-375	2	98,70	G:C	0,122	0,0789	N.S.	*see Table S2.4.*
MTF1-176	2	87,30	G:C	0,123	1,0000	N.S.	*see Table S2.4.*
MTF1-263	2	82,30	G:A	0,185	1,0000	N.S.	*see Table S2.4.*
RHOC-55	2	83,50	T:A	0,053	1,0000	SIGN.	*see Table S2.4.*
RPL12-423	2	88,60	A:T	0,393	1,0000	N.S.	*see Table S2.4.*
CS3-118	3	78,50	T:A	0,315	0,5711	N.S.	ss550227367
CS3-394	3	84,80	G:C	0,015	1,0000	N.S.	*see Table S2.4.*
CS5-395	3	83,50	C:A	0,189	0,0358	N.S.	*see Table S2.4.*
CS5-425	3	91,10	C:T	0,208	0,0607	N.S.	*see Table S2.4.*
FGG-242	3	98,70	G:C	0,359	0,2577	N.S.	*see Table S2.4.*
HIF1A2-3-417	3	97,50	G:C	0,084	0,1010	N.S.	*see Table S2.4.*
OPC02-45	3	92,40	C:T	0,048	1,0000	N.S.	ss550227450

Capítulo III

New nuclear SNP markers unravel the genetic structure and effective population size of albacore tuna (*Thunnus alalunga*)

ESTUDIO II

New nuclear SNP markers unravel the genetic structure and effective population size of albacore tuna (*Thunnus alalunga*)

Urtzi Laconcha, Mikel Iriondo, Haritz Arrizabalaga, Carmen Manzano, Pablo Markaide, Iratxe Montes, Iratxe Zarraonaindia, Igor Velado, Eider Bilbao, Nicolas Goñi, Josu Santiago, Andrés Domingo, Saadet Karakulak, Işık K. Oray, Andone Estonba

Abstract

In the present study we have investigated the population genetic structure of albacore (*Thunnus alalunga*, Bonnaterre 1788) and assessed the loss of genetic diversity, likely due to overfishing, of albacore population in the North Atlantic Ocean. For this purpose, 1,331 individuals from 26 worldwide locations were analyzed by genotyping 75 novel nuclear SNPs. Our results indicated the existence of four genetically homogeneous populations delimited within the Mediterranean Sea, the Atlantic Ocean, the Indian Ocean and the Pacific Ocean. Current definition of stocks allows the sustainable management of albacore since no stock includes more than one genetic entity. In addition, *short-* and *long-term* effective population sizes were estimated for the North Atlantic Ocean albacore population, and results showed no historical decline for this population. Therefore, the genetic diversity and, consequently, the adaptive potential of this population have not been significantly affected by overfishing.

Keywords: Stock definition, sustainable management**,** conservation genetics, fisheries.

1. Introduction

Albacore tuna (*Thunnus alalunga*, Bonnaterre 1788) is distributed in the Atlantic, Pacific and Indian Oceans and in the Mediterranean Sea, extending from 50-55ºN to 40-45ºS (Arrizabalaga *et al.* 2014). This species is the fourth most important one of the *Thunnus* genus with regard to captures (FAO 2014). This fact reflects the high commercial value of the albacore and its related products, which makes this species likely to be exploited beyond its maximum sustainable yield (Arrizabalaga *et al.* 2014). Migrations on this species has been studied for several decades through tag-recapture experiments showing low rate of albacore migration between hemispheres (Arrizabalaga *et al.* 2002), and no transoceanic (Childers *et al.* 2011; Farley *et al.* 2013;

Arrizabalaga *et al.* 2014) neither Atlantic-Mediterranean migrations (Arrizabalaga *et al.* 2004). There are very few studies on spawning areas of this species, because catching larvae or young-of-the-year individuals (reference samples) of this species is not a very common event. One spawning ground has been defined in the western Mediterranean (Duclerc *et al.* 1973; Lalami *et al.* 1973; Dicenta *et al.* 1975), two spawning areas in the North Atlantic Ocean (Beardsley 1969; Carlsson 2007), a single one in the South Atlantic (Beardsley 1969), one spawning area in the Indian Ocean (Koto 1969; Shiohama 1985), and two Pacific separate spawning grounds: north and south (Nakamura 1969; Ueyanagi 1969; Sund *et al.* 1981). According to this knowledge on population dynamics of albacore, six stocks or management units are currently defined by Regional Fisheries Management Organizations (RFMOs): (i) Mediterranean Sea, (ii) North Atlantic Ocean, (iii) South Atlantic Ocean, (iv) Indian Ocean, (v) North Pacific Ocean and (vi) South Pacific Ocean. Many fisheries are regulated in accordance with spatial schemes. However, management units based only on knowledge about migrations do not necessarily correspond to the biological structure of the populations (Ward 2000; Waples & Gaggiotti 2006). In these cases, when fishery management is not based or does not fit the biological structure, changes may occur in the biological attributes, productivity and genetic diversity of the exploited species (Ricker 1981). Therefore, the establishment of an accurate population boundary for a commercial species requires a multidisciplinary approach, and genetic studies can contribute very valuable information in this regard (Pawson & Jennings 1996; Waldman 1999). Thus, studies including population genetic structure assessment together with other population identification methodologies, such as tag-recapture (Arrizabalaga *et al.* 2004) or chemical tags in otoliths (Miller *et al.* 2010), have become more common in the last decade. These multidisciplinary studies allow a more accurate population structure and hence, more sustainable fisheries management policies.

A variety of studies have assessed population structure of albacore species using multiple approaches including: otolith microstructure (García *et al.* 2006; Macdonald *et al.* 2013), tag-recapture methods (Arrizabalaga *et al.* 2004), morphometrics (Bard 1981) and genetic techniques (Suzuki 1962; Graves & Dizon 1989; Chow S & Ushiama 1995, Takagi *et al.* 2001; Viñas *et al.* 2004; Nakadate *et al.* 2005; Davies *et al.* 2011; Montes *et al.* 2012; Albaina *et al.* 2013). The population structure of albacore has been found to exhibit a high dispersal capacity (e.g. Fontaine et al. 2007), similarly to what happens to other marine species such as Atlantic bluefin tuna (*Thunnus thynnus*) (Albaina *et al.* 2013; Riccioni *et al.* 2010) or Atlantic mackerel (*Scomber scombrus*) (Jansen & Gislason 2013). However, despite the number of studies performed since the last decade, genetic structure of albacore is not clear yet, since contradictory information about number of albacore populations and population boundaries have been reported. In this regard, Albaina *et al.* (2013) suggested four albacore populations (one in each ocean and one in the Mediterranean Sea), but Pujolar *et al.* (2003) and Graves and Dizon (1989) found genetic homogeneity between the Atlantic Ocean and Mediterranean Sea and Atlantic and Pacific Oceans, respectively, and Montes *et al.* (2012) found homogeneity between the Atlantic and Indian Oceans. Moreover, genetic structure within oceans remains unclear since heterogeneity within them or within the Mediterranean Sea has

been suggested (Suzuki 1962; Marano *et al.* 1999; Takagi *et al.* 2001; Arrizabalaga *et al.* 2004; Davies *et al.* 2011; Montes *et al* 2012). Comparison between studies is difficult because differences on genetic markers studied and also on geographic areas assessed, which in certain studies are very limited into the bargain. In fact, few studies have addressed the population structure of albacore covering the worldwide distribution range of the species (Arrizabalaga *et al.* 2014; Montes *et al.* 2012; Albaina *et al.* 2013).

The North Atlantic albacore tuna stock was subjected to overfishing conditions between the mid 1960s and mid 2000s. As a result, the spawning stock biomass had been overexploited (below levels associated to the maximum sustainable yield) since the 1980s, but is now recovering over the last decade (ICCAT 2014).After the population genetic structure of a species is defined, an essential parameter that informs about the sustainable management and conservation of exploited species is the effective population size (N_e) (Hauser & Carvalho 2008). While population genetic structure enables a definition of populations, that can be linked to the stock or management unit concept, N_e determines how vulnerable these populations are to losing genetic diversity due to genetic drift (O'Leary *et al.* 2013) and consequently, this variable assesses their responsiveness and adaptation capabilities. Despite the importance of this parameter for populations' conservation, few studies have estimated N_e for tunas (Riccioni *et al.* 2010; Qiu & Miyamoto 2011; Qiu et al. 2013; Bravington et al. 2014; Nomura et al. 2014).

In summary, a number of outstanding issues persist which have direct implications for the sustainable management of albacore. These main questions to be answered include (1) the absence of a consensus about the genetic structure of this species worldwide, and (2) uncertainty about the impact of fishing on the effective population size (N_e) and, therefore, on the genetic diversity of albacore populations. The goal of this study is to obtain a clear definition of the population genetic structure of albacore, and to shed light on its genetic viability via the estimation of N_e for the North Atlantic population, with the aim of providing a more rational foundation for sustainable fishery management. With this objective in mind, we carried out the most extensive sampling of albacore to date, covering its worldwide distribution range (1,331 samples from 26 locations worldwide). The number of markers employed was also the highest used to date, involving 115 novel nuclear SNP markers which we report in albacore tuna through cross-species transcriptome amplification and sequencing.

2. Material and Methods

2.1. Samples and DNA extraction

An exhaustive spatial-temporal sample of 1,331 albacore individuals from 26 locations covering the whole geographical distribution of the species was obtained (Figure 3.1, Table 3.1). The total sample includes 774 individuals from the Atlantic Ocean (12 locations sampled over 24 years), 254

individuals from Mediterranean Sea (7 locations sampled over 12 years), 136 individuals from the Indian Ocean (4 locations within 4 years of sampling), and 167 individuals from the Pacific Ocean (3 locations sampled over 5 years). Individuals were mainly sampled between 2008 and 2012, with some individuals sampled in previous years as far back as 1988. Sampled individuals were provided either by commercial or recreational vessels or by oceanographic institutes that collected the samples during scientific surveys. All fish were collected as part of authorized routine fishing procedures and therefore did not require any special additional permission. Some samples were used in previous studies (Arrizabalaga *et al*. 2004; Davies *et al.* 2011; Montes *et al.* 2012; Albaina *et al.* 2013; Santiago & Arrizabalaga 2005) (Table 3.1). Collected tissues mainly consisted of muscle, fin or heart tissue, and they were stored either frozen at -20ºC or preserved in 96% ethanol at 4ºC. Additionally, spine cuts mounted in Eukitt (O. Kindler GmbH), as well as dried and stained blood samples were collected (Table 3.1). DNA from muscle, fin and heart tissue samples was extracted using NucleoSpin 96 Tissue Kit (Macherey-Nagel). Spine and blood samples were first immersed in xylol, and spine samples were afterwards manually crushed; DNA from these samples was extracted by means of a specific membrane using QIAmp DNA Investigator Kit (Qiagen). DNA from all samples was quantified using both a NanoDrop 1000 spectrophotometer (Thermo Fisher Scientific Inc., Waltham, MA) and a Qubit 2.0 (Invitrogen, Life Technologies) fluorometer. All DNA samples were stored at -20ºC for subsequent analyses.

Table 3.1. Sample number (#), name, location, year of capture, number of individuals (N), stock, geographic coordinates and sample type (T, muscle, fin or heart tissue; B, blood; S, spine)

#	Name	Location	Year	N	Stock	Latitude	Longitude	Sample type
1 [A]	ADR	Adriatic Sea	2006	48	Mediterranean Sea	41.29	17.52	T
2 [A,B,C]	BAL	Balearic Sea	2005	31	Mediterranean Sea	40.00	1.58	T
3	CYP	Cyprus	2011	10	Mediterranean Sea	36.08	33.68	T
4	TUR	Turkey	2011	53	Mediterranean Sea	35.04	26.80	T
5 [A]	TYR	Tyrrhenian Sea	2008	48	Mediterranean Sea	38.88	11.74	T
6	LIG	Ligurian Sea	2011	27	Mediterranean Sea	43.38	9.05	T
7 [D]	ALB	Alboran Sea	1999	37	Mediterranean Sea	36.23	-2.00	B
8 [E]	B88	Bay of Biscay	1988	34	North Atlantic Ocean	45.10	-4.35	S
9 [E]	B89	Bay of Biscay	1989	30	North Atlantic Ocean	45.64	-4.76	S
10 [A,B]	B09	Bay of Biscay	2009	42	North Atlantic Ocean	45.05	-5.28	T
11	B10	Bay of Biscay	2010	240	North Atlantic Ocean	45.71	-5.53	T
12	B11	Bay of Biscay	2011	31	North Atlantic Ocean	44.92	-4.16	T
13	CAN	Canary Islands	2012	41	North Atlantic Ocean	27.73	-17.25	T
14 [A,B]	IRE	Ireland	2008	57	North Atlantic Ocean	54.17	-12.89	T
15	MAU	Mauritania	2010	48	North Atlantic	9.79	-32.16	T

					Ocean			
16	GUI	Gulf of Guinea	1999-2000	32	South AtlanticOcean	1.98	-16.58	B
17	URU	Uruguay	2005, 2007-2012	84	South Atlantic Ocean	-36.19	-53.16	T
18 [A,B]	SCA	South Africa	2009	98	South Atlantic Ocean	-24.56	4.42	T
19	SOA	South Africa	2011	37	South Atlantic Ocean	-34.34	18.00	T
20 [A,B]	I08	Seychelles	2008-2009	23	Indian Ocean	-7.11	54.65	T
21	I10	Seychelles	2010	38	Indian Ocean	-7.27	56.32	T
22	I11	Seychelles	2011	42	Indian Ocean	-7.28	49.06	T
23	I12	Seychelles	2012	33	Indian Ocean	-8.86	49.13	T
24 [A,B]	NPA	California	2008	83	North Pacific Ocean	43.50	-127.00	T
25 [A,B]	SEP	New Caledonia	2004-2005	51	South Pacific Ocean	-19.01	-152.84	T
26 [A,B]	SWP	French Polynesia	2003-2008	33	South Pacific Ocean	-18.53	165.97	T
			1988-2012	Total= 1,331				

[A] Sample previously analyzed in [33]
[B] Sample previously analyzed in [34]
[C] Sample previously analyzed in [32]
[D] Sample previously analyzed in [6]
[E] Sample previously analyzed in [47]

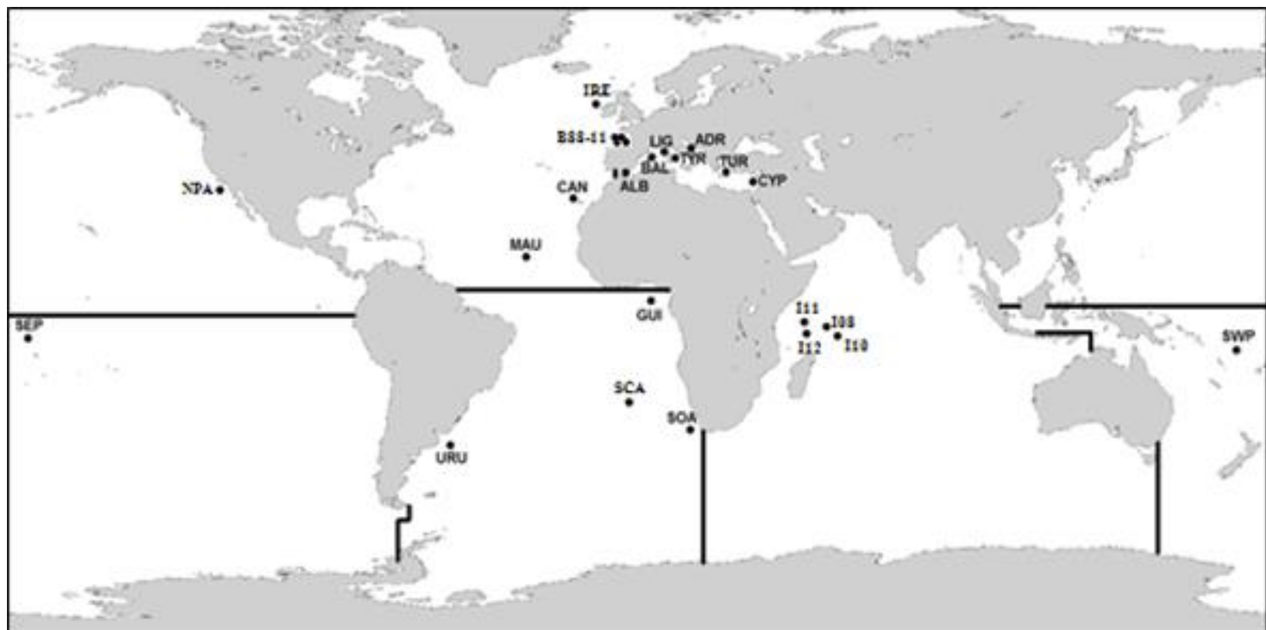


Figure 3.1. Sampling locations. Current stock boundaries delineated with black lines. Sample abbreviations are as defined in Table 1.

2.2. SNP selection and genotyping

The SNPs used in this study were previously discovered in the closely related ABFT species through transcriptome and genome sequencing, using 454 (GS FLEX Titanium) and HiSeq2000 (Illumina), respectively (Cariani *et al*. Personal Communication). Of all the discovered SNPs in the ABFT species, 384 transcriptome SNPs were genotyped using the GoldenGate platform (VeraCode), in 30 albacore samples covering the entire distribution range of the species (5 individuals from each defined management unit: North Atlantic, South Atlantic, Mediterranean Sea, Indian Ocean, North Pacific and South Pacific). From the 384 ABFT SNPs, only those that successfully amplified in albacore (conversion rate) and have Minor Allele Frequency (MAF) values above 0.01 in the latter species were taken into account. From these, only those markers that were compatible with TaqMan OpenArray technology (Life Technologies) were selected for this study. Additionally, two nuclear SNPs, previously described for albacore (Albaina *et al.* 2013; Table S3.1), were included in the final SNP set as a positive control, in order to corroborate the correctness of the genotyping procedure. Thus, a final set of 117 SNPs was designed to genotype 1,331 albacore individuals through TaqMan OpenArray technology. Validation rate was calculated as the proportion of SNP with a MAF > 0.001. In order to ensure genotyping quality, SNPs needed to comply with the following criteria: a call rate higher than 80%, clear genotyping clusters, and compliance with Hardy-Weinberg equilibrium (HWE).

2.3. Statistical analysis

GENEPOP v4.0 (Rousset 2008) software was used to test departures from HWE and to analyze genotypic disequilibrium (GD) between SNPs (p-value < 0.001). Linked SNPs were phased into haplotypes using PHASE v2.1 software (Stephens *et al.* 2001).

In order to assess the genetic population structure of the albacore, Reynolds genetic distance matrices (Reynolds *et al.* 1983) were obtained using Populations v1.2.32 software (Langella 2002). A Neighbor-Net dendrogram was constructed using SPLITSTREE v4.13 (Hudson & Bryant 2008) based on the matrix of genetic distances. Geographic distance was calculated measuring the shortest distance by sea between each pair of sample location using scripts from the Movable Type Ltd webpage (http://www.movable-type.co.uk/scripts/latlong.html). Isolation by distance (IBD) was tested evaluating the correlation between Rousset's genetic distance (Rousset 1996) and geographic distance, using Mantel test implemented in IBDWS (Jensen *et al.* 2005) with 30,000 randomizations. Population genetic structure was also assessed using STRUCTURE v2.3.4 (Pritchard *et al.* 2000) and GENELAND v3.2.2 (Guillot *et al.* 2008) software, which are based on Bayesian clustering algorithms that allow assigning individuals to a group without previous assumption of either population units or population boundaries. STRUCTURE was run using the mixed ancestry model and correlated allele frequencies (Falush *et al.* 2003), using information regarding sampling location. Ten independent runs were simulated for each potential number of

populations (K) with values of K = 1-6, and with a burn-in period of 50,000 Markov chain Monte Carlo (MCMC) steps, followed by 500,000 MCMC steps. The best K was estimated as proposed by Pritchard *et al*. (2000). CLUMPP v1.1.2 (Jakobsson & Rosenberg 2007) was used to determine the optimal assignation of clusters for the analyzed individuals, maximizing similarity between the 10 different STRUCTURE replications for the selected K. Individual membership coefficients were graphically shaped with DISTRUCT v1.1 (Rosenberg 2004). Finally, to test potential weaker structure within the detected major clusters, STRUCTURE analysis was repeated for each of them. While STRUCTURE is based only on the individual genotype data to infer the population structure, GENELAND uses the geographical information of the individuals as an additional parameter in the analysis. In the latter case, K was estimated from 1 to 5, using 500,000 MCMC iterations and 1,000 thinnings. Ten runs with fixed K were then post processed using a burn-in of 50,000 iterations to obtain the posterior probabilities of population membership for each individual and each pixel of the spatial domain.

We searched for candidate *loci* under selection (outlier *loci*) using the Bayesian likelihood method, as implemented in BAYESCAN v2.1 (Foll & Gaggiotti 2008), with 10 pilot runs of 5,000 iterations and an additional burn-in of 50,000 iterations (sample size of 5,000 and thinning interval of 10). Critical values for the test were adjusted with false discovery rate (FDR) procedure (q-value < 0.05) (Benjamini & Hochberg 1995). Pairwise F_{ST} (Weir & Cockerham 1984) values among samples based on neutral markers were estimated with FSTAT v2.9.3 software (Goudet 2001). P-values were weighted using the FDR method for multiple testing (Benjamini & Hochberg 1995).

The statistical power required to detect various levels of differentiation with the SNPs used in this study was estimated using POWSIM version 4.1 (Ryman *et al.* 2006). Since POWSIM is restricted to 50 loci, we selected those 50 loci with highest F_{ST} values. Burn-in consisted of 1000 steps followed by 100 batches of 1000 steps. Chi-square probabilities were used to test the significance of an F_{ST} value for each replicate run. The number of significant F_{ST} values in 1000 replicate simulations provided an estimate of the statistical power for a given level of divergence, which was controlled by allowing frequencies to drift for a given number of generations. Simulated effective populations sizes equaled 2000 fish.

Two different time-scale N_e estimates were obtained for the North Atlantic stock. *Short-term* N_e was estimated from temporal fluctuations in allele frequencies between cohorts (Waples 1989), and a correction for overlapping generations was applied (Jorde & Ryman 1995, 2007; Waples & Yokota 2007). Generation time ($\hat{G}$) was estimated following Felsenstein (1971) from age frequency data of analyzed years (1988–2012), and changes in allele frequencies among cohorts were measured by F_S (Jorde & Ryman 2007). The *long-term* N_e (Beerli & Felsenstein 2001) uses a maximum likelihood estimator based on the coalescence theory. It is a retrospective model of population genetics which traces back for the most recent individual from which all organisms in a group are directly descended, the most recent common ancestor (MRCA). This tool has been employed to estimate historic population sizes for a range of species (Gemmell *et al.* 2004).

Data for North Atlantic albacore were obtained between 1988-2012, which constitutes 4-5 generations of albacore assuming 50% maturity at age 5 (Bard 1981). Age was estimated using length and weight information according to Santiago (1993) and Santiago and Arrizabalaga (2005). We used age-structure data for seven cohorts (Figure 3.2). Cohort analysis was carried out to assess temporal fluctuations in population size. The adult population size (N_c) in the North Atlantic, obtained from the report of the 2013 ICCAT North and South Atlantic albacore stock assessment (ICCAT 2014), was compared with total population size N_e estimates. MIGRATE v3.2.1 software (Beerli 2002) was used for *long-term* N_e estimation, and mutation was modeled by an infinite allele model.

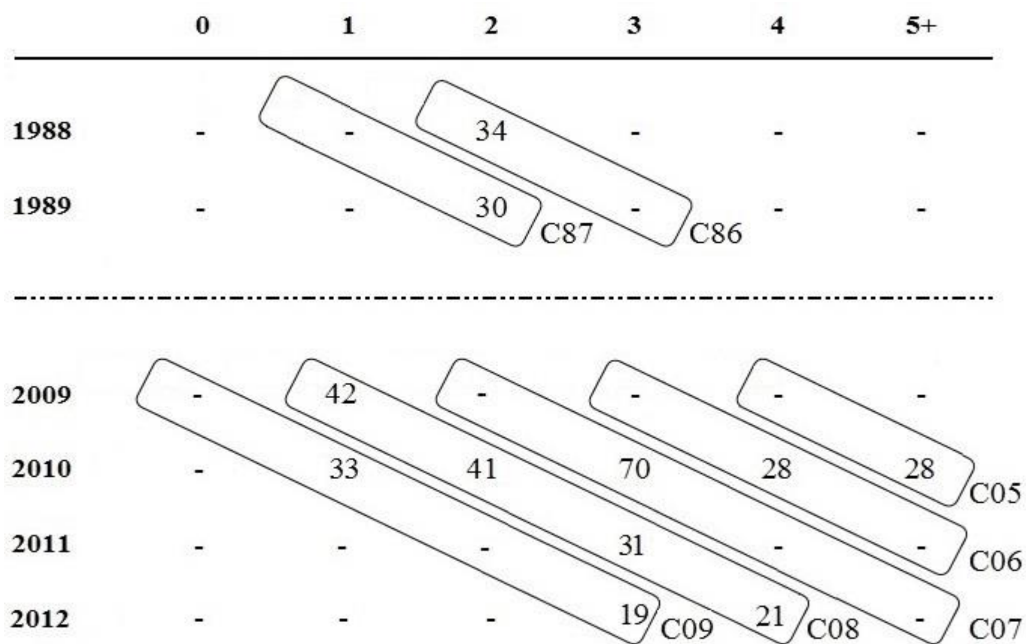


Figure 3.2. Diagram of the defined cohorts, based on the age of the individuals. Columns indicate age of individuals and rows year of capture. Values inside tables are the number of individuals for each age/year combination. Seven cohorts were defined by diagonal frames whose names were based on the hatching year of the individuals: C86, C87, C05, C06, C07, C08 and C09

3. Results

3.1. SNP selection and genotyping

From the 384 ABFT SNPs analyzed in the 30 albacore sample, 311 SNPs (conversion rate = 80.99%) successfully amplified in albacore, and among them, 121 showed MAF > 0.01 (31.51%). From these, 115 SNPs exhibited compatibility with the TaqMan OpenArray technology (Life Technologies), and were further genotyped together with the 2 nuclear SNPs included as a positive control.

Out of 117 nuclear SNPs, 95 were polymorphic (they had a MAF value above 0.001, i.e. the minor allele was observed at least 5 times) and had a clear genotype for the 1,331 albacore individuals (Table S3.1). Therefore, validation rate was 24.61% (95/386). From these, 76 met HWE. The exact tests for genotypic disequilibrium (GD) detected 2 SNPs (ss974292126 and ss974292127) with significant GD probabilities, so these 2 SNPs were phased into one haplotype (ss974292126+ss974292127). Therefore, a set of 75 independent nuclear markers was downstream analyzed.

3.2. Population structure

The Neighbor-Net drawn from Reynolds genetic distances (Figure 3.3) grouped locations according to their geographical region. The Mediterranean Sea samples, grouped into a single cluster, were the most distant from the rest. The samples from the three oceans also grouped by ocean, and those from the Indian Ocean were placed between those of the Atlantic and those of the Pacific. The genetic and geographic distances for the 26 samples showed a significant correlation ($r = 0.4577$, $p < 0.0001$; Figure S3.1). This correlation increased notably when the Mediterranean samples were removed from the analysis ($r = 0.7549$, $p < 0.0001$; 19 locations). Within the Mediterranean, no significant correlation was found between genetic and geographic distances ($r = -0.3210$, $p = 0.0954$; Figure S3.1).

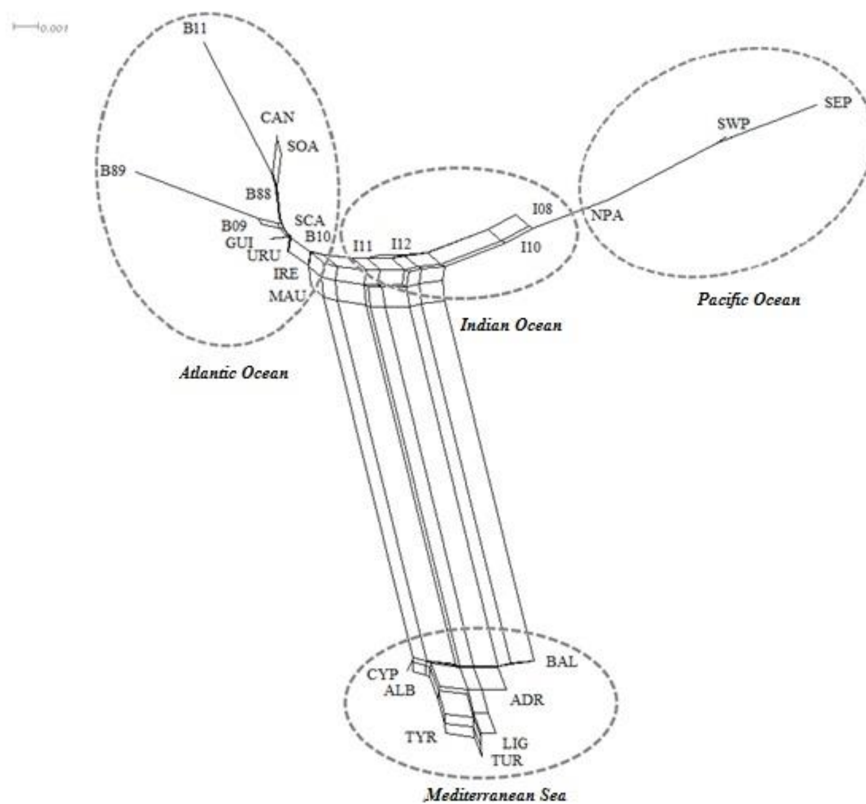


Figure 3.3. Neighbor-Net dendrogram built from Reynolds distances between 26 samples. Sample abbreviations are as defined in Table 1.

With respect to the analysis of individual clustering using the STRUCTURE software, when 2 group clusters were considered (K = 2) a clear distinction could be observed between the samples from the Mediterranean Sea and the others (Figure 3.4). In any event, the best K value obtained was 3 (Figure S3.2a), which clearly distinguished the Pacific Ocean (red) samples from those of the Atlantic Ocean (mostly green; Figure 3.4). The case of samples from the Indian Ocean is special in that we observed intermediate percentages of the components of the Atlantic and the Pacific. In the same way, the GENELAND software also detected K=3 as the most probable number of groups (Figure S3.2b). In this analysis, the 3 clusters were made up of the Mediterranean samples (cluster 1), the Atlantic samples (cluster 2) and the Indo-Pacific samples (cluster 3). When STRUCTURE analysis was repeated for each of the major clusters, no structure was detected within them, since the best K value obtained was 1 for all the analysis.

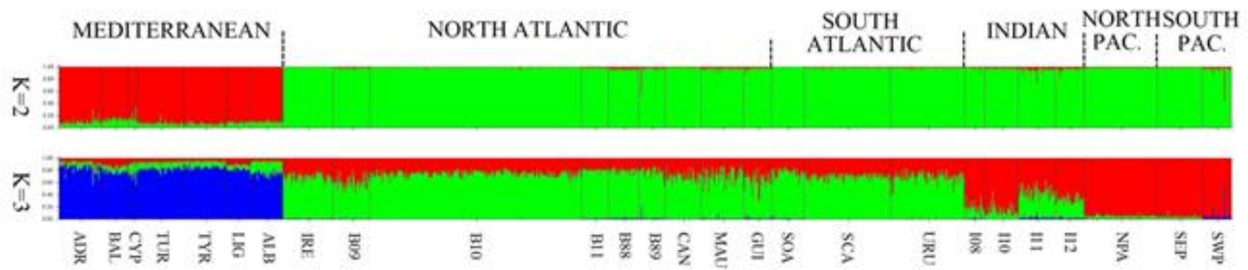


Figure 3.4. Individual clustering analysis implemented with STRUCTURE software for K=2 and K=3. Each vertical bar represents an individual. The 26 locations are separated by horizontal continuous black lines, and the currently accepted 6 stocks are separated by discontinuous horizontal black lines. The color proportions of each bar correspond to individuals' estimated membership fractions of each of the clusters. Sample abbreviations are as defined in Table 1.

A total of 17 out of the 75 independent markers were identified as outliers, therefore, 58 SNPs were defined as neutral SNPs. Heterogeneity analyses performed within stocks based on the 58 neutral SNPs revealed that the 6 stocks defined by the Regional Fisheries Management Organizations (RFMOs) were genetically homogeneous ($p > 0.05$; Table 3.2). POWSIM simulations showed that the 50 SNPs with the highest F_{ST} values together were able to detect significant differences among samples with F_{ST}= 0.0015 in about 95% of the tests, and with F_{ST}= 0.002 in 100% of the tests (Table 3.3). The F_{ST} values between stocks varied from a minimum F_{ST} = 0.001 between the North and South Atlantic and between the North and South Pacific, and a maximum F_{ST} = 0.051 between the South Pacific and the Mediterranean stock. All comparisons were found to be statistically significant, except those obtained between the North and South Atlantic, the North and South Pacific and between the North Pacific and the Indian Ocean (Table 3.2).

Table 3.2. Pairwise F_{ST} values (below the diagonal) and *p*-values (above the diagonal) between the 6 stocks currently recognized by the RFMOs. Stock abbreviations: MED (Mediterranean), NATL (North Atlantic), SATL (South Atlantic), IN (Indian), NPAC (North Pacific) and SPAC (South Pacific). F_{ST} values among locations within stocks are shown on the diagonal, and none of them were significant (*p*-value > 0.05).

	MED	**NATL**	**SATL**	**IN**	**NPAC**	**SPAC**
MED	0.003	<0.001*	<0.001*	<0.001*	<0.001*	<0.001*
NATL	0.035*	0.004	0.783	<0.001*	<0.001*	<0.001*
SATL	0.033*	0.001	0.000	<0.001*	<0.001*	<0.001*
IN	0.038*	0.010*	0.008*	0.002	0.125	0.038*
NPAC	0.049*	0.025*	0.022*	0.002	-	0.405
SPAC	0.051*	0.026*	0.023*	0.003*	0.001	0.004

Regarding adaptation of the populations to the specific environmental conditions of their surroundings, the 17 markers identified as outliers using BAYESCAN were analyzed. The defined haplotype ss974292126+ss974292127 had a positive alpha value and significant high F_{ST} value, suggesting that it may be subject to divergent selection (Foll & Gaggiotti 2008). This haplotype was practically monomorphic in non-Mediterranean samples (fCC = 0.999 and fCA = 0.001), while haplotype frequencies in the Mediterranean Sea were significantly different (fTA = 0.127; fCC = 0.819; fTC = 0.016; fCA = 0.038; F_{ST} = 0.311; $p < 0.0001$). The remaining 16 outlier SNPs showed a negative value for alpha and significant low F_{ST} values, a result consistent with balancing selection (Figure S3). Candidate genes involved in essential metabolic pathways were found by homology between our sequence data surrounding the 18 outlier SNPs, and previously known teleost genes (Table S3.2).

Table 3.3 Probability of detecting a particular level of differentiation (F_{ST}) among populations

F_{ST}	$P\chi^2$
0.0005	0.351
0.0015	0.952
0.0020	**0.999**
0.0025	1.000
0.0050	1.000

3.3. Effective population size

Effective population size (N_e) was estimated for the North Atlantic Ocean analyzing the 58 neutral SNP markers. While *short-term* N_e ranged between 5,466 and 23,330 (C07 and C08 cohorts, respectively) (mean *short-term* N_e = 13,267 ± 6,049; Table S3), *long-term* N_e varied between 13,897 and 20,304 (C08 and C06 cohorts, respectively) (mean *long-term* N_e = 16,729 ± 2,248; Table S4). Mean *short-* and *long-term* N_e were not significantly different (Mann-Whitney U, *p*-value > 0.05). The *short-term* N_e was compared to N_c (Figure 3.5), and N_e/N_c ratio (ratio of effective-to-census size) values were found to range between 2.62×10^{-3} and 9.83×10^{-3} (C07 and C08 cohorts, respectively). Despite the apparent correlation between N_e and N_c, it was not found to be statistically significant (r = 0.383; *p*-value = 0.453).

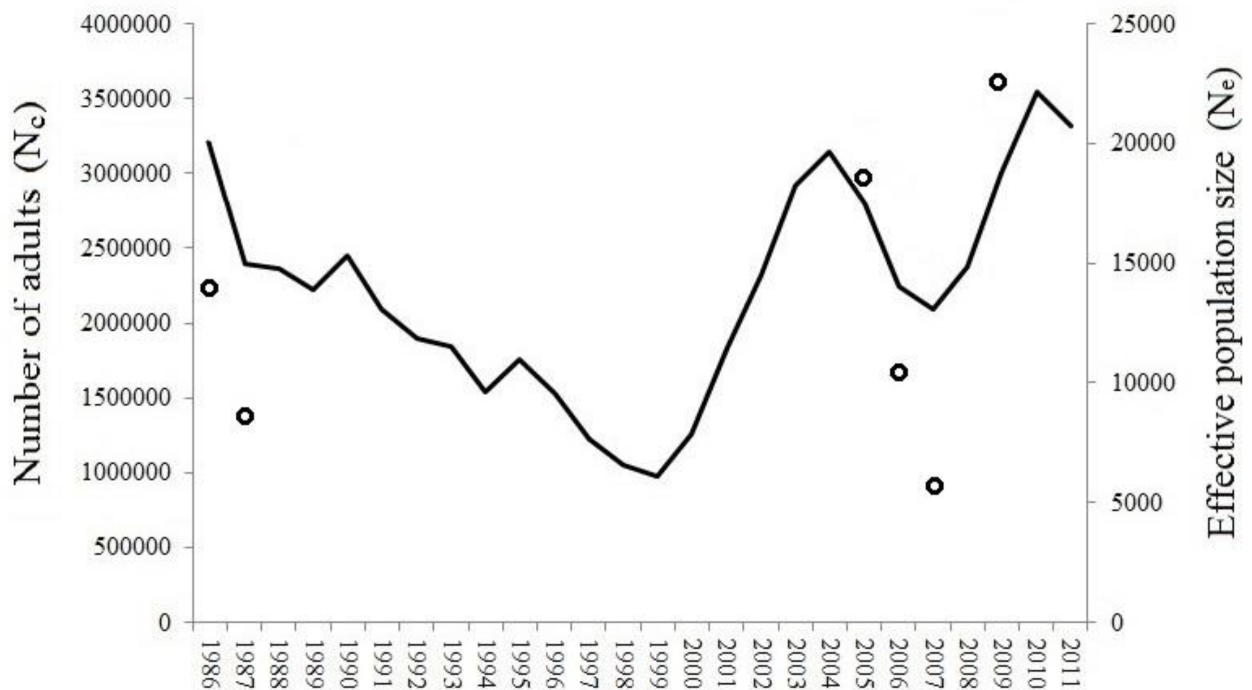


Figure 5. Estimates of the evolution of the number of adults from 1986 to 2011 (solid line, N_c), and the *short-term* effective population size of each cohort in the North Atlantic (dots, N_e).

4. Discussion

A sustainable management of fisheries requires the exploitation of one single population per stock, and accurate population size estimates (Hauser & Carvalho 2008). In this way, two problems that reduce intraspecific diversity are avoided: overexploitation and the risk of losing minority populations when various populations are managed as a single stock. The present study is the most comprehensive genetic study carried out to date of the albacore species worldwide. Overall, 117 novel nuclear SNPs were applied to 1,331 albacore individuals from 26 locations covering the whole distribution area of the species. We have described the genetic structure of the species, provided effective population size estimates for the North Atlantic Ocean population, and reported putative signs of natural selection in the albacore genome. Results obtained indicated that none of the currently defined 6 management units includes more than one genetic population. Regarding population size, N_e estimates ruled out the occurrence of severe historical bottlenecks in the North Atlantic Ocean population, and showed that current levels of genetic diversity are sustainable over the time, thereby corroborating the resiliency and responsiveness of the albacore. All these results on albacore population genetic characteristics should contribute to more rational and sustainable fisheries management policies and programs for this important fish species.

4.1 Cross-species amplification of SNPs

This study has shown that cross-species amplification is a valuable approach to identify SNP markers in the albacore species, with a final validation rate of 24.61%. The reciprocal cross was reported by Albaina *et al.* (2013), who showed that albacore and the ABFT species shared 18% of SNPs. Cross-amplification success between Atlantic herring (*Clupea harengus*) and Pacific herring (*Clupea pallasii*) is even lower, 12% (Helyar *et al.* 2012). The higher success obtained in the present study lies in the high number of individuals and the assortment of their geographical origins. Here, 1,331 albacore individuals worldwide were studied, whereas four Pacific herrings were analyzed by Helyar *et al.* (2012) and 107 Atlantic bluefin tunas by Albaina *et al.* (2013). Cross-amplification success also relies on the design of an appropriate SNP set, such as the 384 SNPs from coding regions analyzed in this study. In conserved regions of the genome, such as coding regions, the similarity between the analyzed sequences of two species is increased and therefore, the chance to share SNPs also increases.

Cross-species amplification is considered a valuable approach to identify SNP markers in non-model organisms. Additionally, when these SNPs are located in genes, as is the case in the present study, they can be used for local adaptation studies. In this regard, ss974292126+ss974292127 SNP haplotype was identified as an outlier in BAYESCAN analysis, being highly polymorphic in the Mediterranean while nearly fixed in the three oceans. The result obtained for the ss974292126+ ss974292127 SNP haplotype in the present study is indicative of a clear pattern of diversifying selection. From an adaptive point of view, this result depicts an environmental scenario in the Mediterranean different with respect to the environmental homogeneity of the three oceans. Unfortunately, no homology was found to known teleost's genes (Table S3.2). We also found 16 outlier SNPs with an anomalous homogeneity within the species. One likely scenario is balancing selection actively maintaining those SNPs in the gene pool of albacore. In fact, the inspection of the sequences surrounding the 18 outlier SNPs revealed 13 candidate genes involved in essential metabolic pathways. However, alternative explanations cannot be ruled out: such outliers could also reflect reduced variation at these loci if the minor allele was rare in all populations around the globe, or they could even be false positives. The fast growing genomic data base in the marine world will help to decipher these findings in the near future.

4.2. Population genetic structure

The genetic structure revealed in the present study is reliable since it was based on quite a large sample size, and on an extensive spatial and temporal distribution of samples. Moreover, POWSIM simulations showed that the 50 SNPs with the highest F_{ST} values together yielded a type II error rate (failure to detect a real difference) of 0% for divergences of F_{ST}= 0.002 or greater. In all, 4 spatial-temporally homogeneous populations were identified for the albacore species: Mediterranean Sea, Atlantic Ocean, Indian Ocean, and Pacific Ocean populations (Table 3.2). When STRUCTURE

analysis was repeated for each of the major clusters (Mediterranean, Pacific and Atlantic), no structure was detected within them. This result bears out the findings of Albaina *et al.* (2013). On the contrary, the genetic heterogeneity within the Mediterranean suggested by others using microsatellite markers (Davies *et al.* 2011; Montes *et al.* 2012), the observation of separate spawning grounds (Marano *et al.* 1999), or differences in isotopic composition (Goñi *et al.* 2011), was not detected using SNPs. Within the Mediterranean no significant correlation was found between genetic and geographic distances. Present findings also contradicted previously suggested heterogeneity for the Pacific Ocean using microsatellite markers (Takagi *et al.* 2001; Montes *et al.* 2012) or studying migrations, spawning areas and seasons as criteria (Lewis 1990). And lastly, the Atlantic Ocean was found to be homogeneous in terms of genetic structure, thus challenging earlier results and interpretations on the basis of blood groups (Arrizabalaga *et al.* 2004), microsatellites (Takagi *et al.* 2001; Davies *et al.* 2011) and migratory features (Arrizabalaga *et al.* 2004; Ortiz de Zárate & Cort 1998). In fact, albacore tuna tagging experiments are very scarce, specially in the South Atlantic, Indian Ocean and throughout the Mediterranean Sea. Thus, there is little information about their migratory behavior and the data available are not very informative about population structure and mixing. Moreover, unfortunately, there is little knowledge about albacore spawning areas and times for the different populations (Ueyanagi 1969), and this makes it difficult to get reference samples of known origin (e.g. larvae or young of the year) for genetic studies. Using samples that might represent transient migrants could, in principle, provide a misleading picture of population structure (e.g. suggesting homogeneity within the Mediterranean, where some structure might exist). This potential problem affects mostly at small scales and less at the scale of ocean basins and/or hemispheres. New knowledge about albacore spawning areas and seasons, as well as increased access to reference samples will allow to design more robust genetic experiments to reveal population structure and mixing at smaller scales.

With respect to the discrepancies between the present study and those using microsatellites, and in regard to the power of the markers, although an individual SNP show less power than do multi-allelic microsatellite loci (Haasl & Payseur 2011), 4–12 nuclear SNPs are expected to have the same power as a single microsatellite locus (Guichoux *et al.* 2011). Moreover, SNP markers have advantages over other markers: the use of single-tube multiplex assays with small PCR products (60–80 bp) could potentially produce better quality data more efficiently than would genotyping multiple microsatellites, and using SNP loci lies in a more representative sample of the entire genome and a reduced interlocus sampling variance (Morin *et al.* 2004).

In this study, neutral SNP variation (Table 3.2, Figs. 3.3, 3.4) and the SNP haplotype putatively under selection showed the Mediterranean group as the most differentiated from the rest of populations. Similarly, extremely different frequencies for *G6PD* locus and mtDNA D-Loop sequences between Mediterranean and Atlantic samples were described by Nakadate *et al*. (2005), and interpreted as indicative of a restricted gene flow. Our results agreed with this, since F_{ST} values between Mediterranean and the rest of populations are the highest, ranging between 0.033 and 0.051 (Table 2). The isolation of the Mediterranean population contrasted with the higher gene flow

that occurs between the three Oceans. In fact, when the three Oceans were analyzed together, correlation between genetic and geographic distances of the different sampling points were found ($r = 0.7549$; p-value < 0.0001; Figure S3.1b), as previously described for the Atlantic herring, another migratory pelagic marine fish (Ruzzante *et al.* 2006). Although there is great evidence that other species (such as Atlantic bluefin tuna and swordfish (*Xiphias gladius*) (Block *et al.* 2005; Alvarado Bremer *et al.* 2005; Aranda *et al.* 2013) migrate substantially across Strait of Gibraltar, migration is negligible for albacore (Arrizabalaga *et al.* 2004; Nakadate *et al.* 2005). Results obtained in the present study, together with those using different methodologies, such as genetic markers (Viñas *et al.* 2004; Nakadate *et al.* 2005; Davies *et al.* 2011; Montes *et al.* 2012; Albaina *et al.* 2013), growth parameters (Megalofonou 2000) and tagging experiments (Arrizabalaga *et al.* 2004) confirm the singularity of the Mediterranean albacore. It is difficult to evaluate whether this singularity is due only to current restricted gene flow, or it may reflect also the demographic history of Mediterranean albacore. According to Kettle *et al*. (2011), the Mediterranean would have served as a refugia for a range of marine species during the last glacial maximum (LGM). Under this latter hypothesis, Mediterranean population would be the result of one major founding event, and would have been isolated from all other populations for a long time. A similar scenario has been proposed for Atlantic herring in the Baltic Sea (Gaggiotti *et al.* 2009). In order to shed light on the controversial genetic relationship of the Indian albacore population with that of the Atlantic or Pacific, we analyzed an ample sample including 774 individuals from 12 Atlantic locations, 167 from three Pacific locations, and 136 individuals from four localities in the Western Indian Ocean. Results indicated that Indian samples appeared genetically closer to North Pacific ones, since the F_{ST} value between these populations was the only no significant comparison (Table 3.2). Our work thus confirms with a large sample of the Indian albacore population the results of Albaina *et al*. (2013), who analyzed, also with SNPs, 24 individuals. This sample was the same as that used in the study with microsatellites by Montes *et al*. (2012), although different results were obtained in both studies, since the analysis with 8 microsatellite markers indicated that the Indian albacore population was closer to the Atlantic than to the Pacific one. We think that in this case results may be biased due to the analysis of highly polymorphic markers in a small sample. In any case, the present study also detected that Indian albacore showed both Atlantic and Pacific components in STRUCTURE (Figure 3.4) and GENELAND analyses. That is, Cape of Good Hope did not represent a definitive barrier to gene flow, as it has been described by other authors (Chow & Ushiama 1995; Arrizabalaga *et al.* 2004; Montes *et al.* 2012).

4.3. Effective population size

Albacore is an overexploited species, whose biomass started decreasing due to overfishing 3 decades ago. Tuna stock assessments based on fishery data are highly uncertain (see Fromentin *et al.* 2014) and albacore is not an exception (Arrizabalaga *et al.* 2007; ICCAT 2014). Albacore is a species with seemingly large populations, however they could be more sensitive to genetic drift and inbreeding from intensive harvests than census sizes would suggest (Poulsen *et al.* 2006; Hare *et al.*

2011). In these cases, management requires the maintenance of a much larger census size than would typically be recommended on the basis of information about population dynamics (Hare *et al.* 2011). This is an assumed problem associated to overfished populations: that the high fishing pressure leads to genetic bottlenecks (Hauser *et al.* 2002; Hutchinson *et al.* 2003). If true, this could have serious implications for management procedures (Poulsen *et al.* 2006; Hare *et al.* 2011; Therkildsen *et al.* 2010). Therefore, estimating of N_e for sustainable management purposes is a good choice, because it integrates genetic effects with the life history of the species, allowing for predictions of a population's current and future viability (Hare *et al.* 2011). Our analyses on population genetic structure showed no statistically significant spatial or temporal fluctuations within each of the four defined populations. This result indicated that (1) migration had failed to alter allele frequencies at each region, and that (2) the effective population size in each region was large enough to prevent microdifferentiation processes driven by genetic drift. This latter hypothesis was supported for the North Atlantic Ocean population; similar *short-* and *long-term* N_e estimates for this population suggested that in spite of the fishing impact on biomass (N_c), genetic diversity remains high and, therefore, viability of the population has not been affected, this is, it has not suffered severe historical bottlenecks.

From a fishery management perspective, *short-term* N_e estimates could provide an approach for generating a fishery-independent indicator of population status. Temporal variations in such an indicator could serve as a prognostic marker of the genetic diversity of exploited albacore tunas and trigger specific well planned management responses to signs of reduced diversity (e.g. drastic reduction of fishing effort until genetic diversity is recovered). Management must often default to apparently simple rules-of-thumb, such as the 50/500 criteria for maintenance of genetic diversity; this means that a *short-term* $N_e \geq 50$ is required to avoid the damaging effects of inbreeding, and a *short-term* $N_e \geq 500$ is necessary to avoid extinctions due to the inability to evolve to cope with environmental change. Taking this rule into account, we have demonstrated that albacore population size in the North Atlantic Ocean is high enough for dealing with both, inbreeding effects and adaptation capabilities. But for management purposes, N_e estimates might be more adequate to better understand how ecological factors reduce or increase the N_e/N_c ratio. With this regard, theory suggests that N_e/N_c ratios in the wild should be above 0.1 (Nunney & Campbell 1993; Frankham 1995; Vucetich *et al.* 1997), and empirical evidence for several wild populations of different non marine species is consistent with this prediction, showing N_e/N_c ratios ranging from 0.10 to 0.14 (Frankham 1995; Palstra & Ruzzante 2008). In the North Atlantic Ocean, the effective population size was three orders of magnitude lower than the adult census size (Table S3.3). These figures are within the range documented for other fish species, such as *Sciaenops ocellatus* (Turner *et al.* 2002), *Pagrus auratus* (Hauser *et al.* 2002) and *Sebastes crameri* (Gomez-Uchida & Banks 2006). A low N_e/N_c ratio could be explained by variance in albacore survival due to high larval and pre-recruit mortality (Whitlock & Barton 1997; Nunney 1999), indicating that few mature adults contribute to each generation. It has been questioned the appropriateness of estimating N_e from temporal data in species with high effective population sizes (Poulsen *et al.* 2006), and whether N_e/N_c ratios reflect

the true dynamics of biological systems (Flowers *et al.* 2002; Therkildsen *et al.* 2010). Nevertheless, it is important to obtain a better understanding of how vulnerable fish populations are to loss of genetic variation and in that respect, the data presented here on temporal stability at neutral markers will serve as an important baseline for future evaluations of N_e/N_c and for monitoring N_e in albacore. In conclusion, N_e estimate, as a fishery-independent index of abundance, provides a valuable complementary tool for monitoring the status of fish populations in order to implement more sustainable management actions.

Acknowledgments

The authors are indebted to technical and human support provided by General Research Services SGIker in the University of the Basque Country (UPV/EHU). We are grateful to the following institutes who provided samples: South African Institute for Aquatic Biodiversity (SAIAB, South Africa), Galway-Mayo Institute of Technology (GMIT, Ireland), Instituto Español de Oceanografía (IEO, Spain). We also are grateful to E. Jimenez, M.A. Pardo, I. Fraile, I. Arregi, J. Lopez, J. Filmarter, J. Areso, A. Delgado, D. Brophy, K.Schaefer, D. Fuller and V. Allain for sampling assistance. This research was supported by the projects ATM2010Hegaluze (351BI20090047) and TUNASNIP (SPE10UN92) funded by the Basque Government. Urtzi Laconcha's work was supported by a PhD grant by the Fundación Centros Tecnológicos Iñaki Goenaga. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. We express our gratitude to ICCAT for funding the identification of ABFT SNPs (GBYP PHASE2 project), and to GBYP consortium members for providing them for cross-amplification. This manuscript was improved thanks to the comments of two anonymous reviewers.

Supporting Information

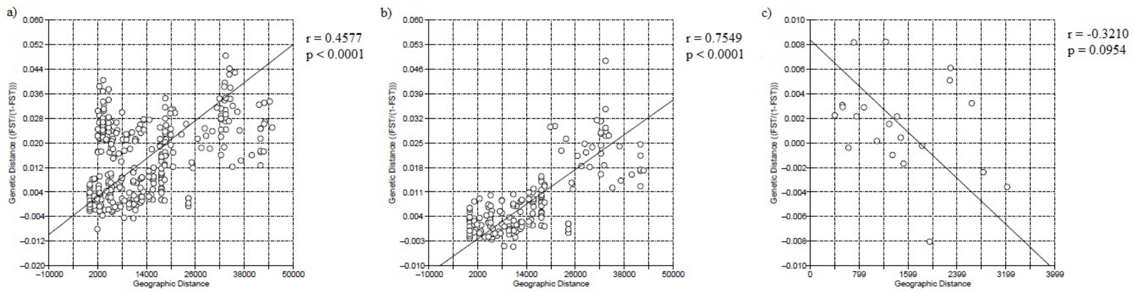


Figure S1. Isolation by distance. a) Regression of pairwise geographic distance and genetic similarity for the 26 locations. b) Similar analysis using 19 locations from the Atlantic, Indian and Pacific Oceans and c) considering only the 7 Mediterranean samples.

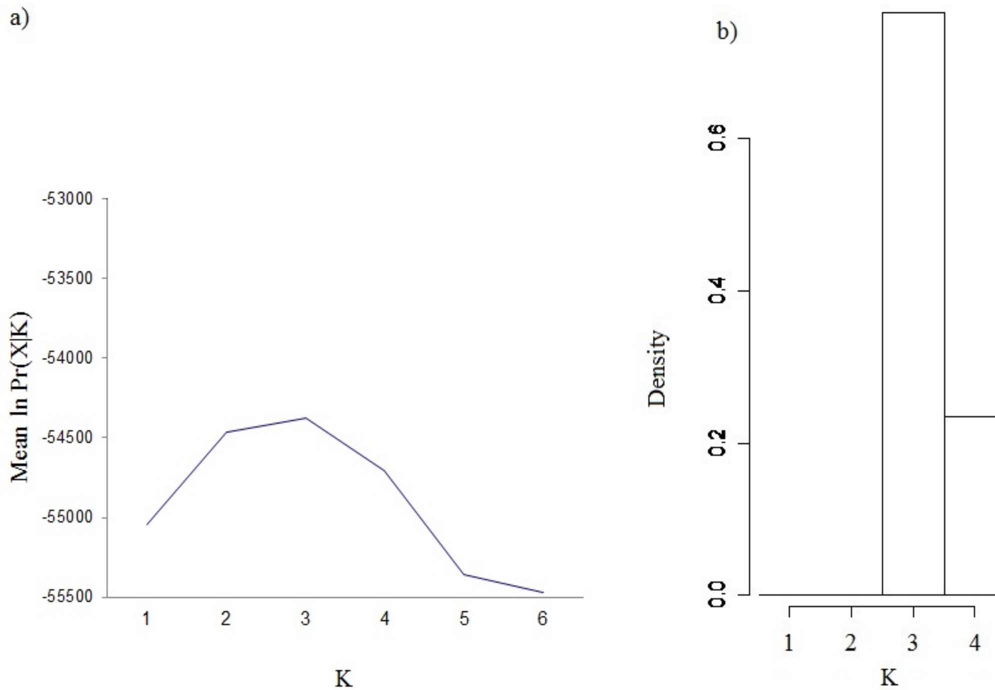


Figure S2. Estimated number of populations from STRUCTURE (a) and GENELAND (b) analyses. (a) Mean probabilities of the data [LnPr(X)|*K*] over 10 STRUCTURE replicated runs plotted as a function of putative number of clusters (*K*). (b) Posterior density distribution of the number of estimated clusters.

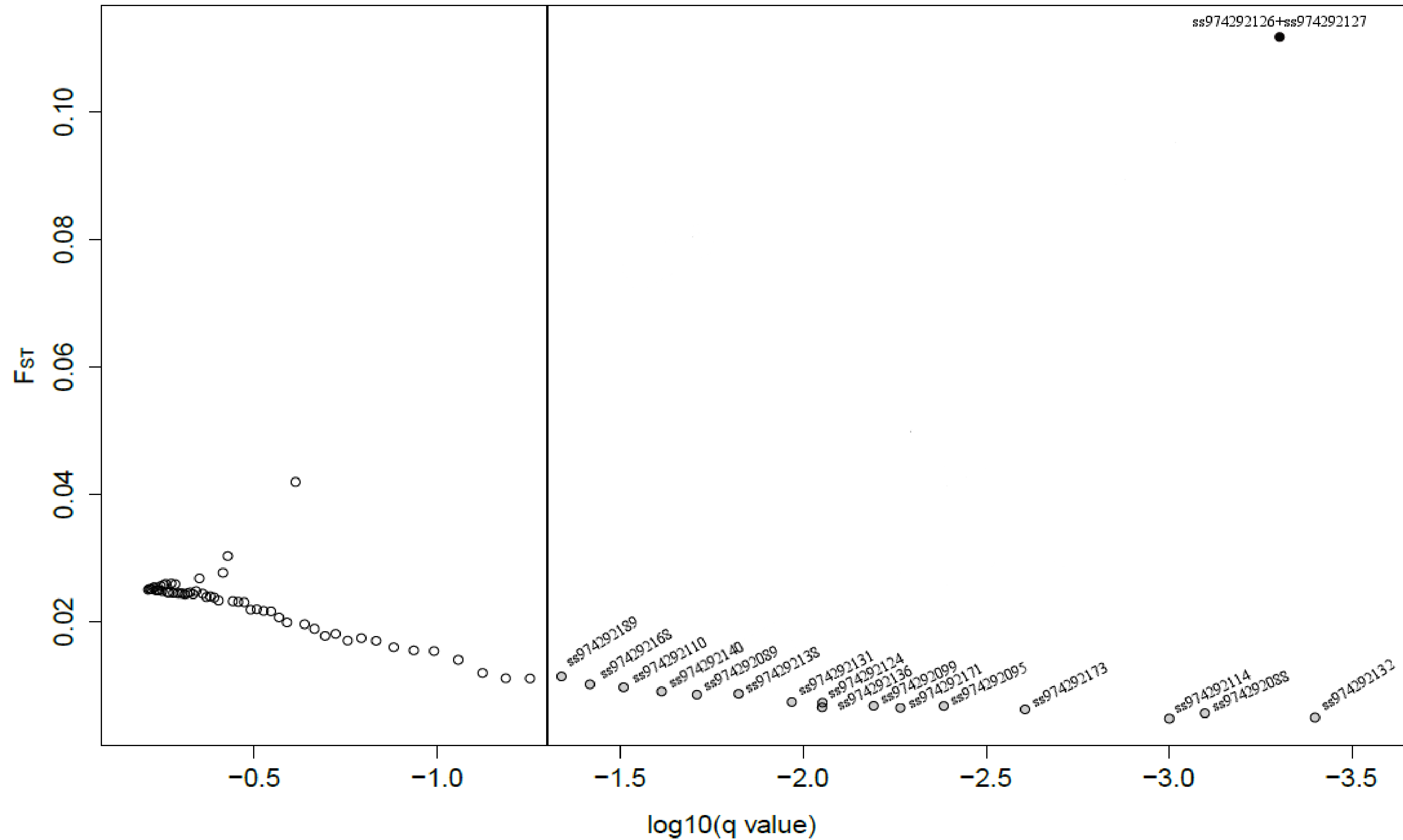


Figure S3. BAYESCAN output plot for outlier identification. Graphical representation of the markers based on F_{ST} values (y axis) against log (*q*-value) (x axis). Candidate markers under selection are those with a *q*-value less than 0.1, represented at the right of the vertical black line. Grey circles represent candidate markers for balancing selection and black circles represent candidate marker for divergent selection; empty circles represent putatively neutral loci.

Table S1. Characteristics of the 117 genotyped SNPs in 1,331 *Thunnus alalunga* samples analyzed in this study. (*) SNPs obtained from Albaina *et al*. [34].

SNP #	NCBI Assay ID	Call rate	Alleles	MAF	HWE
1	ss974292086	90%	A/G	0.472	NS
2	ss974292087	98%	G/G		-
3	ss974292088	98%	A/G	0.095	NS
4	ss974292089	96%	G/T	0.392	NS
5	ss974292090	96%	G/A	0.163	NS
6	ss974292091	9%	A/A		-
7	ss974292092	96%	G/A	0.038	NS
8	ss974292093	98%	G/A	0.004	NS
9	ss974292094	95%	A/G	0.297	NS
10	ss974292095	97%	C/A	0.107	NS
11	ss974292096	99%	C/C		-
12	ss974292097	98%	T/C	0.017	NS
13	ss974292098	72%	C/C		-
14	ss974292099	95%	A/G	0.245	NS
15	ss974292100	95%	T/C	0.032	NS
16	ss974292101	81%	C/G	0.241	SIGN
17	ss974292102	97%	T/C	0.070	NS
18	ss974292103	98%	C/T	0.076	NS
19	ss974292104	98%	A/A		-
20	ss974292105	98%	G/A	0.151	NS
21	ss974292106	93%	G/A	0.263	NS
22	ss974292107	96%	T/C	0.079	NS
23	ss974292108	97%	G/A	0.012	SIGN
24	ss974292109	94%	C/A	0.030	NS
25	ss974292110	97%	G/A	0.251	NS
26	ss974292111	95%	T/C	0.115	NS
27	ss974292459	96%	G/A	0.088	SIGN
28	ss974292460	98%	T/T		-
29	ss974292114	96%	G/A	0.186	NS
30	ss974292115	97%	A/C	0.005	NS
31	ss974292116	98%	G/A	0.017	NS
32	ss974292117	84%	T/C	0.231	NS
33	ss974292118	91%	G/A	0.243	SIGN
34	ss974292119	98%	A/G	0.007	NS
35	ss974292120	96%	G/A	0.032	NS
36	ss974292121	98%	G/G		-
37	ss974292122	97%	T/T		-
38	ss974292123	97%	A/G	0.035	NS
39	ss974292124	97%	G/A	0.075	NS
40	ss974292125	97%	T/T		-
41	ss974292126	95%	T/C	0.030	NS
42	ss974292127	94%	A/C	0.033	NS
43	ss974292128	97%	G/A	0.022	NS

44	ss974292129	55%	T/G		-
45	ss974292130	97%	T/C	0.061	NS
46	ss974292131	93%	C/A	0.277	NS
47	ss974292132	98%	T/C	0.149	NS
48	ss974292133	92%	A/G	0.28	SIGN
49	ss974292134	97%	G/A	0.456	NS
50	ss974292135	98%	G/A	0.018	NS
51	ss974292136	94%	A/G	0.104	NS
52	ss974292137	87%	G/A	0.387	NS
53	ss974292138	97%	C/A	0.062	NS
54	ss974292139	98%	A/G	0.004	NS
55	ss974292140	93%	A/G	0.400	NS
56	ss974292141	90%	A/T	0.297	SIGN
57	ss974292142	98%	G/A	0.050	NS
58	ss974292143	91%	A/G		-
59	ss974292144	89%	T/C	0.153	SIGN
60	ss974292145	91%	A/T	0.216	SIGN
61	ss974292146	98%	G/G		-
62	ss974292147	97%	G/A	0.004	NS
63	ss974292148	95%	A/G	0.488	SIGN
64	ss974292149	93%	T/A	0.470	SIGN
65	ss974292150	98%	A/G	0.003	NS
66	ss974292151	98%	A/A		-
67	ss974292152	95%	G/A	0.048	NS
68	ss974292153	94%	G/A	0.093	NS
69	ss974292154	90%	T/C	0.107	NS
70	ss974292155	97%	G/T	0.017	SIGN
71	ss974292156	94%	G/G		-
72	ss974292157	95%	T/C	0.009	NS
73	ss974292158	98%	G/A	0.006	NS
74	ss974292159	92%	A/G	0.460	NS
75	ss974292160	97%	A/T	0.123	NS
76	ss974292161	97%	G/A	0.025	NS
77	ss974292162	95%	A/G	0.199	NS
78	ss974292163	94%	A/T	0.492	SIGN
79	ss974292164	98%	G/A	0.002	NS
80	ss974292165	98%	T/C	0.042	NS
81	ss974292166	82%	G/A	0.120	SIGN
82	ss974292167	85%	T/C	0.418	SIGN
83	ss974292168	89%	A/T	0.493	NS
84	ss974292169	98%	A/G	0.009	NS
85	ss974292170	97%	G/G		-
86	ss974292171	98%	T/C	0.034	NS
87	ss974292172	95%	G/A	0.197	NS
88	ss974292173	97%	G/A	0.057	NS
89	ss974292174	96%	A/A		-
90	ss974292175	97%	G/A	0.063	NS
91	ss974292176	84%	G/A	0.400	SIGN

92	ss974292177	98%	G/T	0.011	NS
93	ss974292178	96%	A/A		-
94	ss974292179	96%	G/A	0.055	NS
95	ss974292180	98%	G/A	0.059	NS
96	ss974292181	96%	G/A	0.049	NS
97	ss974292182	93%	T/C	0.470	SIGN
98	ss974292183	97%	G/G		-
99	ss974292184	86%	G/A	0.135	NS
100	ss974292185	98%	G/A	0.005	NS
101	ss974292186	95%	C/C		-
102	ss974292187	98%	G/A	0.025	NS
103	ss974292188	95%	A/C	0.096	NS
104	ss974292189	94%	G/T	0.207	NS
105	ss974292190	91%	C/C		-
106	ss974292191	97%	C/T	0.019	NS
107	ss974292192	92%	G/A	0.367	SIGN
108	ss974292193	96%	G/A	0.089	NS
109	ss974292194	97%	G/T	0.005	NS
110	ss974292195	93%	C/G	0.331	SIGN
111	ss974292196	97%	T/C	0.136	SIGN
112	ss974292197	98%	A/C	0.004	NS
113	ss974292198	97%	G/G		-
114	ss974292199	95%	G/A	0.061	NS
115	ss974292200	1%	G/A		-
116	rs366516796*	92%	A/T	0.264	SIGN
117	rs193631170*	95%	A/C	0.253	NS

Table S2. BLASTn results and possible molecular function of eighteen sequences of candidate SNPs under divergent or balancing selection.

NCBI Assay ID	Type of selection	Gene	Id	Molecular function
ss974292088	Balancing	Ribosomal protein L18a	RPL18A	RNA binding
ss974292089	Balancing	Transmembrane 9 superfamily member 3	TM9SF3	Integral component of membrane
ss974292095	Balancing	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 9	NDUFA9	NADH ubiquinone oxidoreductase
ss974292099	Balancing	Probable ribosome biogenesis protein RLP24	RLP24	ATPase activator
ss974292110	Balancing	60S ribosomal protein L27	RPL27	Structural constituent of ribosome
ss974292114	Balancing	60S ribosomal protein L31	RPL31	RNA binding, Structural constituent of ribosome
ss974292124	Balancing	Proline-rich nuclear receptor coactivator 2	PNRC2	regulation of transcription, DNA-templated
ss974292126	Divergent	No Result	-	-
ss974292127	Divergent	No Result	-	-
ss974292131	Balancing	Betaine-homocysteine methyltransferase	BHMT	Methyltransferase activity, zinc ion binding
ss974292132	Balancing	A disintegrin and metalloproteinase with thrombospondin motifs 2	ADAMTS2	Metalloendopeptidase activity, zinc ion binding
ss974292136	Balancing	Vacuolar protein sorting 28	VPS28	Component of the ESCRT-I complex, a regulator of vesicular trafficking process
ss974292138	Balancing	UPF0461 protein C5orf24 homolog	UPF0461	Methyltransferase activity, metal ion binding
ss974292140	Balancing	No Result	-	-
ss974292168	Balancing	No Result	-	-
ss974292171	Balancing	Ribosomal protein L3	RPL3	Structural constituent of ribosome
ss974292173	Balancing	fructose-1,6-bisphosphatase isozyme 2-like, mRNA	PFKFB2	Synthesis and degradation of fructose 2,6-bisphosphate
ss974292189	Balancing	No Result	-	-

Table S3. *Short-term* N_e estimate for the North Atlantic from 1986 to 2009 based on the temporal method. Generation time ($\hat{G}$), F_s values, harmonic means of effective population size ($\check{N}_e$), spawning census population size ($\check{N}_c$), and $\check{N}_e/\check{N}_c$ ratio.

	$\hat{G}$	Fs	F's	Ňe	Ňc	Ňe/Ňc
C86	4.5	0.038	0.007	14,040	3,200,270	4.39×10^{-3}
C87	4.3	0.046	0.012	8,388	2,397,944	3.50×10^{-3}
C05	4.1	0.028	0.005	18,147	2,796,804	6.49×10^{-3}
C06	4.6	0.017	0.005	10,232	2,238,867	4.57×10^{-3}
C07	5.0	0.026	0.016	5,466	2,090,050	2.62×10^{-3}
C08	5.3	0.017	0.004	23,330	2,373,467	9.83×10^{-3}

Table S4. *Long-term* N_e estimate for the North Atlantic based on the coalescent method. Number of samples representing each cohort (N), spawning census population size (N_c) and N_e/N_c ratio values are listed.

	N	Ne	Nc	Ne/Nc
C86	30	20,304	3,200,270	6.34×10^{-3}
C87	34	16,341	2,397,944	6.81×10^{-3}
C05	75	18,862	2,796,804	6.74×10^{-3}
C06	68	14,839	2,238,867	6.63×10^{-3}
C07	70	15,696	2,090,050	7.51×10^{-3}
C08	136	13,897	2,373,467	5.86×10^{-3}
C09	53	17,162	3,006,973	5.71×10^{-3}

Capítulo IV

Phylogeographic patterns in albacore and phylogenetic relationships between *Thunnus* species inferred from direct mtDNA analysis

ESTUDIO III

Phylogeographic patterns in albacore and phylogenetic relationships between *Thunnus* species inferred from direct mtDNA analysis

Urtzi Laconcha, Mikel Iriondo, Haritz Arrizabalaga, Carmen Manzano, Pablo Markaide, Haritz Jimenez, Maria Jesus Sanz, Igor Velado, Andone Estonba

Abstract

In this study, sequences from five mitochondrial regions (CR, 12S, 16S, Cyt *b* and COI), plus nine mitochondrial SNPs were analyzed in samples from all currently defined stocks of albacore (*Thunnus alalunga*). The phylogeographic analyses revealed the presence of two haplogroups that emerged 1 My ago. Haplogroup frequencies were correlated with geographic distances except for the Mediterranean, which shows an unusual level of isolation. In addition, significant differences between the Mediterranean Sea, Atlantic Ocean, Indian Ocean and Pacific Ocean were detected by 9 mitochondrial SNPs (Φ_{ST} ranging from 0.038 to 0.199 between stocks), in line with the genetic structure previously shown through nuclear SNPs. These results display partial historical isolation between the Atlantic and Indo-Pacific Oceans, besides the one in the Mediterranean. Regarding the 8 species included in the *Thunnus* genus, the five mitochondrial regions above and mitochondrial genome-wide sequences don't support actual subdivision of the *Thunnus* genus into two subgenus, *Neothunnus* and *Thunnus.* Furthermore*,* these analyses indicate clearly that albacore constitutes the most ancient group of the genus (bifurcated 4.5 My ago) followed by bigeye tuna (*T. obesus*) (4.0 My). Interestingly, between 2.7 and 2.1 My, a cladogenetic process would have occurred during which all remaining species would emerge during this period, including both lineages of bigeye tuna and longtail tuna (*T. tonggol*). Results also show evidence for historical and/or recent hybridization events between species of the genus *Thunnus*. In this regard, we propose an ancient introgression of albacore mtDNA in the common ancestor of Pacific and Atlantic bluefin which lived originally in the Atlantic Ocean, and around 0.8 My ago, after a founder effect, inhabited the Pacific Ocean and originated the Pacific bluefin tuna.

Keywords: *Thunnus alalunga*, *Thunnus* genus, phylogeny**,** mtDNA, mitochondrial single nucleotide polymorphism (SNP)

1. Introduction

Albacore (*Thunnus alalunga*, Bonnaterre 1788) is the species with the widest distribution in the genus *Thunnus*, since this species is the only one which spans the Pacific, Indian and Atlantic oceans, as well as the Mediterranean Sea, extending from 50-55ºN to 40-45ºS (Arrizabalaga *et al.* 2014). This cosmopolitan distribution in tropical and template waters reflects the adaptation to different oceanographic conditions with high temperature variations. Albacore is present at around 13.5ºC and 25.2ºC and tolerates temperatures as low as 9.5ºC during short periods. It is the smallest among the eight species in the genus *Thunnus* and, the characteristics of the species (longevity, unknown spawning areas and short spawning seasons) added to the high economic value, make albacore especially vulnerable to overfishing (Collette *et al.* 2011). This vulnerability means that any management actions taken must be effective in the long run. Therefore, multidisciplinary information resulting from the different scientific methodologies that contribute towards raised awareness of exploited stocks (population structure, migrations, genetic variability etc.) must be reliable, updated and useful for managers when deciding how to manage the stocks. Inadequate management of fisheries, that is, removed from the definition of management units in accordance with the current population structure of the species, could lead to the reduction of the population and consequently, a reduction in the productivity (Hauser & Carvalho, 2008) which may even bring about the collapse of a few fisheries (Fu & Fanning, 2004). Furthermore, overexploited populations could suffer genetic variability loss (Hauser *et al.* 2002), which could lead to a loss of the adaptive response capacity of the species faced with a continually changing environment.

Regarding the population structure of albacore, results from previous worldwide studies using nuclear SNP markers indicate the existence of 4 different populations: one in the Mediterranean Sea and one in each Ocean (Atlantic, Pacific, and Indian Oceans) (Albaina *et al.* 2013; Laconcha *et al.* 2015a). However, results from different markers do not match this drawing. Thus, according to Montes *et al.* (2012) the use of microsatellites in a worldwide sample did not differentiate between albacores in the Atlantic or Indian Oceans. Using different markers, such as mtDNA sequences, Graves & Dizon (1989) and Viñas *et al.* (1999) failed to find genetic differences between Atlantic and Pacific albacore and Atlantic and Mediterranean albacore, respectively. However, Chow and Ushiama (1995), found differences between Atlantic and Pacific using RFLP markers of the mtDNA ATPase gene, and Nakadate *et al.* (2005), using sequences of the mtDNA D-loop region, differentiated between Atlantic and Mediterranean albacores. These last two studies are in concordance with results from nuclear SNPs.

Furthermore, relationships among closely related *Thunnus* taxa remain unresolved, since the classic subdivision (Collette 1978), based on morphological and ecological differences, is not reflected, overall, in genetic studies. Morphoecologic division of the 8 species from genus *Thunnus* differentiates between two subgenus (Collette 1978): *Thunnus* (*T. alalunga*, *T. obesus T.*

thynnus, *T. orientalis*, *T. maccoyii*) and *Neothunnus* (*T. atlanticus*, *T. tonggol* and *T. albacares*). The separation of the subgenus is conditioned by the presence (subgenus *Thunnus*) or absence (subgenus *Neothunnus*) of a central heat exchanger and several other anatomical characteristics of the red muscle, the liver, the cutaneous artery and the axial skeleton (Nakamura 1965; Gibbs & Collette 1967; Collette 1978; Chow & Kishino 1995). Species of subgenus *Thunnus* are considered temperate species whereas species of subgenus *Neothunnus* are known as tropical species. Although Alvarado Bremer *et al.* (1997) after analysing mitochondrial sequences from Control Region (CR) obtained coincident results with the classification of *Thunnus* species into two subgenus, most of the studies did not find a genetic setting with the morphoecological division by Collette (1978). Thus, Chow & Inoue (1993) -based on RFLP markers of cytochrome b (Cyt *b*), 12S and ATPase mitochondrial genes - and Chow & Kishino (1995) –by analysing Cyt *b* and ATPase genes- situated the Atlantic bluefin tuna species *(Thunnus thynnus*) and Southern bluefin tuna (*Thunnus maccoyii*) phylogenetically more proximate to the subgenus *Neothunnus* than to the subgenus *Thunnus*. It should also be pointed out that both of these studies detected high phylogenetic proximity between Pacific bluefin tuna (*Thunnus orientalis*) and albacore, whereas the analysis carried out with nuclear sequence ITS1 (Chow & Kishino 1995; Chow *et al*. 2006) genetically related Pacific bluefin tuna with Atlantic bluefin tuna and not with albacore. Tsen *et al.* (2012) also concluded that the phylogeny of *Thunnus* species does not fit into the current two subgenera classification pattern analysing complete Cyt *b* sequences from all eight *Thunnus* species. In summary, the existence of the subgenus *Thunnus* and *Neothunnus* based on molecular data and phylogenetic relationships between the species of the genus is still an unresolved issue. Most of studies focused on the analysis of one or two genes to determine the relationships between 8 species of the genus. The approach of the present study, in which an ample sample of *Thunnus* individuals has been analysed for multiple sizes of mitochondrial sequences (complete mitochondrial sequences, various concatenated genes and a single gene), could shed light on the phylogeny of the genus, and temporally reference speciation events of the genus *Thunnus*.

In summary, the goal of this study is two-fold. On the one hand, to test the population genetic structure observed on albacore with nuclear SNPs, analysing for the first time a high number of mitochondrial sequences and SNPs in a worldwide albacore sample. On the other hand, this study aims to shed light on the controversial phylogenetic relationship among *Thunnus* species. For these purposes, 38 albacores were sequenced for 4 mitochondrial genes, 160 individuals were sequenced for Control Region (CR), and 1,254 individuals were genotyped for 10 mitochondrial SNPs located in Cytochrome *b* (CYT *b*), cytochrome oxidase subunit I (COI), 12S and 16S mitochondrial genes. Furthermore, data collection has been done obtaining over 2000 sequences of CR from the 8 species of genus *Thunnus*.

2. Material and methods

2.1. Samples

The sample consisted of 1,254 albacore (*Thunnus alalunga*, Bonnaterre 1788) individuals from the 6 stocks defined for the species (Mediterranean Sea, North Atlantic Ocean, South Atlantic Ocean, Indian Ocean, North Pacific Ocean and South Pacific Ocean), which was compiled by AZTI Marine Research. Individuals were mainly sampled between 2008 and 2012, with some of them sampled between 1988 and 1989. Most samples were muscle, fin or heart tissue samples, either stored frozen at -20ºC or preserved in 96% ethanol at 4 ºC. Additionally, spine cuts mounted in Eukitt (O. Kindler GmbH) and dried and stained blood samples were collected (Table 4.1). DNA of muscle, fin and heart tissue samples was extracted using NucleoSpin 96 Tissue Kit (Macherey-Nagel). Spine and blood samples were first immersed in Xylol, and spine samples were afterwards manually crushed; DNA from these samples was extracted through a specific membrane using QIAmp DNA Investigator Kit (Qiagen). DNA from all samples was quantified using both NanoDrop 1000 spectrophotometer (Thermo Fisher Scientific Inc., Waltham, MA) and Qubit 2.0 (Invitrogen, Life Technologies) fluorometer. All the DNA samples were stored at -20 ºC for further analyses.

Table 4.1. Stock, date of capture, code, number of albacore individuals (N) for each approach: 5 mitochondrial regions (Cyt *b*, COI, 12S, 16S and CR; 5,112 bp), 344bp CR sequences (CR), and 10 SNPs, and sample type (T = Muscle, fin or heart tissue; B = Blood and S = Spine).

Stock	Date	Code	N			Sample type
			5 regions	CR	10 SNP	
Mediterranean Sea	1999-2011	MED	6	27	235	T and B
North Atlantic Ocean	1988-2012	NAT	11	27	492	T and S
South Atlantic Ocean	1999-2012	SAT	7	29	232	T and B
Indian Ocean	2008-2012	IND	6	20	131	T
North Pacific Ocean	2008	NPA	3	29	81	T
South Pacific Ocean	2003-2008	SPA	5	28	83	T
Total			38	160	1,254	

2.2. Sequencing and SNP genotyping

Phylogeographic analysis of the albacore was carried out from three approaches (Table 4.1). Firstly, 38 individuals (6 from Mediterranean, 11 North Atlantic, 7 South Atlantic, 6 Indian, 3 North Pacific and 5 from South Pacific) were sequenced for a total of 5,112 bp, comprising 5

mitochondrial regions: CR, Cyt *b*, COI, 12S and 16S. Secondly, 122 additional individuals (21 from Mediterranean, 16 North Atlantic, 22 South Atlantic, 14 Indian, 26 North Pacific and 23 from South Pacific) were sequenced only for 344 bp in CR. Together with those sequences from the first approach, a total of 160 CR sequences from diverse origins were analysed to infer albacore's phylogeography. Finally, 10 single nucleotide polymorphisms (SNP) were genotyped in 1,254 individuals. SNPs were located in 4 mitochondrial regions sequenced in the first approach: 1 SNP of the gen 12S (ss947846612), 2 SNPs of the gen 16S (ss947846613, ss947846614), 2 of the COI (ss947846615, ss947846616) and 5 of the gen Cyt *b* (ss947846617, ss947846618, ss947846619, ss947846620 and ss947846630) sequence data. The CR region turned out to be unsuitable because of its high variability.

The sequencing was carried out at the Sequencing and Genotyping Unit of the General Research Services of the University of the Basque Country (SGIker), using ABI PRISM BIGDYE v3.1 Terminator Sequencing Reaction (Life Technologies) in an ABI PRISM 3130xl Genetic Analyzer (Life Technologies). Sequences were processed using the SeqScape v3.0 (Life Technologies) software. Finally, the sequences were edited using BioEdit 7.0.1 (Hall 1999) and aligned with MUSCLE (Edgar, 2004). The genotyping of selected SNPs was performed using TaqMan OpenArray technology (Life Technologies) at SGIker. Alleles for each SNP were assigned using TaqMan OpenArray technology (Life Technologies). In order to ensure the genotyping quality, SNPs needed to have a call rate higher than 80% and clear genotyping clusters.

2.2.1. Statistical Analysis

Haplotypic diversity (h), number of haplotypes (nh), SNP number, and nucleotide diversity (π) were estimated with DNAsp v.5 (Rozas *et al.* 2003) and MEGA6 (Tamura *et al.* 2004) softwares. Tajima's D (Tajima 1989) and Fu's FS (Fu 1997) measures of neutrality were estimated with DNAsp v.5 (Rozas *et al.* 2003).

The geographical structure of the albacore was tested using Arlequin 3.5 based on the matrix of Tamura-Nei distances. Pairwise Φ_{ST} was estimated for all pair of regions, using the 160 CR sequences, and 1,254 genotyped individuals for the SNPs. The significance level of each haplotypic correlation examined was tested using 1000 permutations.

For the phylogenetic analysis, we obtained sequences for the 5 regions analysed in the present study (Cyt *b*, COI, 12S, 16S and CR) from GenBank for a total of 55 individuals of 7 *Thunnus* species (all except *T. atlanticus*) and the complete mitochondrial sequences from these 16 individuals, as well as all CR sequences available for the 8 species of the genus. In this case, once all sequences were aligned, and after eliminating those with uncertainties, we obtained a final set composed of 2,113 CR sequences with a length of 344 bp (Table S4.1). Sequences of skipjack

tuna (*Katsuwonus pelamis*), bullet tuna (*Auxis rochei*) and black skipjack (*Euthynnus lineatus*) were also included in the phylogenetic analyses as outgroups.

The best-fit model of evolution was identified for each mitochondrial DNA (mtDNA) data set, using a hierarchical series of likelihood ratio tests as implemented in MODELTEST 3.06 (Posada & Crandall 1998). The best fit was obtained with the Tamura-Nei (1993) model. This substitution model was used to generate maximum likelihood (ML) and neighbour-joining (NJ) trees, using RAxML (Berger & Stamatakis, 2011) and MEGA6 (Tamura *et al.* 2004) softwares, respectively. Additionally, median-joining networks (Bandelt *et al.* 1999) were constructed using the software NETWORK 4.6.1.1 (Fluxus Technology Ltd.). Genetic distances were calculated for CR sequences (344 bp) between albacore haplogroups and also between *Thunnus* species using MEGA6 software, following the Tamura-Nei (1993) model and 1000 replicates of bootstrap. The "Pairwise deletion" option was performed when comparing species and haplogroups to each other, in order to minimize overestimates of distance. Inter-group distances were corrected taking into account intra-group distances. The same estimation was carried out with complete mitochondrial sequences of the species available in the GenBank database. Speciation and cladogenesis event datation was based on CR sequences of 344pb, implementing strict molecular clocks (Faber & Stepien 1997; Stepien & Faber 1998; Munday *et al*. 2004). Two estimates of divergent rates were used: 3.6% proposed for fish CR (Donaldson & Wilson 1999), and 11%, proposed for CR of Atlantic bluefin tuna and swordfish (*Xiphias gladius*) (Alvarado Bremer *et al*. 2005).

3. Results

3.1. Albacore

Analysis of 38 albacore individuals based on the 5 mitochondrial regions —Cyt *b*, COI, 12S, 16S and CR— detected 123 polymorphic sites (60 parsimony-informative sites), which formed 37 haplotypes, one haplotype for each individual, except for one coincident. Analysis of 160 individuals for 344 bp CR sequences reflected 177 polymorphic sites (139 parsimony-informative sites), which formed 145 haplotypes. The latter analysis also represents the extreme mitochondrial haplotypic variability. Thus, the measures of variability reflected values of haplotypic diversity (h) between 0.967-1.000, whereas nucleotide diversity (π) was between 0.003 and 0.057 (Table 4.2).

Table 4.2. Number of haplotypes (nh), haplotype diversity (h), nucleotide diversity (π), and percentage of individuals belonging to A or B haplogroup in 5 mitochondrial regions (5,112 bp), CR (344 bp) and 7 SNPs. Albacore stock abbreviations as defined in Table 4.1.

	5 mitochondrial regions concatenated			**control region**					**selected SNPs**			
	nh	**h**	**π**	**nh**	**h**	**π**	**A%**	**B%**	**nh**	**h**	**A%**	**B%**
MED	5	0.933	0.003	16	0.812	0.025	96.3	3.7	8	0.730	91.1	8.94
NAT	11	1.000	0.003	26	0.997	0.04	88.9	11.1	11	0.725	85.6	14.4
SAT	7	1.000	0.003	28	0.998	0.044	82.8	17.2	9	0.728	87.1	12.9
IND	6	1.000	0.004	20	1.000	0.057	65.0	35.0	8	0.840	68.7	31.3
NPA	3	1.000	0.005	29	1.000	0.057	62.1	37.9	7	0.774	63.0	37.0
SPA	5	1.000	0.004	26	0.995	0.056	60.7	39.3	8	0.790	62.7	37.3
Mean	6	0.989	0.004	24	0.967	0.047	76.0	24.0	9	0.765	76.4	23.6

Concerning mitochondrial lineages, two haplogroups were observed in the albacore. Median-joining network based on both sequences of 5 mitochondrial regions in 38 albacores and 344 bp CR sequences in 160 individuals clearly show two groups of haplotypes (Fig. 4.1; Fig. S4.1). The haplogroup with more haplotypes was defined as haplogroup A, and the one with less number of haplotypes as haplogroup B.

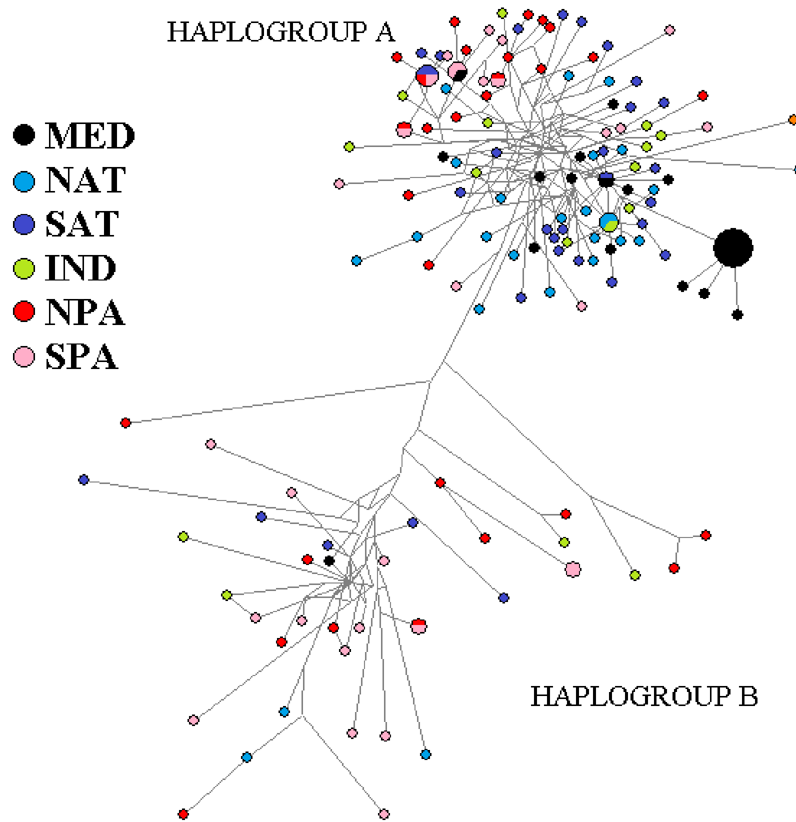


Figure 4.1. Median-joining network of the haplotypes of the 344bp CR sequences in 160 albacore individuals.

Regarding the analysis of 10 SNPs on 1,254 albacore individuals, the SNP ss947846618 failed at the genotyping step. Additionally, for ulterior analysis ss947846614 and ss947846620 were discarded because their high level of polymorphism produced too much noise in haplotype analysis. In this way, 7SNPs were finally selected: 6 SNPs (ss947846612, ss947846613, ss947846615, ss947846617, ss947846619, and ss947846630) defining internal variability between haplotypes, and 1 SNP (ss947846616) that differentiated two previously-defined haplogroups. These 7 SNPs created 13 haplotypes for the 1,254 albacore individuals (Table 4.3). SNP haplotype diversity varies among stocks, with values between h=0.725±0.024 in North Atlantic and h=0.840±0.011 in Indian Ocean (Table 4.2). Tajima's D and Fu's FS measures did not differ significantly from 0 ($D = 1.46$; $p > 0.10$ and $FS = -0.56$; $p > 0.10$), which indicates the neutrality of the loci.

Table 4.3. Mitochondrial SNP haplotypes and haplogoups in albacore. For each haplotype the alleles of the 7 SNPs (ss947846612, ss947846613, ss947846615, ss947846616, ss947846617, ss947846619 and ss947846630) are shown (represented with the latest three numbers in the table). (*) denotes the SNP that differentiated A and B haplogroups. Reference individuals are those assigned to A or B haplogroups, using 5 regions (5,112 bp) and 344 bp CR sequences.

Haplotype	612	613	615	616*	617	619	630	Reference individuals 5,112bp	Reference individuals CR	Haplogroup
H1	G	A	G	A	C	T	A	4	30	A
H2	G	A	A	G	G	C	A	2	17	B
H3	G	A	G	G	G	C	A	1	8	B
H4	G	A	G	A	G	T	A	8	29	A
H5	G	A	G	A	G	T	C	5	24	A
H6	G	A	G	G	G	T	A	1	1	B
H7	G	G	G	A	G	T	A	0	7	A
H8	G	A	A	G	G	T	A	0	1	B
H9	G	A	G	A	G	C	A	0	0	A
H10	A	A	G	A	C	T	A	1	14	A
H11	G	A	A	A	G	T	A	0	0	A
H12	G	G	G	A	C	T	A	0	0	A
H13	G	G	A	G	G	C	A	0	0	B
Total								22	131	

Nine of the 13 SNP haplotypes were also present at the 38 and 160 individuals sequenced for the 5 mitochondrial regions and/or the 344 bp CR sequences, respectively (Table 4.3). In Table 4.4, distribution of the SNP haplotypes is observed for the 1,254 individuals for 6 stocks, and classification by haplogroup A or B. Percentages of individuals belonging to haplogroups A and B for each stock were similar with analysis based on CR sequences and on SNPs (Table 4.2). Haplogroup A was the most abundant, with the highest individual percentages from Mediterranean and Atlantic specimens (83-96%) and the lowest from Indo-Pacific specimens (61-68%) (Table 4.2). It is remarkable that haplotype H10 was found only in individuals (N = 82; 34.9%) from the Mediterranean Sea (Table 4.4).

Table 4.4. SNP Haplotype and haplogroup distribution for each actually assumed albacore stock. N: Number of individuals

Haplotype	N	Individuals per origin						Haplogroup
		MED	NATL	SATL	IN	NPAC	SPAC	
H1	373	83	173	92	23	1	1	A
H2	135	14	37	16	22	24	22	B
H3	50	2	19	4	13	6	6	B
H4	345	29	181	73	33	11	18	A
H5	156	20	45	24	18	26	23	A
H6	28	1	13	7	6	0	1	B
H7	64	0	15	11	15	13	10	A
H8	10	4	2	3	0	0	1	B
H9	7	0	4	2	1	0	0	A
H10	82	82	0	0	0	0	0	A
H11	2	0	2	0	0	0	0	A
H12	1	0	1	0	0	0	0	A
H13	1	0	0	0	0	0	1	B
Total	1,254	235	492	232	131	81	83	

Differentiation between albacore stocks based on 344 bp CR sequences presented a global value of Φ_{ST} = 0.032 significant ($p < 0.01$). Pairwise Φ_{ST} comparisons only detected significant statistical differences between Mediterranean specimens and the rest of stocks ($p < 0.01$) (Table 4.5). When the differentiation degree was estimated based on SNPs, the global Φ_{ST} value was higher (0.087), and also significant ($p < 0.01$). In this case, all pairwise Φ_{ST} comparisons were significant, except those between North and South hemispheres of Atlantic and Pacific individuals (Table 4.5).

Table 4.5. Genetic differentiation matrix between currently assumed albacores stocks stated by Φst, based on 344 bp CR sequences (below diagonal) and 7 SNPs (above diagonal). (* p-value < 0.01)

	MED	**NAT**	**SAT**	**IND**	**NAP**	**SPA**
MED		0.110*	0.098*	0.114*	0.199*	0.182*
NAT	0.095*		0.001	0.038*	0.160*	0.124*
SAT	0.094*	0.002		0.041*	0.161*	0.130*
IND	0.097*	0.000	0.001		0.046*	0.025*
NPA	0.093*	0.001	0.000	0.000		0.000
SPA	0.094*	0.004	0.001	0.003	0.000	

3.2. *Thunnus* genus

Regarding the phylogeny of *Thunnus* genus, our results supported the monophyletic origin of the 8 species. In fact, NJ tree based on 5 mitochondrial regions (5112 bp) showed all *Thunnus* sequences (38 albacore individuals sequenced in this study, along with 16 GenBank sequences for all *Thunnus* species except for *T. atlanticus)* together on a tree branch, separated from the sequences corresponding to the other three genus included in the analysis: *Katsuwonus pelamis*, *Auxis rochei,* and *Euthynnus lineatus* (Fig. 4.2). Moreover, within *Thunnus* genus, albacore and Pacific bluefin tuna sequences depart together from the other 6 species, constituting one group. Similarly, when drastically increasing the number of sequences to 2,113, at the expense of sequence length (CR 344 bp), network and phylogenetic tree, albacore and Pacific bluefin tuna were also shown to form a group separated from the rest (Fig. 4.3; Fig. 4.4). Moreover, the same picture was obtained when a rooted maximum likelihood (ML) phylogenetic tree was built applying 16 mitochondrial genome-wide GenBank sequences from 7 *Thunnus* species (Fig. 4.2). The coincidence in the pattern obtained, by whatever scale used in the analysis, showed reliability, and thus added value to the results obtained in this study.

Figure 4.2. Rooted NJ tree built using sequences of Cyt *b*, COI, 12S, 16S and CR (5,112 bp) of 38 albacores from this study, and sequences of the genus *Thunnus* from GenBank: *T. alalunga* (3 individuals), *T. obesus* (2), *T. thynnus* (4), *T. tonggol* (2), *T. orientalis* (2), *T. albacares* (2) and *T. maccoyii* (2). One individual of each of the species *Katsuwonus pelamis*, *Auxis rochei* and *Euthynnus lineatus* were included as outgroups. Bootstrap support values above 50% are shown.

The high number of sequences used in the phylogenetic analysis allowed a detailed study of the mitochondrial genome of 2,113 individuals, previously classified in one of the 8 species of the *Thunnus* genus based on their morphology (Fig. 4.3, Table 4.6). Firstly, it is noteworthy that most of the sequences were grouped according to their morphological classification, forming clear clades or branches separated by species. Secondly, Atlantic bluefin tuna morphotype sequences were not all grouped together, either in the network or in the phylogenetic tree (Figs. 4.3 and 4.4). Although the majority were grouped in a large species distinctive clade, several sequences of

Atlantic bluefin tuna formed a group apart, in between Pacific bluefin tuna and albacore. Thirdly, a significant number of individual sequences (n=24) appeared interspersed within clades of species that did not correspond to the morphological classification (Fig. 4.3; Table 4.6). Particularly remarkable, mitochondrial haplotypes of 18 Atlantic bluefin tuna morphotypes were grouped in albacore's clade (albacore-like Atlantic Bluefin tuna*)*, and 2 more in *T. albacares*'s clade. Conversely, the mtDNA sequences of 2 Pacific bluefin tuna morphotypes were grouped in Atlantic bluefin tuna's clade (thynnus-like Pacific bluefin tuna). Finally, a sequence from albacore was grouped with *T. albacares*, and another with Pacific bluefin tuna.

Figure 4.3. Median-joining network of the haplotypes of the 344bp CR sequences in 2,113 individuals of *Thunnus* genus species from this paper and GenBank. Dashed circles indicate clusters belonging to the 8 species. In *T. alalunga*, *T. obesus* and *T. tonggol* two haplogroups (A and B) are differentiated. *T. thynnus* (a) and *T. thynnus* (b) indicate two clusters of *T. thynnus*. White circles indicate haplotypes belonging to the clustered species; black circles indicate individuals not belonging to the cluster species (see Table 4.6).

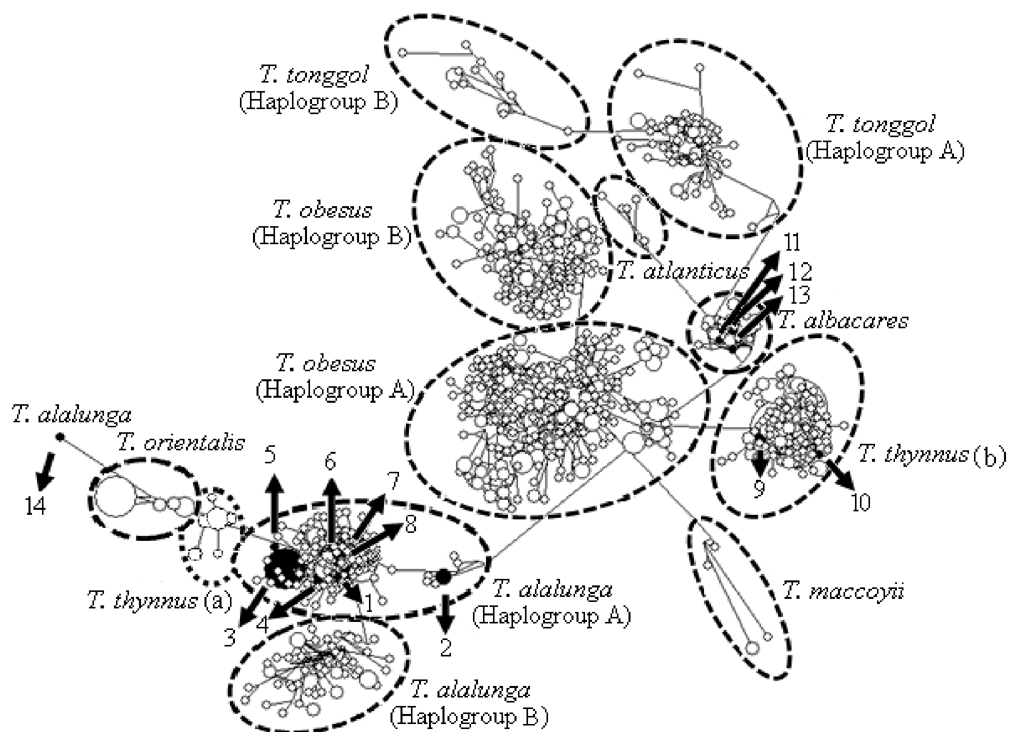


Table 4.6. Information about the 24 sequences that did not group within their species´ cluster in the Median-joining network of CR haplotypes (344 bp) of 2,113 individuals of *Thunnus* genus (Fig. 4.3). **#:** Location of the individuals in Fig. 4.3.;Species: name of the species categorized in GenBank; Group: cluster of species where is included in Fig. 4.3.

#	Species in Genbank	Grouped in this study within	GB Accession number
1	*T. thynnus*	*T. alalunga*	AF390419
1	*T. thynnus*	*T. alalunga*	AY650620
2	*T. thynnus*	*T. alalunga*	AF390379
2	*T. thynnus*	*T. alalunga*	AF390370
2	*T. thynnus*	*T. alalunga*	AY650465
2	*T. thynnus*	*T. alalunga*	AY650452
3	*T. thynnus*	*T. alalunga*	AF390384
3	*T. thynnus*	*T. alalunga*	AF390425
3	*T. thynnus*	*T. alalunga*	AY650713
3	*T. thynnus*	*T. alalunga*	AY650425
3	*T. thynnus*	*T. alalunga*	DQ087541
3	*T. thynnus*	*T. alalunga*	DQ087593
4	*T. thynnus*	*T. alalunga*	AY650502
4	*T. thynnus*	*T. alalunga*	AY650740
5	*T. thynnus*	*T. alalunga*	AY650544
6	*T. thynnus*	*T. alalunga*	AY650659
7	*T. thynnus*	*T. alalunga*	AY699942
8	*T. thynnus*	*T. alalunga*	AY699946
9	*T. orientalis*	*T. thynnus*	GU972554
10	*T. orientalis*	*T. thynnus*	GU972555
11	*T. thynnus*	*T. albacares*	DQ087564
12	*T. thynnus*	*T. albacares*	DQ087565
13	*T. alalunga*	*T. albacares*	JN007604
14	*T. alalunga*	*T. orientalis*	KP412767

Regarding the dating of speciation events on Thunnus genus, Figure 4.4 illustrates the dates obtained by applying two divergence rates, i.e. 3.6 and 11% per million years, to the 2,113 344 bp-long CR mitochondrial sequences. The first bifurcation occurred when the common ancestor of albacore and Pacific bluefin tuna was separated from the rest of the *Thunnus* species, between 4.5 and 1.5 My ago. Interestingly, we found Atlantic bluefin tuna (a) haplotypes within this very same branch. According to our estimates, the common ancestor of these specific Atlantic bluefin tuna (a) haplotypes recently split from Pacific bluefin tuna (0.8 to 0.3 My) being the most recent separation within the genus. Atlantic bluefin tuna (a) sequences in Fig. 4.4 were those that

conform the small group of Atlantic bluefin tuna sequences between albacore and Pacific bluefin tuna in Fig. 4.3 network, apart from the main Atlantic bluefin tuna clade. The second bifurcation within *Thunnus* genus took place soon after: *T. obesus* diverged between 4.0 and 1.3 My ago. However, most of the following speciation events, and even haplogroup differentiation events within *T. tonggol* and *T. obesus* species, would have occurred in a range of between 2.7 and 0.7 My. Finally, much closer in time, between 1.0 and 0.3 My, we found the bifurcation between albacore haplogroups.

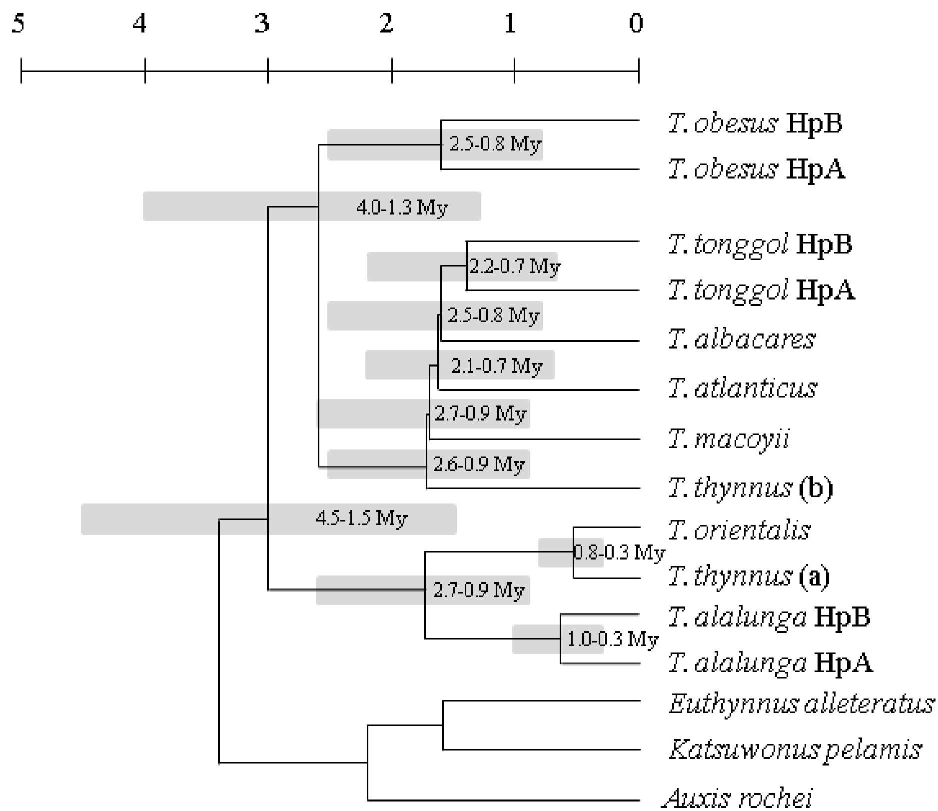


Figure 4.4. Maximum likelihood phylogenetic tree representing 2,113 CR mitochondrial sequences (344 bp), espetiation and cladogenesis events datation based on 3.6 and 11% divergence rates. *T. thynnus* (a) and *T. thynnus* (b) indicate two clusters of *T. thynnus* (as in Fig. 4.3).

4. Discussion

4.1. Population structure in albacore

The results of this study, in which multiple mitochondrial sequences and SNPs were analysed, together with nuclear SNPs information obtained in previous worldwide studies (Albaina *et al.* 2013; Laconcha *et al.* 2015a), allow the reconstruction of the evolutionary process of the albacore.

All the analysis of CR sequences and SNPs from coding mitochondrial regions indicated the existence of two differentiated haplogroups in the albacore. Differentiation between them, based on CR sequences, would have occurred 1.0-0.3 My ago (Fig. 4.4). These results agree with results from Chow *et al.* (1995) who, based on frequencies of the restriction patterns, categorised samples into two large groups. Viñas *et al.* (2004) also defined two phylogroups for these species based on the analysis of CR sequences. In the present study, haplogroup A showed a mean value of genetic distance (0.034) lower than haplogroup B did (0.052). This could indicate that haplogroup A emerged from haplogroup B. A differentiated geographical distribution was detected for haplogroup frequencies. Thus, the Mediterranean Sea and Atlantic Ocean showed frequencies for haplogroup A of between 85.6% and 91.1%, and the Pacific Ocean frequencies of A around 63% (Table 4.2; 7 SNPs). A possible explanation for this distribution of frequencies could be the allopatric evolution of haplogroups, in which case haplogroup A could have emerged 1.0-0.3 My ago in the Mediterranean and/or Atlantic Ocean (Fig. 4.4). A phylogeographic scenario of isolation between de Indo-Pacific and the Atlantic, followed by the unidirectional invasion of the Indo-Pacific clade into the Atlantic was suggested in several other tuna and scombroid fish (Alvarado Bremer *et al*. 1995, 1998; Chow *et al*. 2000). However, despite the asymmetry of the regional distribution of the clades, we do not think that this was the demographic scenario for the albacore since the subsequent gene flow necessary to get the actual distribution of haplogroup frequencies is incompatible with the high genetic distances detected between Oceans with mitochondrial and nuclear SNPs. The alternative hypothesis would be phylogenetic continuity and partial spatial separation, Category V (Avise *et al.* 1987), which coincides better with species with historically intermediate levels of gene flow between geographic populations. In this scenario, two different sets of lineages could have emerged due to the stochastic extinction of haplotypes, generating a bimodal mismatch distribution in a partial geographical isolation of the Atlantic and Indo-Pacific populations (Slatkin and Hudson 1991; Harpending *et al.* 1998).

However, results for Mediterranean albacore would be an exception in this demographic scenario in which populations have remained relatively large and stable for a long time. Some data indicates that Mediterranean albacore is the most genetically differentiated population of the species: mutation accumulations in the CR sequences in each haplogroup (Table 4.5), high frequency (34.9%) of individuals of Mediterranean Sea albacores in an haplogroup (H10)

exclusive from Mediterranean populations (Table 4.4), and level of differentiation between populations obtained from nuclear (Albaina *et al*. 2013; Laconcha *et al.* 2015a) and mitochondrial SNPs. Thus, these data would indicate a differentiation process of the albacore of the Mediterranean Sea ulterior to the emergence of both haplogroups 1.0-0.3 My ago. This population would have been isolated from the rest during some of the last glacial events of the Pleistocene, probably during Riss or Würm glaciations, as proposed for swordfish by Alvarado-Bremer *et al.* (1995). Furthermore, the high frequency of the haplotype H10 and its detection exclusively in the Mediterranean would indicate that actually there is isolation, and that although other species like Atlantic bluefin tuna and swordfish migrate across the Strait of Gibraltar (Block *et al*. 2005; Alvarado Bremer *et al*. 2005; Aranda *et al*. 2013), migration is negligible for albacore.

Pairwise Φ_{ST} comparisons between stock results based on mitochondrial SNPs agreed with the scenarios described above. SNPs detect the phylogenetic continuity and partial spatial separation of the Atlantic and Indo-Pacific described with haplogroups A and B derived from CR sequences, and also the isolation of the Mediterranean population. Thus, higher differentiation occurred between the Mediterranean Sea and the Pacific Ocean (Φ_{ST} = 0.182-0.199; Table 4.5), intermediate between Atlantic and Pacific Oceans (Φ_{ST} = 0.124-0.161) and lower between Mediterranean Sea and Atlantic Ocean (Φ_{ST} = 0.098-0.110). These results agreed with those from Laconcha *et al*. (2015a) based on SNPs from transcriptome. Furthermore, since SNPs are in mitochondrial genes, and therefore their mutation rate is lower than CR sequences, SNPs detected differences that were not reflected in CR sequences. Thus, SNPs from coding mitochondrial regions determine that populations from Atlantic, Indian and Pacific Oceans were genetically internally homogeneous, but significantly different from each other. Therefore, four albacore populations were defined: one in each ocean and one in the Mediterranean Sea. This structure of albacore based on four populations seems to be definitive since it agrees with the results obtained with SNPs from nuclear genome by Albaina *et al*. (2013) and SNPs from transcriptome by Laconcha *et al*. (2015a), and also in part with Montes *et al*. (2012) based on microsatellites. Other studies using nuclear markers in non-worldwide albacore samples also describe similar structuring patterns (Takagi *et al.* 2001; Arrizabalaga *et al.* 2004; Davies *et al.* 2011). In short, results obtained from both nuclear and mitochondrial markers detect the same population structuring, which indicates that evolutionary mechanisms do not affect females and males differently in albacore. Finally, it is remarkable that since Albaina *et al.* (2013) and Laconcha *et al.* (2015a) managed to solve the population structure of the albacore using 53 and 75 nuclear SNPs, respectively, in the present study the same structuring was detected based only on 7 mitochondrial SNPs. However, this structuring was not detected with the analysis of sequences of 5 fragments (5,112 bp) or CR (344 bp) sequences, most likely because the analysed sample was much smaller than those used with mitochondrial SNPs. Another factor that might have had an impact was that in the SNPs selection, the CR region was not taken into account due to hypervariability; the genotyping technology did not allow to design probes in this mitochondrial region. Thus, the analysis of an ample sample based on a relative low number of SNP markers from the coding

mitochondrial region could be a simplified and fast approach for the analysis of the population structure of other *Thunnus* species.

4.2. Evolution of *Thunnus* genus

As regards the phylogeny of *Thunnus* genus, the present study, as well as other studies which have analysed all the species of the genus based on mitochondrial sequences (Chow & Inou 1993; Chow & Kishino 1995; Viñas & Tudela 2009; Tseng *et al.* 2012) or nuclear sequences (Chow *et al.* 2006), showed that the division between the *Neothunnus* and *Thunnus* subgenus proposed by Collette (1978) based on the morphoecologic characteristics did not coincide with the phylogeny of the genus. First of all, the results from our study clearly differentiated the group composed of two of the 5 species from *Thunnus* subgenus, albacore and Pacific bluefin tuna, which formed the most ancient group. Secondly, they highlighted the cluster in a unique branch of the three species of *Neothunnus* subgenus (*T. atlanticus*, *T. albacares* and *T. tonggol*) that differentiated from the other 3 species of *Thunnus* subgenus (Atlantic bluefin tuna, Southern bluefin tuna and *T. obesus).* Thus, the present study would support monophyletic origin of the *Neothunnus* subgenus, but paraphyletic origin for *Thunnus* subgenus species, in accordance with Tseng *et al.* (2012).

The phylogenetic analysis of a large number of mitochondrial haplotypes (mitotypes) showed that there is correlation between genetic and morphological data, when the classification of individuals of the *Thunnus* genus within each of the eight species is taken into account. However, we also found mitotypes "misplaced". That is, we found some odd locations for both group and individual CR sequences that were located outside the mitochondrial branch that corresponds to the morphological classification (Figs. 4.3 and 4.4). These observations demonstrated the existence of hybridisation events between species of the *Thunnus* genus, either historical and/or recent. In this regard, the locations of the Atlantic and Pacific bluefin tuna sequences are remarkable both in themselves and also with respect to the albacore. Most Pacific bluefin tuna and some Atlantic ones -*T. thynnus* (a) sequences- were placed together near albacore, whereas most of the individuals of the Atlantic bluefin tuna and few Pacific ones formed a main clade far away from the first group (Fig. 4.3). On the one hand these results could explain why Chow & Inoue (1993), using RFLP markers from cytochrome b (Cyt *b*), 12S and ATPase mitochondrial genes, and Chow & Kishino (1995), analysing sequences from Cyt *b* and ATPase mitochondrial genes, detected high phylogenetic proximity between Pacific bluefin tuna and albacore, while analysis carried out with ITS1 nuclear sequences (Chow & Kishino 1995; Chow *et al.* 2006) genetically related Pacific bluefin tuna with Atlantic bluefin tuna but not with albacore. On the other hand, these kinds of translocated groups have been previously identified by other authors, and interpreted as introgression between Pacific and Atlantic bluefin tuna (Chow and Kishino 1995; Alvarado Bremer *et al.* 2005; Viñas and Tudela 2009). However, there are data which suggest an alternative hypothesis; an ancient introgression of albacore mtDNA in the common ancestor of Pacific and Atlantic Bluefin tuna. Firstly, most Pacific bluefin tuna and

Atlantic bluefin (a) haplotypes were located within albacore's branch, indicating that these three haplotype groups descended from the first bifurcation event within *Thunnus* genus, so much distant from the main Atlantic bluefin clade (b). Secondly, the absence of contact between the current distribution areas of Pacific and Atlantic bluefin tunas rules out recent hybridisation events between them, and indicates that the two haplotypes of Pacific bluefin (9 and 10 in Fig.4.3) located within the main clade (b) of Atlantic bluefin in the network should be the result of ancient hybridisation events. Thirdly, morphological similarity between Pacific and Atlantic bluefin, which even led to them being considered a single species until recently (Collete 1999), and also the fact they share nearly identical ITS1 sequences well within the range of intraspecific variation (Chow *et al.* 2006), supports the idea of a relative recent and common ancestry between Pacific and Atlantic bluefin tunas. Thus, these observations all together are more concordant on the scenario of an ancient introgression of mtDNA of albacore into the common ancestry between Pacific and Atlantic bluefin tuna. The lower genetic diversity detected in the haplogroup close to the albacore indicated that this was the result of the introgression of the albacore, whereas the mtDNA descendent from common ancestry between Pacific and Atlantic bluefin tuna would be the main clade (b) of Atlantic bluefin tuna situated near to the main species of the *Thunnus* genus. In this way, Atlantic bluefin tuna (a) haplotypes and the Pacific bluefin group's network showed few central main haplotypes, with rarer haplotypes radiating outward. These kinds of networks are associated with founder effect events, including very few individuals, followed by an expansion period with random mutations through time. Thus, present study results most probably reflect very few hybridisation events between males of the common ancestor of Pacific and Atlantic bluefin males, with albacore females, followed by repeated backcrossing of the interspecific hybrids with the Atlantic-Pacific bluefin tuna common parent species.

In fact, a previously described introgression event of the mitochondrial albacore genome into the common ancestor of Pacific and Atlantic bluefin does not appear to be the result of an isolated and unique hybridisation event between species of the *Thunnus* genus. On the contrary, 22 sequences from three different species were detected, in addition to the above-mentioned two, located in clades which did not belong to their own species (Fig. 4.3; Table 4.6), and they are probably reflecting more recent introgression events. We described them as recent since the divergent times estimated for these misplaced sequences fall within the range of the corresponding haplogroup where they are located in the network (data not shown). Among them, 18 albacore-like Atlantic bluefin tuna sequences, and 2 albacore-like *T. albacares* are remarkable. These two are probably the result of recent hybridisation between Atlantic bluefin tuna males and albacore (n=18) or *T. albacares* (n=2) females.

The cladogenetic and speciation events that happened in the *Thunnus* genus about 2.7-0.9 My ago (Fig. 4.4) are remarkable. If we take into account the divergence rate of 3.6%, this process would have happened in 600,000 years, around 2.7-2.1 My ago, which is coincident with the beginning of the Gelasian and the Pleistocene Ages (2.58-1.80 My; Gibbard *et al.* 2010). If a divergence rate of 11% is considered, the cladogenetic events would have happened in 200,000

years, around 0.9-0.7 My, at the end of the Calabrian Age (1.80-0.78 My) and beginning of the Ionian Age (0.78-0.12 My; Gibbard *et al.* 2010). It makes sense that a process of this magnitude must be correlated with some biogeographic or climatic event. In this way, we could identify this event at the beginning of the glaciation events of glaciation cycles, 2.58 My ago, conditioned by the cyclic growth and decay of continental ice sheets, the associated climate and environmental changes and rises and falls in sea levels. This date is concordant with the range obtained with the divergence rate of 3.6%, proposed for the fish CR (Donaldson & Wilson 1999). Thus, if results obtained with the divergence rate of 3.6% for My (Fig. 4.4), the separation of the albacore with respect to the rest of the *Thunnus* species would have happened around 4.5 My ago. After that, 4.0 My ago, there would have been the separation of *T. obesus*. As we have previously stated, between 2.7 and 2.1 My ago a cladogenetic process would have happened where the rest of species have emerged, and also would have originated the two mitochondrial lineages of *T. obesus* and *T. tonggol*. Temporally situated within that temporal range would be the emergence of the ancestry species of the actual Pacific and Atlantic bluefin tunas, and not much later than 2.7 My ago would have been the introgression event of mtDNA of albacore into the common ancestry between Pacific and Atlantic bluefin tuna species. Furthermore, the haplotypes of these two species next to the albacore allow us to estimate the time of separation between them. These divergences would have happened around 0.8 My ago, probably in an interglacial period from Pleistocene, when the temperature and sea levels were increasing and tuna populations could move via the Arctic Ocean. The fact that most Pacific bluefin tunas were close to the cluster of albacore, and that the haplotypic diversity was remarkably higher in the Atlantic bluefin tuna than in the Pacific one (Table S4.1), suggested that the ancient population of both species originally inhabited the Atlantic Ocean and, after a founder effect, populated the Pacific Ocean generating Pacific bluefin tuna.

Acknowledgments

The authors are indebted to technical and human support provided by General Research Services SGIker in the University of the Basque Country (UPV/EHU). We are grateful to the following institutes who provided samples: South African Institute for Aquatic Biodiversity (SAIAB, South Africa), Galway-Mayo Institute of Technology (GMIT, Ireland), Instituto Español de Oceanografía (IEO, Spain). We also are grateful to E. Jimenez, M.A. Pardo, I. Fraile, I. Arregi, J. Lopez, J. Filmarter, J. Areso, A. Delgado, D. Brophy, K.Schaefer, D. Fuller and V. Allain for sampling assistance.

Supplementary material

Table S4.1. Individuals of each species according to three analysis considered: Complete mtDNA, 5 regions (Cyt *b*, COI, 12S, 16S and CR) and 344 bp CR sequences (CR). Number of haplotypes (nh), haplotype diversity (h) and nucleotide diversity (π) were calculated for CR sequences.

	Complete	5 regions	CR			
	N	**N**	**N**	**nh**	**h**	**π**
Thunnus obesus (BET)	2	2	1052	744	0.999	0.057
Thunnus thynnus (BFT)	4	4	630	311	0.992	0.031
Thunnus alalunga (ALB)	3	41	274	224	0.995	0.041
Thunnus tonggol (LOT)	2	2	93	84	0.998	0.043
Thunnus orientalis (PBF)	2	2	32	7	0.593	0.037
Thunnus albacares (YFT)	2	2	22	18	0.983	0.027
Thunnus maccoyii (SBF)	2	2	5	4	0.900	0.042
Thunnus atlanticus (BLF)	-	-	5	5	1.000	0.041
Total	17	55	2,113			

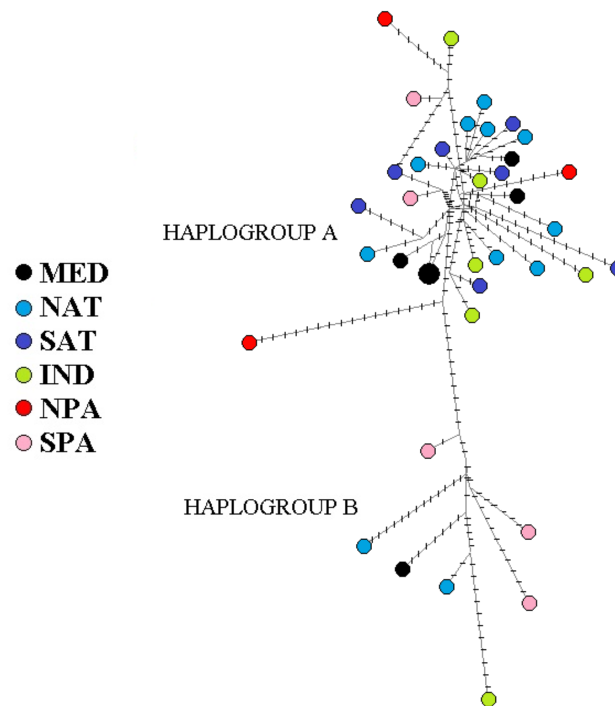


Figure S4.1. Median-joining network built using sequences of Cyt *b*, COI, 12S, 16S and CR (5,112 bp) of 38 albacores. Haplotypes with their stock of origin denoted by colors (see key). Mutated positions denoted by dashes.

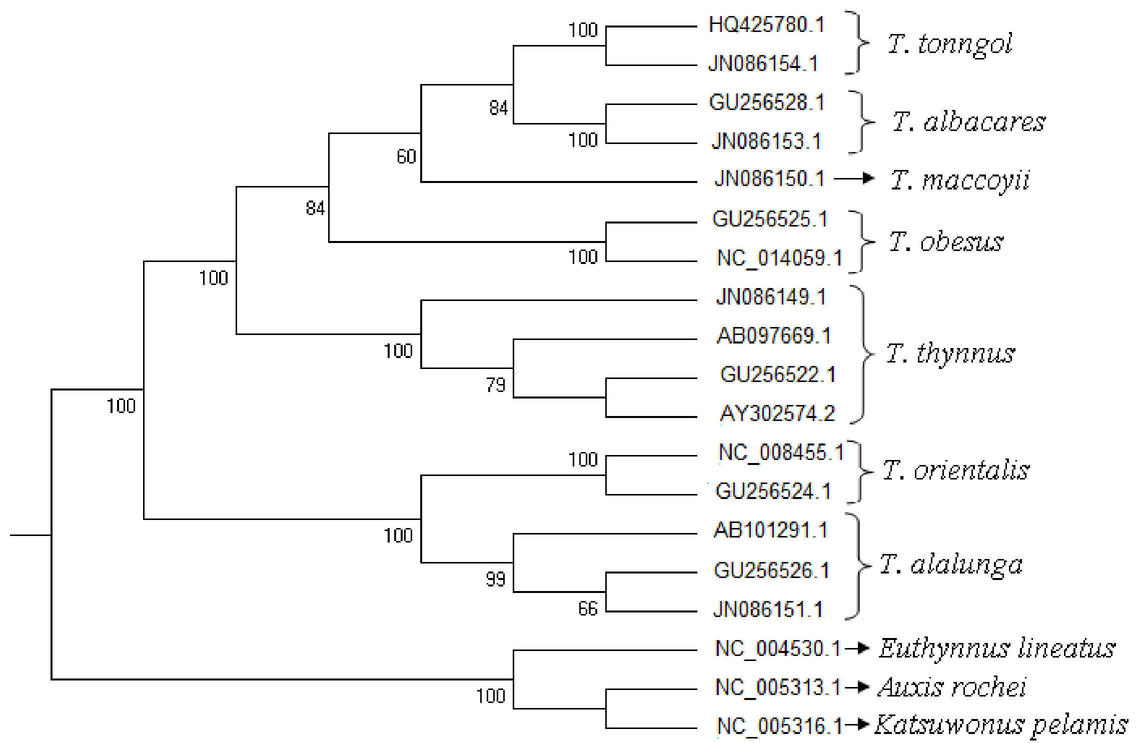


Figure S4.2. Rooted maximum likelihood (ML) tree built using complete mitochondrial sequences of 16 *Thunnus* individuals from 7 species obtained from GenBank. Sequences from *Katsuwonus pelamis*, *Auxis rochei* and *Euthynnus lineatus* were included as outgroups. Bootstrap support above 50% is shown by the branches.

Capítulo VI

Discusión general

DISCUSIÓN GENERAL

Según los criterios de la Unión Internacional para la Conservación de la Naturaleza (IUCN), el 18% de las especies de escómbridos y peces vela presentan deficiencia de datos para su correcta evaluación. El 64% de estas especies presenta una situación de "Preocupación Menor", el 7% se consideran "Casi Amenazadas" y el 11% se clasifican en alguna de las categorías de "Amenazada" (en grave riesgo de extinción, amenazada o vulnerable) (Collette *et al.* 2011). Actualmente, la mayoría de los stocks del atún blanco y del atún rojo del Atlántico se encuentran sobreexplotados (Collette 2011). Esta situación de sobrepesca hace que se agudice la necesidad de desarrollar metodologías para la evaluación precisa de los stocks, y la delimitación de las unidades de gestión – o stocks- debería ser compatible con la estructura poblacional real de las especies explotadas. De hecho, una gestión pesquera orientada hacia el rendimiento sostenible, además de contemplar la dinámica y ecología de las especies, requiere considerar la estructura genética de sus poblaciones a la hora de definir los stocks. En ese sentido, el atún blanco ha sido estudiado por varios grupos, incluyendo el nuestro. Los estudios realizados detectan heterogeneidad entre océanos, si bien adolecen de la representación poblacional y del número de marcadores que incluye el estudio presentado en este trabajo.

El peligro de sobreexplotación del atún blanco trae consigo la necesidad de implantar medidas de protección y control del origen de los productos derivados. Además, el atún blanco del océano Atlántico está muy valorado en el mercado por sus características sensoriales de firmeza y sabor, y sin embargo sólo supone el 20% de las capturas mundiales. Existe, por tanto, riesgo de fraude en el etiquetado y en la declaración de desembarcos, y la dimensión de este fraude puede alcanzar una importancia económica considerable. Estas prácticas irregulares, además de defraudar al consumidor, dificultan la evaluación de stocks, y ponen de manifiesto la necesidad de implantar medidas de protección en las que se controle el origen de la pesca y de sus productos derivados. En este sentido, son totalmente necesarios el diseño y refuerzo de las herramientas que supongan una mejora en el control del origen de los productos derivados del atún blanco, y las herramientas de ADN son un medio muy útil para trazar el origen geográfico de un producto pesquero. La mayoría de los estudios genéticos realizados hasta la fecha están centrados en la discriminación a nivel de especie, y el único método existente para discriminar las diferentes poblaciones de atún blanco (Patente ES 2 392 293 B1, que utiliza 18 SNPs) es una herramienta mejorable.

Por otro lado, las relaciones filogenéticas en el género *Thunnus* siguen siendo una cuestión sin resolver (Alvarado Bremer *et al.* 1997; Tsen *et al.* 2012), y apenas existen dataciones de los eventos de especiación dentro del género. El análisis de un elevado número de secuencias correspondientes a varias regiones mitocondriales ayudará a arrojar luz sobre la historia evolutiva del género *Thunnus,* y sobre la controversia taxonómica existente acerca de los subgéneros *Thunnus*

y *Neothunnus.* Resulta importante resolver definitivamente si esta subdivisión del género está apoyada o no por los datos genéticos.

En vista de lo anteriormente expuesto, y teniendo en cuenta la escasez de estudios genéticos que hayan abordado de una manera global la estructura poblacional del atún blanco, el primer objetivo general planteado en esta Tesis Doctoral ha sido caracterizar la estructura genética del atún blanco (*Thunnus alalunga*) en toda su área de distribución. El conocimiento generado al alcanzar este objetivo ha sido útil para delimitar las unidades de gestión en esta especie, y también para desarrollar un sistema de autentificación del origen geográfico del atún blanco. El segundo objetivo general ha sido profundizar en el conocimiento de la historia evolutiva del género *Thunnus*, tanto en lo referente a las relaciones filogenéticas entre las especies que lo conforman, como a la datación de los eventos de especiación y cladogénesis.

Los resultados obtenidos en la presente Tesis Doctoral se han recogido en cuatro estudios. En los estudios I y II se han desarrollado recursos genéticos para el análisis del atún blanco, que han permitido abordar el estudio de la diversidad genética y de la estructura poblacional de la especie desde la perspectiva del ADN nuclear. El estudio III está basado en recursos genéticos derivados del ADN mitocondrial, y tiene el objetivo de esclarecer la estructura poblacional del atún blanco y arrojar luz sobre las relaciones filogenéticas del género *Thunnus*. Por último, el estudio IV representa un ejemplo de aplicación práctica de una herramienta genética, que busca conseguir la autentificación del origen geográfico para las diferentes poblaciones que incluye la especie.

-ESTUDIO I: **Single nucleotide polymorphism discovery in albacore and Atlantic bluefin tuna provides insights into worldwide population structure**

Este estudio describe por primera vez SNPs nucleares en el atún blanco, y los aplica en la caracterización de la estructura genética de esta especie. Además, mediante amplificación cruzada, valida en el atún rojo del Atlántico varios de los SNPs descubiertos en el atún blanco. Los métodos que utilizamos para el descubrimiento de SNPs en este estudio se basaron en la secuenciación Sanger. Para el estudio genético-poblacional, se analizaron muestras de los seis stocks descritos por el ICCAT para el atún blanco y los dos del atún rojo del Atlántico, examinando un total de 460 individuos para el atún blanco y 107 individuos para el atún rojo del Atlántico. Detectamos diferencias genéticas significativas entre los stock, delimitando al menos cuatro stocks diferenciados para el atún blanco y dos para el atún rojo del Atlántico.

-ESTUDIO II: **New nuclear SNP markers unravel the genetic structure and effective population size of albacore tuna (*Thunnus alalunga*)**

En este estudio se caracterizaron nuevos marcadores SNP nucleares en el atún blanco, derivados de los descubiertos mediante secuenciación masiva en el transcriptoma del atún rojo del Atlántico. Además de analizar nuevos marcadores SNP, en este estudio, centrado en el atún blanco, se aumentó notablemente el número de ejemplares analizados por muestra respecto al ESTUDIO I, y se añadieron muestras recogidas hasta con 24 años de diferencia, alcanzando un total de 1331 individuos. Todo ello permitió, además de comprobar la estructura genética propuesta en el estudio anterior, realizar una estima del tamaño efectivo del atún blanco del Atlántico norte. A este respecto los resultados indican que la diversidad genética y, por consiguiente, el potencial de adaptación del stock no se han visto afectados significativamente por la sobrepesca.

-ESTUDIO III: **Phylogeographic relationships of worldwide albacore (*Thunnus alalunga*) stocks and phylogenetic relationships of *Thunnus* genus based on mitochondrial DNA**

En este estudio se analizaron las secuencias de 5 regiones y 10 SNPs mitocondriales generados en este estudio para el atún blanco. Se consideraron también las correspondientes secuencias mitocondriales existentes en la base de datos GenBank, para todas las especies pertenecientes al género *Thunnus*. El análisis fitogeográfico evidenció la existencia de dos haplogrupos en el atún blanco, cuyas frecuencias mostraban correlación con la distancia geográfica, excepto para el Mediterráneo, que mostraba un alto grado de aislamiento. Además, los análisis basados en los SNPs mitocondriales coinciden con los resultados obtenidos con los SNPs nucleares, al diferenciar cuatro poblaciones: Mar Mediterráneo, Océano Atlántico, Océano Índico y Océano Pacífico (Φ_{ST} =0.038-0.199). Por otro lado, el análisis de las cinco regiones mitocondriales mencionadas y el genoma mitocondrial completo de las especies del género *Thunnus* mostró que la actual subdivision del género en los subgéneros *Neothunnus* y *Thunnus* no se corresponde con los datos moleculares. Mediante estos análisis también quedó patente que el atún blanco constituye el grupo más antiguo del género, separado hace 4.5 Ma. Por último, los resultados evidencian eventos de hibridación históricos y recientes entre especies del género. En este sentido, los resultados apuntan hacia una antigua introgresión entre el ADNmt del atún blanco y el ancestro común del atún rojo del Atlántico y del atún rojo del Pacífico (*Thunnus orientalis*). Tras este evento introgresivo, hace alrededor de 0.8 Ma, tras un efecto fundador, el ancestro común de los dos atunes rojos poblaría el océano Pacífico dando lugar al atún rojo del sur.

-ESTUDIO IV: **A genetic traceability tool for Atlantic albacore authentication based on SNP markers**

El cuarto estudio constituye un ejemplo aplicado del uso de marcadores genéticos para la asignación del origen geográfico de ejemplares pescados de atún blanco, y/o a sus productos derivados. Es decir, se ha desarrollado una herramienta de trazabilidad útil y de fácil aplicación en la cadena de producción del atún blanco. La herramienta consiste en un panel de 32 marcadores SNP: 30 seleccionados a partir de los ESTUDIOS I, II y III, y 2 SNPs mitocondriales adicionales descritos previamente para la diferenciación de la especie atún blanco respecto al resto de especies del género *Thunnus*. El método propuesto es capaz de discriminar entre los atunes con origen en las poblaciones del Atlántico, Mediterráneo e Indo-Pacífico, en el 68% de los casos con el 100% de probabilidad de acierto.

1. Descubrimiento de marcadores SNP en el Atún blanco

Los primeros trabajos moleculares para el estudio de la estructura genética del atún blanco se basaron en el análisis de alozimas (Suzuki, 1962; Serene, 1969; Hallaire y Dao, 1971; Hue, 1979; Pujolar *et al.* 2003), RFLPs (Chow y Ushiama, 1995) y grupos sanguíneos (Arrizabalaga *et al.* 2004). A partir de los años 80 los estudios genéticos pasaron a analizar secuencias del ADN mitocondrial y, en menor medida, nuclear (Graves y Dizon, 1989; Viñas *et al.* 1999; Viñas *et al.* 2004; Nakadate *et al.* 2005), así como *loci* microsatélites (Takagi *et al.* 2001; Davies *et al.* 2011; Montes *et al.* 2012). No obstante, los marcadores SNP están sustituyendo a los microsatélites y secuencias de ADN debido a las ventajas que presentan en varios campos de la investigación, como son la ecología, la conservación y la acuicultura (Brumfield *et al.* 2003; Morin *et al.* 2004). Sin embargo, para los organismos no-modelo, como es el caso de los túnidos, la falta de información sobre la secuencia completa del genoma ha sido un factor limitante para el descubrimiento de marcadores SNP. Por ello, en el caso del atún blanco se han utilizado metodologías como la secuenciación comparativa de fragmentos amplificados al azar (McLenachan *et al.* 2000; Bensch *et al.* 2002).

En esta Tesis Doctoral se han utilizado 3 aproximaciones metodológicas para el descubrimiento de un total de 751 SNPs en el Atún blanco:

1.- La metodología EPIC-PCR (Exon-Primed Intron-Crossing) (Palumbi y Baker, 1994; Lyons *et al.* 1997; Shubitowski *et al*. 2001; Primmer *et al*. 2002) a partir de secuencias del género *Thunnus* y de otras especies de teleósteos obtenidas de la bibliografía y de bases de datos (GenBank y Ensembl) (ESTUDIO I).

2.- Secuenciación comparativa de fragmentos del genoma del atún blanco obtenidos a partir de primers degenerados de secuencias conservadas en teleósteos (ESTUDIO I), y en el género *Thunnus* (ESTUDIO I y ESTUDIO III).

3.- Amplificación cruzada entre especies (Cross species amplification) a partir de SNPs descubiertos en el atún rojo del Atlántico (ESTUDIO II).

Así, mediante la aplicación de las dos primeras estrategias citadas se obtuvieron 54 nuevas secuencias a lo largo del genoma del atún blanco: 19 regiones intrónicas mediante la metodología EPIC-PCR, 17 mediante la utilización de primers degenerados y 18 regiones a partir de primers diseñados en base a secuencias del género *Thunnus*. Además, se proporcionó un panel validado de 67 SNPs nucleares en equilibrio de Hardy- Weinberg (H-W) que constituían un total de 41 marcadores SNP independientes (32 SNP individuales y 9 bloques haplotípicos). Se seleccionó un solo SNP o un solo bloque haplotípico por cada uno de los 54 fragmentos, para mantener la independencia entre los marcadores.

Mediante la aplicación de la tercera estrategia, se validaron otros 75 marcadores nucleares más, independientes y en equilibrio de H-W. Éstos incluían un único bloque haplotípico constituido por 2 SNPs.

Todas las estrategias garantizan el descubrimiento de SNPs neutrales, básicos para estimar procesos evolutivos como la deriva genética y el flujo génico, que se ven sesgados con la inclusión de marcadores bajo selección (Beaumont y Nichols, 1996; Laval *et al.* 2010). Las teorías convencionales de genética de poblaciones y la mayoría de los estadísticos que son comúnmente aplicados, asumen la neutralidad de los marcadores, por lo que estas tres estrategias resultaban atractivas cara al estudio de la estructura poblacional del atún blanco que se pretendía realizar. De hecho, en este trabajo se definieron un total de 96 (38+58) *loci* nucleares y 9 SNP mitocondriales validados con variación neutral.

Ahora bien, las secuencias analizadas en las estrategias 2 y 3 son regiones del genoma altamente conservadas entre teleósteos o túnidos, y por tanto con potencial para estar seleccionadas en cuanto que implican función. De hecho, de los 41 *loci* nucleares estudiados, tres son candidatos a estar bajo el efecto de la selección diversificadora. Del mismo modo, la estrategia 3, proporcionaba SNPs derivados del transcriptoma del atún rojo del Atlántico. El hecho de que los SNPs descubiertos estén conservados en el transcriptoma de las dos especies de túnidos, abre la posibilidad de descubrir SNPs funcionales implicados en procesos biológicos adaptativos importantes en túnidos. De hecho, de los 75 *loci* estudiados, 17 son candidatos a estar bajo el efecto de la selección balanceadora.

Además de los marcadores nucleares, en el ESTUDIO III se secuenciaron 5 regiones mitocondriales en 38 atunes blancos pertenecientes a los 6 stocks definidos actualmente en la

especie (un total de 190 secuencias), y se describieron 159 SNPs mitocondriales. De éstos, se seleccionaron 10 y se validaron 9 SNPs mitocondriales. La información del genoma mitocondrial es un recurso evolutivo valioso y complementario al nuclear en el sentido que hace referencia a una escala temporal diferente (Bohonak y Vandergast 2011; Wang 2011).

Los porcentajes de validación de SNPs obtenidos son satisfactorios (25-72%). Además, el conjunto de SNPs validados no estaría sujeto a grandes sesgos derivados de su descubrimiento, ya que el panel de individuos seleccionado para el descubrimiento de marcadores incluía individuos de todos los stocks de la especie, lo que los hace adecuados para futuras investigaciones a nivel global.

Podemos considerar que las tres estrategias arriba mencionadas resultaron ser efectivas y satisfactorias. De hecho, se generaron un total de 163 secuencias referencia del genoma nuclear y mitocondrial del atún blanco, y se analizó la variación de 137 SNPs a nivel mundial. De esta forma, se ha ampliado la cobertura del genoma en esta especie no-modelo, proporcionando un amplio número de marcadores neutrales y selectivos que pueden ser utilizados en estudios de genética de poblaciones. Sin embargo, habría que señalar que su coste en términos económicos y temporales no resulta ser el óptimo frente a las nuevas estrategias que se están desarrollando en los últimos años en torno a la Secuenciación de Nueva Generación (NGS), ya que estas metodologías permiten el descubrimiento de un número superior de SNPs y microsatélites, en un tiempo menor, abaratando costes y automatizando los análisis (Seeb *et al.* 2011).

2. Estructura genética del atún blanco

El gran tamaño poblacional y la alta capacidad migratoria hacen que muchas especies marinas muestren diferenciaciones débiles entre poblaciones, en comparación con las halladas en especies de agua dulce y especies anádromas (Waples, 1998; DeWoody y Avise, 2000). No obstante, son numerosos los trabajos que encuentran diferenciación poblacional significativa en especies pelágicas marinas (Carlsson *et al.* 2004; Castillo *et al.* 2004; Ruzzante *et al.* 2006; Gaggiotti *et al*. 2009; Poulsen *et al.* 2011).

En los casos en los que la diferenciación genética es presumiblemente baja, es muy importante reforzar dos aspectos a la hora de planificar un estudio: 1) utilizar un número de marcadores con resolución suficiente como para detectar estructuración genética en la especie objetivo, si la hubiera y 2) obtener un amplio número de muestras representativo de cada población y de un amplio espectro temporal.

Existen varios trabajos que contemplan el estudio de la estructura poblacional del atún blanco, tanto estudios merísticos y morfológicos (Godsil 1948; Kurogane y Hiyama 1959; Ishii 1965; Yeh *et al.* 1996) como genéticos. La mayoría de ellos han abarcado la distribución del atún blanco sólo parcialmente. Los ESTUDIOS I, II y III de la presente Tesis Doctoral son los únicos,

junto con el estudio de Montes *et al.* (2012), que abarcan la práctica totalidad del rango de distribución de la especie. De este modo, los ESTUDIOS I, II y III aportan en conjunto una visión genética global de la especie, con el análisis de muestras que cubren de forma satisfactoria la distribución global de la misma mediante un número suficiente de marcadores: tres paneles de marcadores SNPs, siendo dos nucleares (ESTUDIO I y II), y uno mitocondrial (ESTUDIO III).

Los estudios I, II y III agruparon las muestras analizadas en 4 poblaciones genéticamente diferenciadas entre sí: 1) mar Mediterráneo, 2) océano Atlántico, 3) océano Índico, y 4) océano Pacífico. Mediante el análisis de correlación entre distancias genéticas y distancias geográficas llevada a cabo en el ESTUDIO II se encontró una correlación significativa desde el Atlántico Norte hasta el Pacífico Norte, pasando por el océano Índico. Sin embargo, cuando las muestras del Mar Mediterráneo fueron incluidas en el análisis, la correlación no resultaba significativa (Figura S3.1b). Estos análisis sugieren un modelo de aislamiento por distancia en el que la deriva genética superaría el efecto del flujo génico a medida que aumenta la distancia geográfica (Figura S3.1a), a excepción del stock del mar Mediterráneo que se muestra excepcionalmente aislado. Además, tanto los análisis de las secuencias de la región control como de los SNPs de la zona mitocondrial codificante, mostraron la existencia de dos haplogrupos diferenciados en el atún blanco, que mostraron una distribución diferencial de sus frecuencias (Tabla 4.2). Aunque una posible explicación para esta distribución sería una evolución alopátrica de los haplogrupos, no pensamos que sea éste el escenario demográfico del atún blanco, ya que la cantidad de flujo génico posterior necesario para lograr la distribución actual de frecuencias de los haplogrupos es incompatible con las distancias genéticas tan elevadas que se detectaron entre los Océanos, tanto con SNPs mitocondriales (Tabla 4.5) como con SNPs nucleares (Tabla 2.4 y Tabla 3.2). Una posible hipótesis explicativa es la siguiente: en un escenario de aislamiento geográfico parcial entre el Atlántico y el Indo-Pacífico, podría haber emergido una distribución bimodal de haplogrupos (dos linajes o haplogrupos) a partir de la extinción estocástica de haplotipos.

2.1. Mar Mediterráneo

No obstante, los resultados obtenidos para el atún blanco del Mediterraneo en los tres estudios indican que esta población podría constituir una excepción al escenario demográfico mencionado anteriormente, en el que las poblaciones habrían permanecido relativamente grandes y estables en el tiempo. Hay una serie de datos que indican que el atún blanco del Mar Mediterráneo es la población genéticamente más diferenciada dentro de la especie. Esta marcada diferenciación de la población mediterránea respecto al resto de poblaciones queda patente en el análisis de correlación entre distancias genéticas y geográficas (Figura S3.1), pero también en el resto de análisis. Los análisis de clustering individual llevados a cabo con los programas STRUCTURE (2.3 y Figura 3.4) y GENELAND, y el dendrograma basado en las distancias Reynolds (Figura 3.3) también indican que la población del Mar Mediterráneo es la más distante genéticamente. En los

mismos términos los valores F_{ST} calculados tanto en el ESTUDIO I como el ESTUDIO II, muestran que el Mar Mediterráneo está genéticamente más diferenciado respecto al resto de poblaciones (Tabla 2.4 y Tabla 3.2). Los análisis basados en el ADNmt también evidencian la mayor diferenciación del Mediterráneo respecto al resto de poblaciones; por un lado el 34.9% de los atunes blancos genotipados para los SNPs mitocondriales seleccionados comparten un haplotipo (H10) exclusivo del Mediterráneo (Tabla 4.4). Por otro lado, los valores Φ_{ST} de la comparación de las secuencias del gen CR solo resultaron significativos en las comparaciones de las muestras del Mediterráneo frente al resto de poblaciones y para los SNPs mitocondriales seleccionados los valores medios de Φ_{ST} eran mayores para el Mediterráneo que para el resto (Tabla 4.5).

La diferenciación de la población mediterránea ha sido descrita previamente por otros autores: Viñas *et al.* (2004) mediante secuencias del ADNmt, Nakadate *et al.* (2005) mediante secuencias del ADNmt y RFLP nucleares, y Davies *et al.* (2011) y Montes *et al.* (2012) mediante marcadores microsatélites. Aunque estos estudios describieron heterogeneidad entre el Medíterráneo y el resto de poblaciones, Viñas *et al.* (1999) en base a secuencias del ADNmt y Pujolar *et al.* (2003) mediante alozimas, no encontraron diferencias entre el océano Atlántico y el Mediterráneo, debido seguramente al bajo número de muestras en el primero y la baja resolución de los marcadores en el segundo.

Por otro lado, son dos los estudios que han encontrado estructuración dentro del Mediterráneo: Davies *et al.* (2011) y Montes *et al.* (2012). Ambos trabajos utilizaron marcadores microsatélites y encontraron diferencias entre el Mediterráneo occidental y el Mediterráneo central. En el ESTUDIO II de esta Tesis sin embargo, no se obtuvo ninguna diferenciación significativa entre las 7 muestras mediterráneas analizadas (Figura 3.4), ni tampoco se encontró relación significativa entre la distancia genética y la distancia geográfica dentro del Mediterráneo (Figura S3.1).

El alto grado de diferenciación genética que muestra el atún blanco del Mediterráneo respecto al resto de poblaciones refleja el aislamiento de esta población, indicando que el Estrecho de Gibraltar supone una fuerte barrera al flujo génico para el atún blanco, aunque otras especies migradoras como son el atún rojo del Atlántico y el pez espada (*Xiphias gladius*) sean capaces de atravesar esta barrera oceánica (Block *et al*. 2005; Alvarado Bremer et al. 2005; Aranda *et al.* 2013). Este hecho está apoyado por los estudios de marcado y recaptura, donde se describe la migración entre el Atlántico y el Mediterráneo como prácticamente nula (Arrizabalaga *et al.* 2004). Sin embargo, la homogeneidad descrita a lo largo del Mediterráneo en el ESTUDIO II indicaría que, según nuestro estudio, ninguna otra barrera oceánica supone una interrupción para el flujo génico para la especie dentro de dicho mar. Es difícil de evaluar si la diferenciación que muestra la población del Mediterráneo es debida sólo a la restricción del flujo génico actual, o si refleja también la historia demográfica del atún blanco de esta región. Según Kettle *et al.* (2011) el Mediterráneo sirvió de refugio para muchas especies en el último máximo glacial (LGM). De

acuerdo con esta última hipótesis, la población Mediterránea habría sido el resultado de un evento fundador mayor y habría estado aislada del resto de poblaciones durante largo tiempo, dando lugar al alto grado de diferenciación actual.

2.2. Océano Atlántico y océano Pacífico

Suzuki (1962) mediante estudios serológicos, Chow y Ushiama (1995) con RFLPs y, Takagi *et al.* (2001) y Montes *et al.* (2012) con microsatélites describieron una alta diferenciación genética entre el Atlántico y el Pacífico.

En cuanto a la estructuración dentro de cada océano, Chow y Ushiama (1995) no encontraron diferencias entre hemisferios, mientras que Takagi *et al.* (2001) y Montes *et al.* (2012) sí lo hacían. Mientras que Takagi *et al.* (2001) describieron diferencias entre el Atlántico Norte y Sur, y describieron diferencias significativas entre las tres muestras del Pacífico, Montes *et al.* (2012) no definieron subestructuración en el Atlántico pero sí entre el Norte y el Sur del Pacífico. El trabajo basado en microsatélites de Davies *et al.* (2011) encontró subestructuración tanto temporal como espacial en el Nordeste Atlántico.

En los ESTUDIOS I, II y III de la presente Tesis Doctoral las poblaciones del hemisferio norte y sur de los océanos tanto Atlántico como Pacífico se muestran homogéneas e independientes del resto (Figura 2.3; Figura 3.4; Tabla 4.5). Por tanto, los marcadores SNP tanto nucleares como mitocondriales indican que no existiría ninguna barrera que impidiese el flujo génico entre los hemisferios Norte y Sur.

En el ESTUDIO II, el océano Atlántico fue la región mejor muestreada con un total de 12 muestras, 774 individuos y 24 años de diferencia temporal máxima (Tabla 3.1). Los análisis llevados a cabo a partir de los 75 marcadores SNP no revelaron subestructuración espacial ni temporal en el Atlántico. Los valores de F_{ST} basados en los marcadores neutrales (Tabla 3.2) también apoyan la hipótesis de la homogeneidad en el Atlántico. La estabilidad temporal hallada en nuestro estudio hace que los resultados obtenidos en cuanto a estructuración genética sean robustos.

2.3. Océano Índico

Arrizabalaga *et al.* (2004) mediante el estudio de grupos sanguíneos y Montes *et al.* (2012) mediante microsatélites encontraron que las muestras del Índico se sitúan genéticamente más cercanas al Atlántico que al Pacífico. En el caso de Arrizabalaga *et al.* (2004) la muestra del Índico era de 20 individuos y en Montes *et al.* (2012) de 26, y puede que el tamaño muestral pequeño haya podido provocar un sesgo en el resultado. De todas formas, la misma muestra del Índico de Montes *et al.* (2012) fue analizada en el ESTUDIO I mediante SNPs nucleares y el resultado fue diferente,

ya que la población del Ìndico resultó ser independiente pero más cercana al Pacífico que al Atlántico.

Este mismo resultado fue corroborado con el nuevo panel de SNPs descrito en el ESTUDIO II y con un aumento sustancial en la representación muestral del Índico (N=136). En este estudio se corroboró la cercanía genética existente entre el Índico y el Pacífico Norte (Figura 3.4 y Tabla 3.2), con lo que no podemos descartar la existencia del flujo génico entre estas dos regiones a través del archipiélago malayo. En este estudio se abarcaron 4 años de muestreo en el Índico, lo que le da mayor robustez a los resultados obtenidos.

En definitiva, la actual estructura genética del atún blanco detectada con los marcadores nucleares y mitocondriales podría explicarse por un patrón filogenético continuo pero con una separación espacial parcial. Este modelo denominado categoría V por Avise *et al.* (1987), se da en especies que históricamente han presentado niveles de flujo génico intermedios entre poblaciones geográficas. El atún blanco del Mediterráneo, que es la población genéticamente más diferenciada dentro de la especie, constituiría una excepción a este escenario demográfico de relativa estabilidad poblacional.

3. Aplicación en la gestión del atún blanco

3.1. Definición y delimitación de stocks de atún blanco

Los atunes son especialmente vulnerables a la sobrepesca debido a varios aspectos biológicos como son la madurez sexual tardía, longevidad, conducta gregaria y su alto valor económico (Collette *et al.* 2011). Estas características hacen que la evaluación y gestión de estos grandes depredadores pelágicos resulte especialmente complicada.

Actualmente la gestión del atún blanco a nivel mundial se basa en la definición de 6 stocks: (i) Mar Mediterráneo, (ii) Atlántico Norte, (iii) Atlántico Sur, (iv) Índico, (v) Pacífico Norte y (vi) Pacífico Sur. Aunque el Mediterráneo es considerado como un único stock a efectos de la gestión pesquera por el ICCAT, varios trabajos basados en posibles zonas de puestas diferenciadas (Marano *et al.* 1999), y subestructuración genética (Davies *et al.* 2011, Montes *et al.* 2012) apuntan hacia la posible existencia de varios stocks dentro del Mediterráneo. Sin embargo, el ESTUDIO II de esta tesis, que analiza el tamaño muestral más elevado y el mayor número de marcadores hasta la fecha, y es el único que contempla un muestreo de todo el área de distribución de la especie en el Mediterráneo, no ha encontrado estructuración en esta región.

Lo mismo ocurre con el océano Índico, ya que se considera un único stock independiente por las RFMOs. El estudio de Yeh *et al.* (1996), sin embargo, basándose en datos morfológicos y secuencias mitocondriales, propuso la existencia de dos stocks en el océano Índico delimitados por

el meridiano 90ºE. Por lo contrario, Montes *et al.* (2012) y Arrizabalaga *et al*. (2004) observaron homogeneidad entre el atún blanco del Índico y del Atlántico. El muestreo en el océano Índico en ambos trabajos no es el óptimo (N=20 y N=26), y ambos trabajos sugieren la necesidad de un muestreo mayor en este área de estudio para poder llegar a conclusiones más robustas. En cuanto a los estudios de esta Tesis Doctoral, basados en SNPs nucleares y mitocondriales, podemos decir que coinciden con el planteamiento actual de un único stock independiente para el océano Índico. Sin embargo, la posibilidad de la existencia de flujo génico entre el Índico y el Pacifico Norte (Tabla 3.2) y la ausencia de muestras en la zona este del Índico (a partir del meridiano 90ºE), indican la necesidad de un análisis más robusto en esta región.

Respecto al Atlántico y al Pacífico la gestión se basa en la definición de stocks diferentes en el hemisferio Norte y el hemisferio Sur de ambos océanos. Diversos estudios han sugerido la existencia de subestructuración en el Pacífico y en el Atlántico: ausencia de migraciones entre hemisferios (Arrizabalaga 2003), diferencias en las estadísticas pesqueras (Nakamura 1969), estudios histológicos de las gónadas (Otsu y Uchida 1959), diferencias morfológicas (Kurogane y Hiyama 1958, 1959; Ishii 1965), y estudios genéticos basados en microsatélites (Takagi *et al.* 2001; Davies *et al.* 2011; Montes *et al*. 2012). Sin embargo, los resultados de esta Tesis Doctoral no apoyan, a nivel genético, esta diferenciación entre hemisferios ni en el Atlántico ni en el Pacífico. El estudio de Chow y Ushiama (1995) tampoco observa heterogeneidad dentro de los océanos Atlántico y Pacífico en los análisis de RFLP y secuencias mitocondriales.

En resumen, los resultados genéticos de esta Tesis Doctoral respaldan la existencia de al menos 4 poblaciones: Mediterránea, Atlántica, Pacífica e Índica que están separadas entre ellas en diferentes grados de aislamiento. Estos resultados apoyan los planes de gestión en los que las poblaciones oceánicas, incluido el Mediterráneo, se gestionan de modo separado, pero discrepa de la actual definición de seis stocks. Sin embargo, el modelo de gestión actual representa una aproximación conservativa que reduciría la sobrepesca involuntaria de algunas poblaciones.

3.2. Estimación del tamaño efectivo del atún blanco del Atlántico Norte

Debido a que el atún blanco del Atlántico Norte ha sufrido la sobrepesca durante décadas, con el objetivo de poder medir la "salud" genética de este stock, y comprobar si la variabilidad genética había disminuido con el paso de los años, en el ESTUDIO II se calculó el tamaño poblacional efectivo (N_e) para este stock. Aunque tradicionalmente los estudios que han realizado estimas de N_e se han basado en marcadores microsatélites o secuencias mitocondriales (Turner *et al.* 1999; Hauser y Carvalho 2008; Qiu *et al.* 2013), en este estudio se han utilizado marcadores SNP nucleares para dicho fin. Para optimizar los análisis se llevó a cabo un muestreo temporal exhaustivo en los que se definieron 7 cohortes (Figura 3.2) separados por un máximo de 23 años (1986-2009).

Para la estimación del N_e se utilizaron dos métodos de escala temporal diferente. Por un lado se calculó el N_e contemporáneo, que se define por la variabilidad en las frecuencias alélicas debida a la deriva génica. Con la definición de cohortes con una distancia temporal máxima de 23 años se evitó el posible sesgo en los resultados debido a la asunción de generaciones discretas en la especie. Por otro lado, se estimó el N_e histórico basado en el método coalescente. Esta estimación del N_e hace referencia a una escala temporal mayor que la contemporánea (Whiteley *et al.* 2010), anterior a la época de sobrepesca de la especie, es decir, un periodo evolutivo mucho más lejano.

Los tamaños poblacionales efectivos contemporáneo e histórico no resultaron significativamente diferentes (Tabla S3.3 y Tabla S3.4). Esta similitud entre ambas estimaciones de N_e indica que el tamaño poblacional del atún blanco en el Atlántico Norte en la actualidad y en el pasado lejano no fue tan diferente. Por otro lado, el ratio N_e/N_c (tamaño poblacional censal) estimado varía entre 2.62×10^{-3} y 9.83×10^{-3}, la razón por la que el ratio N_e/N_c sea tan bajo en las especies marinas puede ser explicada por las altas tasas de mortalidad en las fases pre-reclutamiento y que hacen que unos pocos individuos maduros sean los que contribuyan genéticamente a las siguientes generaciones (Whitlock y Barton (1997); Nunney (1999)).

Las estimaciones de N_e en los estudios de túnidos varían dependiendo de si se calcularon en base a secuencias mitocondriales (entre 317.929 y 407.900.000 individuos) o marcadores microsatélites (entre 500 y 44.903 individuos) (Riccioni *et al*. 2010; Qiu y Miyamoto 2011; Qiu *et al*. 2013; Nomura *et al*. 2014). Según Qiu *et al.* (2013) una posible explicación a esta discrepancia podría ser la diferencia en el comportamiento de machos y hembras en las especies de túnidos. Existen dos estimaciones de N_e para el atún blanco basadas en microsatélites: 23.250 individuos para el Atlántico y 20.183 individuos para el Pacífico (Qiu *et al*. 2013). Las estimaciones medias obtenidas de N_e en el ESTUDIO II son 13.267 para el contemporáneo y 16.729 para el histórico. Estas dos estimas no son significativamente diferentes, de todas formas sería conveniente introducir la monitorización de la diversidad genética con la estimación anual del N_e. En este sentido, en el ESTUDIO II se presenta un panel de 58 SNPs neutrales útiles para este fin. Nuestros resultados se sitúan en el mismo rango que las estimaciones llevadas a cabo con microsatélites por Qiu *et al.* (2013). Las diferencias que presentan unos y otros marcadores en la tasa mutacional, podría hacer que las estimaciones de N_e varíen según la utilización de un tipo marcador u otro. En todo caso es evidente que los SNPs suponen una herramienta igual de útil y fiable que los microsatélites para la estimación de tamaño poblacional efectivo de los túnidos.

En resumen, el ESTUDIO II evidencia que el N_e del atún blanco del Atlántico Norte se ha mantenido estable mientras que el tamaño censal ha sufrido variaciones más marcadas (Figura 3.5). Las estimaciones de N_e indican que los niveles de diversidad genéticas se han mantenido estables y por encima de los 500 individuos, definido como el mínimo crítico. Por consiguiente, se descartan eventos de cuello de botella importantes que pudiesen poner en peligro la viabilidad del stock del Atlántico Norte. De todas formas, es importante profundizar en el entendimiento de la

vulnerabilidad de las poblaciones frente a la perdida de la variabilidad genética. En este sentido, las estimaciones del ratio N_e/N_c se deben considerar con cautela, ya que la dinámica poblacional de las especies que presentan tamaños poblacionales tan elevados son difíciles de comprender (Flowers *et al.* 2002; Therkildsen *et al.* 2010). A este respecto, teóricamente, en las poblaciones salvajes el ratio N_e/N_c contemporáneo debería estar por encima de 0.1 (Nunney y Campbell 1993; Frankham 1995; Vucetich *et al.* 1997) para asegurar una buena salud genética. Sin embargo, varios trabajos han cuestionado la idoneidad de estimar N_e en especies que presentan grandes tamaños poblacionales (Poulsen *et al.* 2006) y hasta qué punto N_e/N_c el ratio reflejan la dinámica real de estas especies (Therkildsen *et al.* 2010; Flowers *et al.* 2002).

En cualquier caso, los resultados de estabilidad temporal basados en marcadores neutrales presentados en el ESTUDIO II de la presente Tesis Doctoral, pueden servir como referente para evaluaciones futuras del ratio N_e/N_c y monitorización del N_e para las poblaciones del atún blanco. Además, estos índices de abundancia independientes de la pesca, suponen una herramienta complementaria en la monitorización de las poblaciones pesqueras con el objetivo de implementar acciones relacionadas directamente con la gestión sostenible.

4. Relaciones filogenéticas y evolutivas del género *Thunnus*

4.1. Estructura poblacional del atún blanco basado en el ADNmt

Los análisis de las cinco regiones mitocondriales (CR, 12S, 16S, Cyt *b* y COI) como de los SNPs de las regiones mitocondriales codificantes mostraron la existencia de dos haplogrupos diferenciados en el atún blanco, coincidiendo con Viñas *et al.* (2004) que también definieron dos filogrupos en la especie mediante el análisis de las secuencias del CR y con Chow y Ushiama (1995), que categorizaron las muestras en dos grandes grupos mediante las frecuencias de RFLPs. Según nuestros análisis basados en 160 secuencias del CR la diferenciación entre los haplogrupos habría ocurrido hace 1.0-0.3 Ma (Figura 4.4).

En nuesto trabajo, el haplogrupo A presentó una media de los valores de distancias genéticas (0.034) menor que la del haplogrupo B (0.052), indicando que el haplogrupo A pudo surgir a partir del haplogrupo B. Las frecuencias de los haplogrupos siguieron una distribución geográfica diferencial: el Mar Mediterráneo y el Océano Atlántico presentaron frecuencias del haplogrupo A entre el 85.6% y el 91.1%, y el Océano Pacífico frecuencias del haplogrupo A entorno al 63% (Tabla 4.2). La distribución actual de los haplogrupos podría derivar de la evolución alopátrica de los haplogrupos, tal y como se ha descrito para otras especies de escómbridos (Alvarado Bremer *et al*. 1995, 1998; Chow *et al.* 2000).

Sin embargo, y a pesar de la distribución asimétrica de los haplogrupos, el nivel de flujo génico posterior necesario para lograr la distribución actual de las frecuencias de los haplogrupos es

incompatible con las distancias genéticas elevadas entre los Océanos descritas tanto con SNPs mitocondriales como nucleares, esto sugiere otro escenario filogeográfico alternativo para la especie. La hipótesis alternativa se basa en una continuidad filogenética y una separación espacial parcial, definida como categoría V en Avise *et al.* (1987), más acorde con especies con niveles de flujo génico intermedios. Según esta hipótesis la aparición de los dos linajes se daría a partir de la extinción estocástica de haplotipos, después, el aislamiento geográfico parcial del Atlantico y el Indo-Pacífico daría lugar a una distribución bimodal dispar de los linajes (Slatkin y Hudson, 1991; Harpending *et al.* 1998).

No obstante, los resultados obtenidos para el Mediterraneo serían una excepción a este escenario demográfico en el cual las poblaciones han mantenido un tamaño relativamente grande y estable en el tiempo. La acumulación de mutaciones en la CR dentro de cada haplogrupo (Tabla 4.5), la elevada frecuencia de individuos del Mediterráneo (34.9%) en un único haplotipo (H10) exclusivo (Tabla 4.4) y el grado de diferenciación entre poblaciones obtenido con SNPs nucleares (ESTUDIO I y II) indican que el Mar Mediterráneo es la población genéticamente más diferenciada de la especie. Estos datos indican que en el Mediterráneo pudo haber ocurrido un proceso de diferenciación posterior al surgimiento de los dos haplogrupos hace 1.0-0.3 Ma. Esta población se habría quedado aislada durante alguno de los últimos eventos glaciales del Pleistoceno, probablemente durante la glaciación de Riss o la glaciación de Würm, tal y como se ha sugerido para el pez vela (*Xiphias gladius*) (Alvarado-Bremer *et al.* 1995).

Los resultados de las diferencias genéticas (Φ_{ST}) entre stocks basadas en los SNPs mitocondriales (Tabla 4.6) son congruentes con los escenarios que hemos planteado, ya que muestran el alto grado de diferenciación del Mediterráneo, la homogeneidad dentro de los océanos y heterogeneidad entre sí. En definitiva, los SNPs de las regiones codificantes del genoma mitocondrial detectaron cuatro poblaciones en el atún blanco: mar Mediterráneo, Océano Atlántico, Océano Índico y Océano Pacífico, coincidiendo así con los obtenidos en el ESTUDIO II con SNPs nucleares.

En base a estos resultados se puede concluir que la estructura del atún blanco en cuatro poblaciones está apoyada en gran medida por los resultados con microsatélites de Montes *et al.* (2012), SNPs del genoma nuclear por (ESTUDIO I; ESTUDIO II) y SNPs mitocondriales (ESTUDIO III). No obstante, esta estructuración no se evidencia al analizar las secuencias de 5 fragmentos (5.112 bp) ni la region CR (344 bp) debido en gran medida al tamaño muestral menor (n= 38 y n=160 respectivamente, frente a n=1.254 para los SNPs) y debido también a que no se han podido utilizar SNPs de la región hipervariable CR. Finalmente, se ha podido describir un panel de SNPs mitocondriales capaz de inferir la estructuración poblacional de atún blanco, resultaría interesante aplicar estos marcadores en otras especies del género con el mismo fin.

4.2. Evolución del género Thunnus

La división del género *Thunnus* en los subgéneros *Neothunnus* y *Thunnus* propuesta por Collette (1978) en base a características morfoecológicas, no coincide con la filogenia del género descrita en el presente trabajo. Los resultados descritos en el ESTUDIO III concuerdan con los estudios que han analizado secuencias mitocondriales (Chow y Inou 1993; Chow y Kishino 1995; Viñas y Tudela 2009; Tseng *et al*. 2012) o nucleares (Chow *et al*. 2006) del género *Thunnus*. Los resultados del ESTUDIO III describen como, en primer lugar, se diferenciaron atún blanco y el atún rojo del Pacífico, constituyendo el grupo más antiguo del género. En segundo lugar, se agruparon las tres especies constitutivas del subgénero *Neothunnus* (el atún de aleta negra (*Thunnus atlanticus*), el rabil (*Thunnus albacares*) y el atún tongol (*Thunnus tonggol*)) que se diferenció de las otras tres especies del subgénero *Thunnus* (el atún rojo del Atlántico, el atún rojo del sur (*Thunnus maccoyii*) y el patudo (*Thunnus obesus*)).De este modo el presente trabajo apoyara el origen monofilético del subgénero *Neothunnus,* pero define como parafilético el subgénero *Thunnus*, tal y como está actualmente aceptado.

Por otro lado, al analizar un extenso número de secuencias del CR de las especies del género hemos apreciado que un número de haplotipos pertenecientes en principio a una especie no aparecían situados junto a individuos de su especie, sino que en haplogrupos de diferentes especies. Estos agrupamientos, en principio "equivocados" son individuos que contienen un genoma mitocondrial perteneciente a una especie diferente al genoma nuclear, es decir, son individuos "híbridos". (Figuras. 4.3 y 4.4). La existencia de episodios de hibridación (o de introgresión) entre especies del género *Thunnus*, han sido descritas por varios autores (Chow y Kishino 1995; Carlsson *et al*. 2004; Alvarado Bremer *et al.* 2005; Lowenstein *et al*. 2009; Viñas y Tudela 2009). Resulta destacable la localización de las secuencias de los atunes rojos del Atlántico y del Pacífico y el atún blanco. En esta Tesis Doctoral proponemos una introgresión histórica del ADNmt del atún blanco en el ancestro común de los atunes rojos del Atlántico y del Pacífico. Por un lado, la mayoría de los haplotipos de los atunes rojos del Pacífico y algunos del Atlántico (a) se agrupan junto a las secuencias del atún blanco (Figura 4.3) indicando que estos tres grupos de haplotipos descienden de la primera separación de especies del género *Thunnus*. Por otro lado, la ausencia de áreas de contacto entre las especies de atún rojo del Atlántico y del Pacífico descartan un posible evento introgresivo reciente ente ellos, esto indicaría que las dos secuencias de atún rojo del Pacífico (nº 9 y 10 en la Figura 4.3) localizadas dentro del haplogrupo (b) del atún rojo del Atlántico sería el resultado de un evento introgresivo antiguo. Por último, que los atunes rojos del Atlántico y del Pacífico fuesen consideradas de la misma especie hasta hace poco (Collete 1999) debido a que fenotípicamente no difieran y que comparten secuencias nucleares del gen ITS1 (Chow *et al.* 2006), apoyaría la idea de un ancestro común relativamente reciente entre estas dos especies.

La menor diversidad genética detectada en los haplotipos del atún rojo del Atlántico cercanos al atún blanco, indicaría que estos serían los derivados de la introgresión histórica, mientras que el ADNmt descendiente del ancestro común de atunes rojos del Atlántico y del Pacífico sería el correspondiente al haplogrupo (b) del atún rojo del Atlántico situado junto a la mayoría de las especies de *Thunnus*. En este sentido, el agrupamiento de haplotipos del atún rojo del Atlántico (a) y la mayoría de haplotipos del atún rojo del Pacífico muestran escasos grandes haplotipos centrales y más haplotipos que se extienden periféricamente. Este tipo de formas se asocian a eventos de efecto fundador a partir de pocos individuos seguido de una expansión con mutaciones aleatorias. En base a estas evidencias parece que las hibridaciones que actualmente observamos debieron ocurrir a partir de cruces entre machos del ancestro común de atunes rojos del Atlántico y del Pacífico y hembras de atún blanco. El cruzamiento entre híbridos en las dos especies de atunes rojos aseguró la perduración del genoma mitocondrial del atún blanco en las otras dos especies hasta el presente.

La hibridación dentro del género *Thunnus* parece ser un proceso común, así lo demuestran las 24 secuencias de tres especies diferentes ubicadas en clados que no pertenecían a la especie a la cual pertenecen fenotípicamente (Figura 4.3 y Tabla 4.6). Los tiempos de divergencia estimados para estas secuencias "descolocadas" se encuentran dentro del rango de variación correspondiente al clado en el que se sitúan en la figura network (excepto las secuencias de las especies de atún rojo anteriormente mencionadas), lo cual indica que son eventos introgresivos relativamente recientes. Entre ellos, destacan las 18 secuencias de atún rojo del Atlántico dentro del haplogrupo del atún blanco y 2 secuencias de atún rojo del Atlántico dentro del haplogrupo del rabil. Estos son probablemente el resultado de hibridaciones recientes entre machos de atún rojo del Atlántico y hembras de atún blanco (n=18) y rabil (n=2).

Al emplear un reloj molecular que asume una tasa de divergencia del 3.6% por Ma, propuesta para la CR en peces (Donaldson y Wilson 1999), la separación del atún blanco respecto al resto de especies del género habría sucedido hace unos 4.5 Ma. Entre 2.7 y 2.1 Ma habría ocurrido un proceso cladogenético en el cual habrían surgido todas las especies restantes del género, y se habrían originado también los dos linajes mitocondriales del patudo y el atún tongol. Dentro de ese rango temporal se situaría la aparición de la especie antepasada de las actuales especies morfológicamente hermanas atún rojo del Atlántico y del Pacífico, y no mucho después de 2.7 Ma habría ocurrido la introgresión del ADNmt del atún blanco en el antecesor común de los dos atunes rojos mencionados tal y como hemos descrito. Además las diferencias genéticas entre los haplotipos cercanos al atún blanco (Figura 4.3) permiten datar en 0.8 Ma la divergencia entre el atún rojo del Atlántico y el atún rojo del Pacífico. Probablemente sucedió en algún periodo interglaciar del Pleistoceno, cuando las temperaturas y nivel del mar subieron y las poblaciones de atún pudieron moverse a través del océano Ártico. Tal y como sugiere la mayor diversidad genética encontrada en el atún rojo del Atlántico, el ancestro común de ambas especies habría vivido en el Atlántico y tras

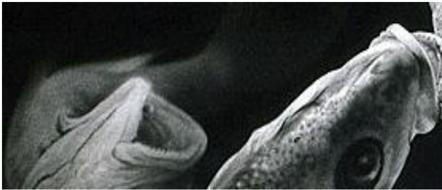
un efecto fundador pudo haber poblado el océano Pacífico dando lugar así al atún rojo del Pacífico.

5. Desarrollo de una herramienta genética para la autentificación del origen del atún blanco

La selección de los SNPs más informativos descubiertos y validados en los ESTUDIOS I, II y III dieron como resultado un panel de SNPs útil para la trazabilidad de productos derivados del atún blanco. Se ha obtenido una herramienta que asigna correctamente el 100% de los individuos que se consiguen discriminar genéticamente entre los orígenes Mediterráneo, Atlántico e Indo-Pacífico (el 68% de los individuos testados).

Con el objetivo de desarrollar una marca del atún blanco del Atlántico, la Universidad del País Vasco y AZTI desarrollaron y publicaron una patente (ES 2 392 293 B1). Sin embargo, esta herramienta sólo conseguía clasificar el 45.5% de los individuos, de los cuales el 91.2 % se clasificaba de forma correcta. Con la herramienta que se propone en el presente trabajo, se aumenta notablemente el porcentaje de individuos clasificados (del 45.5% al 67.8%) y el 100% de individuos quedan correctamente clasificados, por lo que se ha mejorado notoriamente la herramienta de trazabilidad. Además, se ha conseguido alcanzar el objetivo de desarrollar un panel con el mínimo número de marcadores que logre el mayor poder discriminatorio, ya que de los 137 SNPs iniciales se consiguió reducir el panel a tan solo 32 SNPs.

Este panel de SNPs es muy útil para alcanzar dos objetivos: 1) comprobar que el etiquetado de cualquier producto, fresco o tratado, derivado del atún blanco cumple con la normativa establecida en cuanto a la identificación de la especie y zona de captura (Mediterráneo, Atlántico o Índico-Pacífico); 2) comprobar si las capturas desembarcadas en cualquier punto han sido pescadas en la zona de captura en las que se han declarado.



Conclusiones

CONCLUSIONES

METODOLOGÍA

- Las estrategias clásicas de secuenciación Sanger, EPIC-PCR y primers degenerados de secuencias conservadas en teleósteos, han resultado ser efectivas y satisfactorias para el descubrimiento de SNPs en el atún blanco. Se ha ampliado la cobertura del genoma proporcionando 54 secuencias nucleares y 5 mitocondriales, además de 151 SNPs validados y de acceso público para el análisis genético-poblacional de la especie.
- La amplificación cruzada entre el atún blanco y el atún rojo del Atlántico también ha proporcionado un alto número de marcadores SNP del transcriptoma, que aplicados en una amplia muestra han dibujado de forma robusta la estructura poblacional mundial del atún blanco, y en las series temporales han permitido la estimación del tamaño efectivo poblacional (N_e).

ESTRUCTURA GENETICA

- El atún blanco presenta una estructuración poblacional evidente constituida por al menos cuatro poblaciones: i) Mediterránea, ii) Atlántica, iii) Índica y iv) Pacífica. Así, el análisis genético, agrupa los stocks de los hemisferios norte y sur tanto en el Atlántico, como en el Pacífico, reduciendo de este modo a cuatro los seis stocks establecidos por las Organizaciones Regionales de Ordenación Pesquera. En principio, la gestión de una población en dos stocks diferenciados no tendría necesariamente consecuencias negativas, máxime considerando la existencia de zonas de puesta diferenciadas y la aparente ausencia de migración entre ellas. No obstante, resulta muy recomendable la coordinación entre gestores de ambos hemisferios.
- La población del Mediterráneo es genéticamente la más singular de las poblaciones debido muy probablemente a que el estrecho de Gibraltar supone una fuerte barrera al flujo génico. El Índico y el Pacífico muestran la mayor similitud, tal vez reflejo del flujo génico acontecido a través del archipiélago Malayo.
- La presencia y distribución de dos grandes haplogrupos (A y B) en el atún blanco coincide con el patrón filogeográfico de aislamiento genético parcial entre el Atlántico y el Indo-Pacífico, y el alto grado de aislamiento del Mediterráneo. La diferenciación de la población Mediterránea respecto al resto queda patente por la presencia mayoritaria del haplogrupo A en el Mediterráneo (>90%), la existencia de un haplotipo exclusivo que incluye al 35% de los individuos del Mediterráneo y la mayor diferenciación (Φ_{ST}) del mismo respecto al resto.

GENÉTICA DE LA CONSERVACION

- El ratio N_e/N_c obtenido mediante cohortes del Atlántico norte indica que un número relativamente pequeño de individuos da lugar a la descendencia, debido probablemente a la alta variación en la supervivencia inherente a las fases de prerreclutamiento.
- A pesar de que el stock del Atlántico norte ha sido sobreexplotado y el número censal se ha visto afectado, el descenso del tamaño efectivo de la población (N_e) no es significativo, indicando que la diversidad genética de la especie no ha sufrido un descenso drástico a lo largo de su historia más reciente. De todos modos, sería recomendable la monitorización del N_e a fin de establecer si su reducción es una tendencia estable con consecuencias a largo plazo, en cuyo caso resultaría importante tomar medidas inmediatas para paliar la sobrepesca.

RELACIONES FILOGENÉTICAS EN EL GÉNERO *Thunnus*

- El origen monofilético del subgénero *Neothunnus* queda patente por el análisis de secuencias de diferentes regiones del ADN mitocondrial, mientras que el subgénero *Thunnus* sería parafilético. Por lo tanto, queda demostrado que los datos moleculares no apoyan la actual división del género *Thunnus* basado en características morfoecológicas.
- El análisis de un gran número de secuencias del CR pertenecientes a las ocho especies del género *Thunnus* ha permitido constatar que los eventos introgresivos son abundantes dentro del género, y se han podido constatar sucesos de hibridación entre especies que no se habían descrito hasta el momento. Proponemos un evento de hibridación histórico del ADN mitocondrial del atún blanco en el antecesor común de las especies de atún rojo del Atlántico y del Pacífico, y la separación de estos últimos hace unos 0.8 Ma.
- Asumiendo el reloj molecular de 3.6% de divergencia por Ma, las dataciones de 4.5 Ma planteada para la separación del atún blanco del resto del género, 4.0 Ma para la separación del patudo y, entre 2.7 y 2.1 Ma para el resto de eventos de especiación son coincidentes con las glaciaciones globales que pudieron condicionar dichos eventos.

TRAZABILIDAD

- Hemos desarrollado una herramienta genética altamente discriminante para la autentificación del origen poblacional del atún blanco. Está compuesta por 32 SNPs, y mediante su aplicación el 68% de los individuos testados fueron asignados a su origen geográfico con un 100% de fiabilidad.

Bibliografía

BIBLIOGRAFIA

Ackerman MW, Habicht C, Seeb LW (2011) Single-Nucleotide Polymorphisms (SNPs) under diversifying selection provide Increased accuracy and precision in mixed-stock analyses of sockeye salmon from the Copper River, Alaska. *T Am Entomol Soc* **140**: 865-381.

Agnew DJ, Pearce J, Pramod G, Peatman T, Watson R et al. (2009). Estimating the worldwide extent of illegal fishing. *PLoS ONE* **4**: e4570.

Albaina A, Iriondo M, Velado I, Laconcha U, Zarraonaindia I et al. (2013) Single nucleotide polymorphism discovery in albacore and Atlantic bluefin tuna provides insights into worldwide population structure. *Anim Genet* **44**(6): 678-92. doi:10,1111/age.12051.

Alvarado Bremer JR, Baker AJ, Mejuto J (1995). Mitochondrial DNA control region sequences indicate extensive mixing of swordfish (*Xiphias gladius*) populations in the Atlantic Ocean. *Can J Fish Aquat Sci* **52**: 1720-1732.

Alvarado Bremer JR, Naseri, I Ely B (1997). Orthodox and unorthodox phylogenetic relationships among tunas revealed by the nucleotide sequence analysis of the mitochondrial DNA control region. *J Fish Biol*, 50: 540-554.

Alvarado Bremer JR, Stéquert B, Robertson NW, Ely B (1998). Genetic evidence for inter-oceanic subdivision of bigeye tuna (*Thunnus obesus*) populations. *Mar Biol* **132**: 547-557.

Alvarado Bremer JR, Viñas J, Mejuto J, Ely B, Pla C (2005) Comparative phylogeography of Atlantic bluefin tuna and swordfish: the combined effects of vicariance, secondary contact, introgression, and population expansion on the regional phylogenies of two highly migratory pelagic fishes. *Mol Phylogenet Evol* **36**: 169-187.

André C, Larsson LC, Laikre L, Bekkevold D, Brigham J et al. (2011) Detecting population structure in a high gene-flow species, Atlantic herring (*Clupea harengus*): direct, simultaneous evaluation of neutral vs putatively selected loci. *Heredity* **106**: 270-280.

Antao T, Lopes A, Lopes RJ, Beja-Pereira A, Luikart G (2008) LOSITAN: A workbench to detect molecular adaptation based on a FST-outlier method. *BMC Bioinformatics* **9**: 323-327.

Aranda G, Abascal FJ, Varela JL, Medina A (2013) Spawning behaviour and post-spawning migration patterns of Atlantic bluefin tuna (*Thunnus thynnus*) ascertained from satellite archival tags. *PLoS ONE* **8**: e76445.

Arena P, Potoschi A, Cefali A (1980) Risultati prelimiari di studi sull'etá, l'accrescimento a la prima maturitá sessuale dell'alalunga *Thunnus alalunga* (Bonn., 1788) del Tirreno. *Mem Biol Mar Oceanogr* **10**: 71-81

Arrizabalaga H, Lopez-Rodas V, Oritz de Zarate V, Costas E, Gonzalez-Garcés A (2002) Study on the Migrations and Stock Structure of Albacore (*Thunnus alalunga*) from the Atlantic Ocean and the Mediterranean Sea Based on Conventional Tag Release- Recapture Experiences. *ICCAT Col Vol Sci Pap* **54**: 1479-1494.

Arrizabalaga H (2003) Estructura poblacional del atún blanco (*Thunnus alalunga* Bonn. 1788): una aproximación multidisciplinar. PhD thesis. University of Vigo

Arrizabalaga H, Costas E, Juste J, González-Garcés A, Nieto B et al. (2004) Population structure of albacore *Thunnus alalunga* inferred from blood groups and tag-recapture analyses. *Mar Ecol Prog Ser* **282**: 245−252

Arrizabalaga H, Lopez-Rodas V, Costas E, Gonzalez-Garcés A (2007) Use of genetic data to assess the uncertainty in stock assessments due to the assumed stock structure: the case of albacore (*Thunnus alalunga*) from the Atlantic Ocean. *Fish Bull* **105**: 140-146.

Arrizabalaga H, Dufour F, Kell L, Merino G, Ibaibarriaga L et al. (2014) Global habitat preferences of commercially valuable tuna *Deep-Sea Res Pt II* **113***:* 102-112.

Avise JC, Arnold J, Ball RM, Bermingham E, Lamb T et al (1987). Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annu Rev Ecol Syst* **1987**:489-522.

Bard FX (1981) La thon germon (*Thunnus alalunga*) de l'Ocean Atlantique De la dynamique de population à la stratégie démographique Thèse Doctorat ès Sciences Naturelles, Universitè de Paris VI, 330P.

Bartlett SE, Davidson WS (1991). Identification of *Thunnus* tuna species by the polymerase chain reaction and direct sequence analysis of their mitochondrial cytochrome *b* genes. *Can J Fish Aquat Sci* **48**: 309-317.

Beardsley GL (1969) Proposed Migrations of Albacore, *Thunnus alalunga*, in the Atlantic. *Ocean T Am Fish Soc* **98**: 589-598.

Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. *P Roy Soc Lond B Bio* **263**: 1619-1626.

Beddington JR, Agnew DJ, Clark CW (2007) Current problems in the management of marine fisheries. *Science* **316**: 1713-1716.

Beerli P (2002) MIGRATE: documentation and program, part of LAMARC, Version 1.5 Available at http://evolution.genetics/washington.edu/lamarc.html

Beerli P, Felsenstein J (2001) Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. *P Natl A Sci* **98**: 4563-4568.

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Roy Stat Soc B Met* **57**: 289-300.

Bensch S, Åkesson S, Irwin DE (2002). The use of AFLP to find an informative SNP: genetic differences across a migratory divide in willow warblers. *Mol Ecol* **11**: 2359-2366.

Benton MJ. (Ed.) (1993). The fossil record 2 (2nd ed., pp. 681-715). London: Chapman & Hall.

Berger SA. Stamatakis A (2011). Aligning short reads to reference alignments and trees. *Bioinformatics* **27**: 2068-2075.

Block BA, Dewar H, Blackwell SB, Williams TD, Prince ED et al. (2001) Migratory movements, depth preferences, and thermal biology of Atlantic bluefin tuna. *Science* **293**: 1310-1314.

Block BA, Teo SL, Walli A, Boustany A, Stokesbury MJ et al. (2005) Electronic tagging and population structure of Atlantic bluefin tuna. *Nature* **434**: 1121-1127.

Block BA, Dewar H, Blackwell SB, Williams TD, Prince ED et al. (2001). Migratory movements, depth preferences, and thermal biology of Atlantic bluefin tuna. *Science* **293**: 1310-1314.

Block BA, Teo SL, Walli A, Boustany A, Stokesbury et al. (2005). Electronic tagging and population structure of Atlantic bluefin tuna. *Nature* **434**: 1121-1127.

Bohonak AJ, Vandergast AG (2011). The value of DNA sequence data for studying landscape genetics. *Mol Ecol* **20**: 2477-2479.

Boustany AM, Reeb CA, Block BA (2008) Mitochondrial DNA and electronic tracking reveal population structure of Atlantic bluefin tuna (*Thunnus thynnus*). *Mar Biol* **156**: 13-24.

Bravington MV, Grewe PM, Davies CR (2014) Fishery-independent estimate of spawning biomass of Southern Bluefin Tuna through identification of close-kin using genetic markers. FRDC Report 2007/034 CSIRO, Australia.

Brumfield RT, Beerli P, Nickerson DA, Edwards SV (2003). The utility of single nucleotide polymorphisms in inferences of population history. *Trends Ecol Evol* **18**: 249-256.

Carlson CS, Eberle MA, Rieder MJ, Yi Q, Kruglyak L et al. (2004). Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. *Am J Hum Genet* **74**: 106-120.

Carlsson J, McDowell JR, Carlsson JEL, Graves JE (2007) Genetic Identity of YOY Bluefin Tuna from the Eastern and Western Atlantic Spawning Areas. *J Hered* **98**: 23-28.

Carroll R L (1988). Vertebrate paleontology and evolution. *Freeman*.

Castillo AG, Martinez JL, Garcia-Vazquez E (2004) Fine spatial structure of Atlantic hake (*Merluccius merluccius*) stocks revealed by variation at microsatellite loci. *Mar Biotechnol* **6**: 299-306.

Childers J, Snyder S, Kohin S (2011) Migration and behavior of juvenile North Pacific albacore (*Thunnus alalunga*). *Fish Oceanogr* **20**: 157-173.

Chow S, Inoue S (1993). Intra-and interspecific restriction fragment length polymorphism in mitochondrial genes of *Thunnus* tuna species. *Bull Natl Res Inst Far Seas Fish* **30**: 207-225.

Chow S, Kishino H (1995). Phylogenetic relationships between tuna species of the genus *Thunnus* (Scombridae: Teleostei): inconsistent implications from morphology, nuclear and mitochondrial genomes. *J Mol Evol* **41**: 741-748.

Chow S, Ushiama H (1995) Global population structure of albacore (*Thunnus alalunga*) inferred by RFLP analysis of the mitochondrial ATPase gene. *Mar Biol* **123**: 39-45.

Chow S, Okamoto H, Miyabe N, Hiramatsu K, Barut N (2000). Genetic divergence between Atlantic and Indo-Pacific stocks of bigeye tuna (*Thunnus obesus*) and admixture around South Africa. *Mol Ecol* **9**: 221-227.

Chow S, Nakagawa T, Suzuki N, Takeyama H, Matsunaga T (2006) Phylogenetic relationships among *Thunnus* species inferred from rDNA ITS1 sequence. *J Fish Biol* **68**: 24-35.

Collette BB (1999) Mackerels, molecules, and morphology. *Proc. 5th Indo-Pac. Fish Conf, Soc. Fr. Ichtyol* **1999**:149–164.

Collette BB, Nauen C (1983) FAO species catalogue. Vol. 2 Scombrids of the world. An annotated and illustrated catalogue of tunas, mackerels, bonitos and related species known to date. FAO *Fisheries Synopsis* **125** Volume 2. FAO, Rome.

Collette BB, Carpenter KE, Polidoro BA, Juan-Jordá MJ, Boustany A et al. (2011) High Value and Long Life—Double Jeopardy for Tunas and Billfishes. *Science* **333**: 291-292.

Collette BB (1978) Adaptation and systematics of the mackerels and tunas. In: Sharp GD, Dizon AE (eds) The physiological ecology of tunas. *Academic Press* New York: 7-39

Cort JL, Arregui I, Estruch VD, Deguara S (2014). Validation of the Growth Equation Applicable to the Eastern Atlantic Bluefin Tuna, *Thunnus thynnus* (L.), Using L max, Tag-Recapture, and First Dorsal Spine Analysis. *Rev Fish Sci Aqu* **22**: 239-255.

Davies CA, Gosling EM, Was A, Brophy D, Tysklind N (2011) Microsatellite analysis of albacore tuna (*Thunnus alalunga*):population genetic structure in the North-East Atlantic Ocean and Mediterranean Sea. *Mar Biol* **158**: 2727-2740.

De Roos AM, Persson L (2002) Size-dependent life-history traits promote catastrophic collapses of top predators. *P Natl Acad Sci USA* **99**: 12907-12912.

DeWoody JA, Avise JC (2000). Microsatellite variation in marine, freshwater and anadromous fishes compared with other animals. *J Fish Biol* **56**: 461-473.

Diaz M, Wethey D, Bulak J, Ely B (2000). Effect of harvest and effective population size on genetic diversity in a striped bass population. *T Am Fish Soc* **129**: 1367-1372.

Dicenta A, Piccinetti C, Piccinetti Manfrin G (1975) Observaciones sobre la reproducción de los túnidos en las islas Baleares. *Bol Inst Espa Oceanogr* **204**: 25-37.

Dickhut RM, Deshpande AD, Cincinelli A, Cochran MA, Corsolini S, et al. (2009) Atlantic Bluefin Tuna (*Thunnus thynnus*) population dynamics delineated by organochlorine tracers. *Environ Sci Technol* **43**: 8522-8527.

Donaldson KA, Wilson RR (1999). Amphi-panamic geminates of snook (Percoidei: Centropomidae) provide a calibration of the divergence rate in the mitochondrial DNA control region of fishes. *Mol Phylogenet Evol* **13**: 208-213.

Duclerc J, Sacchi J, Piccinetti C, Piccinetti Manfrin G, Dicenta A, et al. (1973) Nouvelles donne sur la reproduction du thon rouge (*Thunnus thynnus* L.) et d'autres espèces de Thonidés en Méditerranée. *Rev Trav Inst Peches Marit* **37**: 2.

Edgar RC (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**: 1792-1797.

Efron B (1983) Estimating the error rate of a prediction rule: improvement on cross-validation. *J Am Stat Assoc* **78**: 316-331. doi:10,1080/01621459.1983.10477973.

Elliott NG, Ward RD (1995). Genetic relationships of eight species of Pacific tunas (Teleostei: Scombridae) inferred from allozyme analysis. *Mar Freshwater Res* **46**: 1021-1032.

Ely B, Stoner DS, Bremer AJ, Dean JM, Addis P, et al. (2002) Analyses of nuclear ldhA gene and mtDNA control region sequences of Atlantic Northern bluefin tuna populations. *Mar Biotechnol* **4**: 583-588.

Faber JE, Stepien CA (1997). The utility of mitochondrial DNA control region sequences for analyzing phylogenetic relationships among populations, species, and genera of the Percidae. *Mol Syst Fish*, 129-143.

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **164**: 1567-1587.

FAO 2011 Review of the state of world marine fishery resources. *FAO Fisheries and Aquaculture Technical Paper* **569** Rome, 2011

FAO 2014 Fisheries and aquaculture software FishStatJ - software for fishery statistical time series In: FAO Fisheries and Aquaculture Department [online] Rome Updated 22 July 2014 http://www.fao.org/fishery/statistics/software/fishstatj/en

Farley JH, Williams AJ, Hoyle SD, Davies CR, Nicol SJ (2013) Reproductive Dynamics and Potential Annual Fecundity of South Pacific Albacore Tuna (*Thunnus alalunga*). *PLoS ONE* **8**: e60577.

Felsenstein J (1971) Inbreeding and variance effective numbers in populations with overlapping generations. *Genetics* **68**: 581-597.

Flowers, JM, Schroeter SC, Burton RS (2002) The recruitment sweepstakes has many winners: Genetic evidence from the sea urchin *Strongylocentrotus purpuratus*. *Evolution* **56**:1445-1453.

Foll M, Gaggiotti OE (2008) A genome scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics* **180**: 977-993.

Fontaine MC, Baird SJE, Piry S, Ray N, Tolley KA et al. (2007) Rise of oceanographic barriers in continuous populations of a cetacean: the genetic structure of harbour porpoises in Old World waters. *BMC Biol* **5**: 30.

Frankham R (1995) Effective population-size adult population size ratios in wildlife: A review. *Genet Res* **66**: 95-107.

Fromentin JM, Fonteneau A (2001) Fishing effects and life history traits: A case study comparing tropical versus temperate tunas. *Fisheries Res* **53**: 133-150.

Fromentin JM, Powers JE (2005) Atlantic bluefin tuna: Population dynamics, ecology, fisheries, and management. *Fish Fish* **6**: 281-306.

Fromentin JM (2009) Lessons from the past: Investigating historical data from bluefin tuna fisheries. *Fish Fish* **10**: 197-216.

Fromentin JM, Bonhommeau S, Arrizabalaga H, Kell LT (2014) The spectre of uncertainty in management of exploited fish stocks: The illustrative case of Atlantic bluefin tuna Mar Policy **47**: 8-14.

Fu C, Fanning LP (2004). Spatial considerations in the management of Atlantic cod off Nova Scotia, Canada. N Am J Fish Man **24**: 775-784.

Fu YX (1997). Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147**: 915-925.

Gaggiotti OE, Bekkevold D, Jørgensen HBH, Foll M, Carvalho GR et al (2009) Disentangling the effect of evolutionary, demographic and environmental factors influencing genetic structure of natural populations: Atlantic herring as a case study. *Evolution* **63**: 2939-2951.

Galuardi B, Royer F, Golet W, Logan J, Neilson J et al. (2010) Complex migration routes of Atlantic bluefin tuna (*Thunnus thynnus*) question current population structure paradigm. *Can J Fish Aquat Sci* **67**: 966-976.

García A, Alemany F, Rodríguez JM (2002) Distribution of tuna larvae off the Balearic sea: preliminary results of the TUNIBAL 0600 larval survey. *ICCAT Col Vol Sci Pap* **54**: 554-560

García A, Cortés D, Ramírez T, Fehri-Bedoui R, Alemany F et al. (2006) First data on growth and nucleic acid and protein content of field-captured Mediterranean bluefin (*Thunnus thynnus*) and albacore (*Thunnus alalunga*) tuna larvae: a comparative study. *Sci Mar* **70**: 67-78.

García SM, Grainger RJ (2005) Gloom and doom? The future of marine capture fisheries. *Philos T Roy Soc B* **360**: 21-46.

Garvin MR, Saitoh K, Gharrett AJ (2010) Application of single nucleotide polymorphisms to non-model species: a technical review. *Mol Ecol Resour* **10**: 915-934.

Gattepaille LM, Jokobsson M (2012) Combining markers into haplotypes can improve population structure inference. *Genetics* **190**: 159-174.

Gemmell NJ, Schwartz MK, Robertson BC (2004) Moa were many. *P Roy Soc Lond B Bio* **271**: 430-432.

Gibbard PL, Head MJ, Walker MJ (2010) Formal ratification of the Quaternary System/Period and the Pleistocene Series/Epoch with a base at 2.58 Ma. *J Quaternary Sci* **25**: 96-102.

Gibbs RHJ, Collette BB (1967) Comparative anatomy and systematics of the tunas, genus *Thunnus*. *Fish Wildl Ser Fish Bull* **66**:65-130

Godsil HC (1948) A preliminary population study of the yellowfin and the albacore. *Calif Dep Fish Game Fish Bull* **70**:1-90

Gomez-Uchida D, Banks MA (2006) Estimation of Effective Population Size for the Long-Lived Darkblotched Rockfish (*Sebastes crameri*). *J Herd* **97**: 603-606.

Goñi N, Logan J, Arrizabalaga H, Jarry M, Lutcavage M (2011) Variability of albacore (*Thunnus alalunga*) diet in the Northeast Atlantic and Mediterranean Sea. *Mar Biol* **158**: 1057-1073.

Goudet J (2001) FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3) Available at: http://www2.unil.ch/popgen/softwares/fstat.htm

Graves JE, Dizon AE (1989) Mitochondrial DNA sequence similarity of Atlantic and Pacific albacore tuna (*Thunnus alalunga*). *Can J Fish Aquat Sci* **46**: 870-873.

Guichoux E, Lagache L, Wagner S, Chaumeil P, Léger P, et al (2011) Current trends in microsatellite genotyping. *Mol Ecol Resour* **11**: 591-611.

Guillot G, Santos F, Estoup A (2008) Analysing georeferenced population genetics data with Geneland: a new algorithm to deal with null alleles and a friendly graphical user interface. *Bioinformatics* **24**: 1406-1407.

Haasl R, Payseur B (2011) Multi-locus inference of population structure: a comparison between single nucleotide polymorphisms and microsatellites. *Heredity* **106**: 158-171.

Hall TA (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic acids sym se* **41**: 95-98.

Hallaire L, Dao JC 1971 Etude serologique du germon du nord-est Atlantique (résultats préliminaires). *ICCAT* **35**

Hare MP, Nunney L, Schwartz MK, Ruzzante DS, Burford M et al. (2011) Understanding and Estimating Effective Population Size for Practical Application in Marine Species Management. *Conserv Biol* **25**: 438-449.

Harpending, HC, Batzer MA, Gurven M, Jorde LB, Rogers AR et al. (1998). Genetic traces of ancient demography. *P Natl Acad Sci* **95**: 1961-1967.

Hauser L, Seeb JE (2008). Advances in molecular technology and their impact on fisheries genetics. *Fish Fish* **9**: 473-486.

Hauser L, Adcock GJ, Smith PJ, Bernal Ramirez JH, Carvalho GR (2002) Loss of microsatellite diversity and low effective population size in an overexploited population of New Zealand snapper (*Pagrus auratus*). *P Natl Acad Sci USA* **99**: 742-747.

Hauser L, Carvalho GR (2008) Paradigm shifts in marine fisheries genetics: ugly hypotheses slain by beautiful facts. *Fish Fish* **9**: 333-362.

Hayes B, Lærdahl J, Lien S, Berg P, Davidson W et al. (2004) Detection of single nucleotide polymorphisms (SNPs) from Atlantic salmon expressed sequence tags (ESTs). In: Proceedings of the 55th Annual Meeting of the European Association for Animal Production: 5-9 September 2004; Bled, Slovenia. European Federation of Animal Science.

Helyar SJ, Hemmer-Hansen J, Bekkevold D, Taylor MI, Ogden R et al. (2011) Application of SNPs for populations genetics of non-model organism: new opportunities and challenges. *Mol Ecol Resour* **11**: 123-136.

Helyar SJ, Limborg MT, Bekkevold D, Babbucci M, Van Houdt J, et al. (2012) SNP Discovery Using Next Generation Transcriptomic Sequencing in Atlantic Herring (*Clupea harengus*) *PLoS ONE* **7**: e42089.

Hoarau G, Boon E, Jongma DN, Ferber S, Palsson J et al. (2005). Low effective population size and evidence for inbreeding in an overexploited flatfish, plaice (*Pleuronectes platessa* L.). *P Roy Soc Lond B Bio* **272**: 497-503.

Hudson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies *Mol Biol Evol* **23**: 254-267.

Hue SB (1979) Electrophoretic studies on the heterogeneity of the albacore stock (*Thunnus alalunga*) in the northeast Atlantic. *ICCAT Col Vol Sci Pap* **8**: 265-271

Hutchinson WF, Van Oosterhout C, Rogers SI, Carvalho GR (2003) Temporal analysis of archived samples indicates marked genetic changes in declining North Sea cod (*Gadus morhua*). *Proc R Soc B* **270**: 2125-2132.

ICCAT [14-04] recommendations for eastern (14-04) BFT.

ICCAT 2014. Report of the 2013 ICCAT North and South Atlantic albacore stock assessment meeting (Sukarrieta, Spain - June 17 to 24, 2013). *ICCAT Col Vol Sci Pap*. In press.

Ihssen PE, Booke HE, Casselman JM, McGlade JM, Payne NR et al. (1981) Stock identification: Materials and methods. *Can J Fish Aquat Sci* **38**: 1838-1855

Ishii T (1965) Morphometric analysis of the Atlantic albacore populations mainly her eastern areas. *B Jpn Soc Sci Fish* **31**: 333-339

Iwai T, Nakamura I (1964) Branchial skeleton of the bluefin tuna, with special reference to the gill rays. *Bull Misaki Mar Biol Inst Kyoto Univ* **6**:21-25

Iwai T, Nakamura I, Matsubara K (1965) Taxonomic study of the tunas. *Misaki Mar Biol Inst Kyoto Univ* **2**:1-51

Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* **23**: 1801-1806.

Jamieson A (1973) Genetic "tags" for marine fish stocks. *Sea Fish Res* Elek Science, London: 91-99

Jansen T, Gislason H (2013) Population Structure of Atlantic Mackerel (*Scomber scombrus*) *PLoS ONE* **8**: e64744.

Jensen JL, Bohonak AJ, Kelley ST (2005) Isolation by distance, web service. *BMC Genet* **6**: 13.

Jones S, Silas EG (1960). Indian tunas-A preliminary review, with a key for their identification. *Indian J Fish* **7**: 369-393.

Jorde PE, Ryman N (1995) Temporal allele frequency change and estimation of effective size in populations with overlapping generations. *Genetics* **139**: 1077-1090.

Jorde PE, Ryman N (2007) Unbiased estimator for genetic drift and effective population size. *Genetics* **177**: 927-935.

Jørgensen HBH, Hansen MM, Bekkevold D, Ruzzante DE, Loeschcke V (2005) Marine landscapes and population genetic structure of herring (*Clupea harengus* L.) in the Baltic Sea. *Mol Ecol* **14**: 3219-3234.

Juan-Jordá MJ, Mosqueira I, Cooper AB, Freire J, Dulvy NK (2011) Global population trajectories of tunas and their relatives. *Proc Natl Acad Sci USA* **108**: 20650-10655.

Juan-Jordá MJ, Mosqueira I, Freire J, Dulvy NK (2013). Life in 3-D: life history strategies in tunas, mackerels and bonitos. *Rev Fish Biol Fisher* **23**: 135-155.

Kaminski S, Olech W, Olenski K, Nowak Z, Rusc A (2012) Single nucleotide polymorphisms between two lines of European bison (*Bison bonasus*) detected by the use of Illumina Bovine 50 K BeadChip. *Conserv Genet Resour* **4**: 311-314.

Kettle AJ, Morales-Muniz A, Rosello-Izquierdo E, Heinrich D, Vøllestad LA (2011) Refugia of marine fish in the northeast Atlantic during the last glacial maximum: concordant assessment from archaeozoology and palaeotemperature reconstructions. *Clim Past* **7**: 181-201.

Knutsen H, Jorde PE, André C, Stenseth NC (2003) Fine-scaled geographical population structuring in a highly mobile marine species: the Atlantic cod. *Mol Ecol* **12**: 385-394.

Knutsen H, Jorde PE, Albert OT, Hoelzel AR, Stenseth NC (2007) Population genetic structure in the North Atlantic Greenland halibut (*Reinhardtius hippoglossoides*): influenced by oceanic current systems? *Can J Fish Aquat Sci* **64**: 857-866.

Knutsen H, Olsen EM, Jorde PE, Espeland SH, André C et al. (2011) Are low but statistically significant levels of genetic differentiation in marine fishes 'biologically meaningful'? A case study of coastal Atlantic cod. *Mol Ecol* **20**: 768-783.

Koto T (1969) Studies on the albacore-XIV Distribution and movement of the albacore in the Indian and the Atlantic Oceans based on the catch statistics of Japanese tuna long-line fishery *Bull Far Seas Fish Res Lab*, Japan **1**:115-129.

Kunal SP, Kumar G, Menezes MR, Meena RM (2014). Genetic homogeneity in longtail tuna *Thunnus tonggol* (Bleeker, 1851) from the northwest coast of India inferred from direct sequencing analysis of the mitochondrial DNA D-loop region. *Mar Biol Res* **10**: 738-743.

Kurogane K, Hiyama Y (1958) Morphometric comparison of the albacore from the northwest, the equatorial and the southwest Pacific. *Rec Oceanogr Wks* **4**: 67-75

Kurogane K, Hiyama Y (1959) Morphometric comparison of the albacore from the Indian and the Pacific Ocean. *Rec Oceanogr Wks* **5**: 68-84

Laconcha U, Iriondo M, Arrizabalaga H, Manzano C, Markaide P et al. (2015a) New nuclear SNP markers unravel the genetic structure and effective population size of albacore tuna (*Thunnus alalunga*). *PLoS ONE* **10**: e0128247.

Laconcha U Iriondo M, Arrizabalaga H, Manzano C, Markaide P (2015b) Phylogeographic pattern in albacore and phylogenetic relationship between *Thunnus* species inferred from direct mtDNA analysis. In prep.

Lalami Y, Tellai S, Barrois JM, Piccinetti C, Piccinetti Manfrin G (1973) Observations sur les oeufs et larves des thonides des cotes algeriennes Pelagos IV (2): 54-65.

Langella O (2002) Populations 1.2.32. CNRS UPR 9034.

Laval G, Patin E, Barreiro LB, Quintana-Murci L (2010) Formulating a historical and demographic model of recent human evolution based on resequencing data from noncoding regions. *PloS ONE* **5**: e10284.

Le Gall JY (1974) Expose synoptique des données biologiques sur le germon *Thunnus alalunga* (Bonaterre, 1788) de l'Ocean Atlantique. *Synopsis FAO sur les péches* **109**: 70

Lenfant P, Planes S (2002). Temporal genetic changes between cohorts in a natural population of a marine fish, *Diplodus sargus*. *Biol J Linn Soc* **76**: 9-20.

Lewis AD (1990) South Pacific albacore stock structure: a review of available information South Pacific Commission, Noumea, New Caledonia, p 13

Lourie SA, Vincent AC (2004). Using biogeography to help set priorities in marine conservation. *Conserv Biol* **18**: 1004-1020.

Lowenstein, JH, Amato G, Kolokotronis SO (2009). The real maccoyii: identifying tuna sushi with DNA barcodes–contrasting characteristic attributes and genetic distances. *PLoS ONE* **4**: e7866.

Lutcavage ME, Brill RW, Skomal GB, Chase BC, Howey PW (1999) Results of pop-up satellite tagging on spawning size class fish in the Gulf of Maine: Do North Atlantic bluefin tuna spawn in the mid-Atlantic? *Can J Fish Aquat Sci* **56**: 173-177.

Lyons LA, Laughlin TF, Copeland NG, Jenkins NA, Womack J E et al. (1997) Comparative anchor tagged sequences (CATS) for integrative mapping of mammalian genomes. *Nat Genet* **15**: 47-56.

Macdonald JI, Farley JH, Clear NP, Williams AJ, Carter TI et al (2013) Insights into mixing and movement of South Pacific albacore *Thunnus alalunga* derived from trace elements in otoliths. *Fish Res* **148**: 56-63.

Magnuson JJ, Block BA, Deriso RB, Gold JR, Grant WS et al. (1994) An assessment of Atlantic bluefin tuna. *Natl Acad Press*, Washington, DC.

Malhi RS, Trask JS, Shattuck M, Johnson J, Chakraborty D et al. (2011) Genotyping single nucleotide polymorphisms (SNPs) across species in Old World monkeys. *Am J Primatol* **73**: 1031-1040.

Marano G, De Zio V, Pastorelli AM, Rositani L, Ungaro N et al. (1999) Studio sinottico sulla biologia e pesca di *Thunnus alalunga* (Bonnaterre, 1788) *Biol Mar Medit* **6**: 192-214.

Mather FJ, Mason JM. Jones AC (1995) Historical document: life history and fisheries of Atlantic bluefin tuna. *NOAA Technical Memorandum* **370**, NOAA, Miami.

McCluskey SM, Lewison RL (2008) Quantifying fishing effort: a synthesis of current methods and their applications. *Fish Fish* **9**: 188-200.

McLenachan PA, Stöckler K, Winkworth RC, McBreen K, Zauner S, Lockhart PJ (2000). Markers derived from amplified fragment length polymorphism gels for plant ecology and evolution studies. *Mol Ecol* **9**: 1899-1903.

Megalofonou P (2000) Age and growth of Mediterranean albacore. *J Fish Biol* **57**: 700-715.

Mesnick SL, Taylor BL, Archer FI, Martien KK, Treviño SE et al. (2011) Sperm whale population structure in the eastern and central North Pacific inferred by the use of single-nucleotide polymorphisms, microsatellites and mitochondrial DNA. *Mol Ecol Resour* **11**: 278-298.

Metzker ML (2010). Sequencing technologies—the next generation. *Nat Rev Genet* **11**: 31-46.

Michelini E, Cevenini L, Mezzanotte L, Simoni P, Baraldini M et al. (2007) One-step triplex-polymerase chain reaction assay for the authentication of yellowfin (*Thunnus albacares*), bigeye (*Thunnus obesus*), and skipjack (*Katsuwonus pelamis*) tuna DNA from fresh, frozen, and canned tuna samples. *J Agr Food Chem* **55**: 7638-7647.

Miller JM, Poissant J, Kijas JW, Coltman DW (2011) A genome-wide set of SNPs detects population substructure and long range linkage disequilibrium in wild sheep. *Mol Ecol Resour* **11**: 314-322.

Miller JA, Gray A, Merz J (2010) Quantifying the contribution of juvenile migratory phenotypes in a population of Chinook salmon *Oncorhynchus tshawytscha. Mar Ecol-Prog Ser* **408**: 227-240.

Moen T, Hayes B, Nilsen F, Delghandi M, Fjalestad KT et al. (2008) Identification and characterisation of novel SNP markers in Atlantic cod: evidence for directional selection. *BMC Genet* **9**: 18.

Molecular Ecology Resources Primer Development Consortium, Abreu AG, Albaina A, Alpermann TJ, Apkenas VE et al. (2012) Permanent genetic resources added to Molecular Ecology Resources Database 1 October 2011-30 November 2011. *Mol Ecol Resour* **12**: 374-376.

Montes I, Iriondo M, Manzano C, Arrizabalaga H, Jiménez E, et al.(2012) Worldwide genetic structure of albacore (*Thunnus alalunga*) revealed by microsatellite DNA markers. *Mar Ecol-Prog Ser* **471**: 183-191.

Morin PA, Luikart G, Wayne RK (2004) SNPs in ecology, evolution and conservation. *Trends Ecol Evol* **19**: 208–216.

Munday CJ, Whitehead GM, Todd NJ, Campbell M, Hawkey PM (2004). Predominance and genetic diversity of community-and hospital-acquired CTX-M extended-spectrum β-lactamases in York, UK. *J Antimicrob Chemoth* **54**: 628-633.

Nakadate M, Viñas J, Corriero A, Clarke S, Suzuki N et al. (2005) Genetic isolation between Atlantic and Mediterranean albacore populations inferred from mitochondrial and nuclear DNA markers. *J Fish Biol* **66**: 1545-1557.

Nakamura I (1965) Relationships of fishes referable to the subfamily Thunninae on the basis of the axial skeleton. *Bull Misaki Mar Biol Inst Kyoto Univ* **8**:7-38

Nakamura H (1969) Tuna distribution and migration *Fishing News* Ltd, London: 76

Nielsen EE, Hemmer-Hansen J, Larsen PF, Bekkevold D (2009) Population genomics of marine fishes: identifying adaptative variation in space and time. *Mol Ecol* **18**: 3128-3150.

Nielsen EE, Cariani A, Mac Aoidh E, Maes GE, Milano I et al.(2012) Gene-associated markers provide tools for tackling illegal fishing and false ecocertification. *Nat Commun* **3**: 851. doi:10,1038/ncomms1845

Nishikawa Y, Honma M, Ueyanagi S, Kikawa S. (1985) Average distribution of larvae of oceanic species of scombrid fishes 1956–1981. *Nat Res I Far Seas Fish*, Shimizu: 99

Nomura S, Kobayashi T, Agawa Y, Margulies D, Scholey V et al. (2014) Genetic population structure of the Pacific bluefin tuna *Thunnus orientalis* and the yellowfin tuna *Thunnus albacares* in the North Pacific Ocean. *Fisheries Sci* **80**: 1193-1204.

Nunney L (1999) The effective size of a hierarchically structured population. *Evolution* **53**: 1-10.

Nunney L, Campbell KA (1993) Assessing minimum viable population size: demography meets population genetics. *Trends Ecol Evol* **8**: 234-239.

Ogden R (2011) Unlocking the potential of genomic technologies for wildlife forensics. *Mol Ecol Resour* **11**: 109-116.

Ogden R, Baird J, Senn H, McEwing R (2012) The use of cross-species genome-wide arrays to discover SNP markers for conservation genetics: a case study from Arabian and scimitar-horned oryx. *Conserv Genet Resour* **4**: 471-473.

O'Leary SJ, Hice LA, Feldheim KA, Frisk MG, McElroy AE et al. (2013) Severe Inbreeding and Small Effective Number of Breeders in a Formerly Abundant Marine Fish. *PLoS ONE* **8**: e66126.

Ortiz de Zárate V, Cort JL (1998) Albacore (*Thunnus alalunga*, Bonnaterre) stock structure in the Atlantic Ocean, as inferred from distribution and migration patterns. Proceedings of the ICCAT Tuna Symposium. *ICCAT Tuna Symposium* **1**: 251-260.

Otsu T. y Uchida R.N. (1959) Sexual maturity of albacore in the Pacific ocean. *Fish Bull* **59**: 787-805

Ovenden JR (1990) Mitochondrial DNA and marine stock assessment: a review. *Aust J Mar Freshw Res* **41**: 835-853

Palstra FP, Ruzzante DE (2008) Genetic estimates of contemporary effective population size: what can they tell us about the importance of genetic stochasticity for population persistence in the wild? *Mol Ecol* **17**: 3428-3447

Palumbi SR, Baker CS (1994) Contrasting population structure from nuclear intron sequences and mtDNA of humpback whales. *Mol Biol Evol* **11**: 426-435.

Pardo MA, Pérez-Villareal B (2004). Identification of commercial canned tuna species by restriction site analysis of mitochondrial DNA products obtained by nested primer PCR. *Food Chem* **86**: 143-150.

Pawson MG, Jennings S (1996) A critique of methods for stock identification in marine capture fisheries. *Fish Res* **25**: 203-217.

Petit E, Balloux F, Goudet J (2001) Sex-biased dispersal in a migratory bat: a characterization using sex-specific demographic parameters. *Evolution* **55**: 635-640.

Piccinetti C, Manfrin GP (1993) Distribution of Thunnidae larvae in the Mediterranean. *ICCAT Col Vol Sci Pap* **40**: 164-172

Piry S, Alapetite A, Cornuet JM, Baudouin L, Estoup A (2004) GENECLASS2: A software for genetic assignment and first-generation migrant detection. *J Hered* **95**: 536-539. doi:10,1093/jhered/esh074.

Posada D, Crandall KA (1998). Modeltest: testing the model of DNA substitution. *Bioinformatics* **14**: 817-818.

Poulsen NA, Nielsen EE, Schierup MH, Loeschcke V, Grønkjær P (2006) Long-term stability and effective population size in North Sea and Baltic Sea cod (*Gadus morhua*). *Mol Ecol* **15**: 321-331.

Poulsen NA, Hemmer-Hansen J, Loeschcke V, Carvalho GR, Nielsen EE (2011) Microgeographical population structure and adaptation in Atlantic cod *Gadus morhua*: spatio-temporal insights from gene-associated DNA markers. *Mar Ecol-Prog Ser* **436**: 231-243.

Primmer CR, Borge T, Lindell J, Saetre GP (2002) Single-nucleotide polymorphism characterization in species with limited available sequence information: high nucleotide diversity revealed in the avian genome. *Mol Ecol* **11**: 603-612.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* **155**: 945-959.

Pujolar JM, Roldan MI, Pla C (2003) Genetic analysis of tuna populations, *Thunnus thynnus* and *Thunnus alalunga*. *Mar Biol* **143**: 613-621.

Qiu F, Miyamoto MM (2011) Use of nuclear DNA data to estimate genetic diversity and population size in Pacific Bluefin and Yellowfin Tuna (*Thunnus orientalis* and *Thunnus albacares*). *Copeia* **2011**: 264-269.

Qiu F, Kitchen A, Beerli P, Miyamoto MM (2013) A possible explanation for the population size discrepancy in tuna (genus *Thunnus*) estimated from mitochondrial DNA and microsatellite data. *Mol Phylogenet Evol* **66**: 463-468.

Rannala B, Mountain JL (1997) Detecting immigration by using multilocus genotypes. *Proc Natl Acad Sci USA* **94**: 9197-9221. doi:10,1073/pnas.94.17.9197.

Reynolds J, Weir BS, Cockerham CC (1983) Estimation of the coancestry coefficient: basis for a short-term genetic distance. *Genetics* **105**: 767-779.

Riccioni G, Landi M, Ferrara G, Milano I, Cariani A et al. (2010) Spatio-temporal population structuring and genetic diversity retention in depleted Atlantic bluefin tuna of the Mediterranean Sea. *Proc Natl Acad Sci USA* **107**: 2102-2107.

Richards WJ (1984) Kinds and abundances of fish larvae in the Caribbean Sea and adjacent areas. *US Dep Commer NOAA Tech Rep* **776:** 54

Richter-Boix A, Quintela M, Segelbacher G, Laurila A (2011) Genetic analysis of differentiation among breeding ponds reveals a candidate gene for local adaptation in *Rana arvalis*. *Mol Ecol* **20**: 1582-1600.

Ricker WE (1981) Changes in the average size and average age of Pacific salmon. *Can J Fish Aquat Sci* **38**: 1636-1656.

Roest CH, Weissenbach J (2005). Fish genomics and biology. *Genome Res* **12**: 1675-1682.

Rooker JR, Secor DH, De Metrio G, Rodríguez-Marín E (2006) Evaluation of population structure and mixing rates of Atlantic bluefin tuna from chemical signatures in otoliths. The International Commission for the Conservation of Atlantic Tuna. *Col Vol Sci Pap* **59**: 813-818.

Rooker JR, Alvarado Bremer JR, Block BA, Dewar H, de Metrio G et al. (2007) Life history and stock structure of Atlantic Bluefin Tuna (*Thunnus thynnus*). Reviews in *Fisheries Sci* **15**: 265-310.

Rooker JR, Secor DH, De Metrio G, Schloesser R, Block BA, Neilson JD (2008) Natal homing and connectivity in Atlantic bluefin tuna populations. *Science* **322**: 742-744.

Rosenberg NA, Li LM, Ward R, Pritchard JK (2003) Informativeness of genetic markers for inference of ancestry. *Am J Hum Genet* **73**: 1402-1422.

Rosenberg NA (2004) Distruct: a program for the graphical display of population structure *Mol Ecol Notes* **4**: 137-138.

Rousset F (1996) Equilibrium values of measures of population subdivision for stepwise mutation processes. *Genetics* **142**: 1357-1362.

Rousset F (2008) genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Mol Ecol Resour* **8**: 103-106. doi:10,1111/j.1471-8286.2007.01931.x.

Rozas J, Sánchez-DelBarrio JC, Messeguer X, Rozas R (2003). DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19**: 2496-2497.

Rozen S, Skaletsky HJ (2000) Primer3 on the WWW for general users and for biologist programmers. *Method Mol Biol* **132**: 365-386.

Ruzzante DE, Mariani S, Bekkevold D, André C, Mosegaard H, et al. (2006) Biocomplexity in a highly migratory pelagic marine fish, Atlantic herring. *P Roy Soc B-Biol Sci* **273**: 1459-1464.

Ruzzante DE (1997) A comparison of several measures of genetic distance and population structure with microsatellite data: bias and sampling variance. *Can J Fish Aquat Sci* **55**: 1–14.

Ryman N, Palm S, André C, Carvalho GR, Dahlgren TG et al. (2006) Power for detecting genetic divergence: differences between statistical methods and marker loci. *Mol Ecol* **15**: 2031-2045.

Ryynänen HJ, Primmer CR (2006) Single nucleotide polymorphism (SNP) discovery in duplicated genomes: intron-primed exon-crossing (IPEC) as a strategy for avoiding amplification of duplicated loci in Atlantic salmon (*Salmo salar*) and other salmonid fishes. *BMC Genomics* **7**: 192.

Santaclara FJ, Velasco A, Pérez-Martín RI, Quinteiro J, Rey-Méndez M et al.(2014) Development of a multiplex PCR-ELISA method for the genetic authentication of *Thunnus* species and *Katsuwonus pelamis* in food products. *Food Chem* **180**: 9-16.

Santiago J (1993) A new length-weight relationship for the North Atlantic albacore. *ICCAT Col Vol Sci Pap* **40**: 316-319.

Santiago J, Arrizabalaga H (2005) An integrated growth study for North Atlantic albacore (*Thunnus alalunga* Bonn 1788) *ICES J Mar Sci*: Journal du Conseil **62**: 740-749.

Schwartz MK, Tallmon DA, Luikart G (1998). Review of DNA-based census and effective population size estimators. *Anim Conserv* **1**: 293-299.

Seeb JE, Carvalho G, Hauser L, Naish K, Roberts S et al. (2011) Single-nucleotide polymorphism (SNP) discovery and applications of SNP genotyping in nonmodel organisms. *Mol Ecol Resour* **11**: 1-8.

Serene PH (1969) Esterase of the north east Atlantic albacore stock. *ICES Special meeting on the biochemical and serological identification of fish stocks* **33**: 1-6

Sharp GD, Pirages S (1978) The distribution of red and white swimming muscles, their biochemistry, and the biochemical phylogeny of selected scombrid fishes. In: Sharp GD, Dizon AE (eds) The physiological ecology of tunas. *Academic Press*, New York: 41-78

Shiohama T (1985) Overall fishing intensity and length composition of albacore caught by long line fishery in the Indian Ocean, 1952-1982. IPTP TWS/85/22 91-109.

Shubitowski DM, Venta PJ, Douglass CL, Zhou RX, Ewart SL (2001). Polymorphism identification within 50 equine gene-specific sequence tagged sites. *Anim Genet* **32**: 78-88.

Slate J, Gratten J, Beraldi D, Stapley J, Hale M et al.(2009) Gene mapping in the wild with SNPs: guidelines and future directions. *Genetica* **136**: 97-107.

Slatkin M, Hudson RR (1991). Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics* **129**: 555-562.

Smith CT, Elfstrom CM, Seeb L, Seeb JE (2005) Use of sequence data from rainbow trout and Atlantic salmon for SNP detection on Pacific salmon. *Mol Ecol* **14**: 4193-4203.

Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* **68**: 978-989.

Stepien CA, Faber JE. (1998). Population genetic structure, phylogeography and spawning philopatry in walleye (*Stizostedion vitreum*) from mitochondrial DNA control region sequences. Mol Ecol **7**: 1757-1769.

Sund PN, Blackburn M, Williams F (1981) Tunas and their environment in the Pacific Ocean: A review. *Oceanogr Mar Biol* **19**: 443-512.

Suzuki A (1962) Serologocal studies of the races of tuna VI Bigeye-3 antigen occurred in the albacore. *Rep Nankai Reg Fish Res Lab* **16**: 67-70.

Tajima F (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**: 585-595.

Takagi M, Okamura T, Chow S, Taniguchi N (1999) PCR primers for microsatellite loci in tuna species of the genus *Thunnus* and its application for population genetic study. *Fish Sci* **65**: 571-576.

Takagi M, Okamura T, Chow S, Taniguchi N (2001) Preliminary study of albacore (*Thunnus alalunga*) stock differentiation inferred from microsatellite DNA analysis. *Fish Bull* **99**: 697-701.

Takeyama H, Chow S, Tsuzuki H, Matsunaga T (2001) Mitochondrial DNA sequence variation within and between tuna *Thunnus* species and its application to species identification. *J Fish Biol* **58**: 1646-1657.

Tamura K, Nei M, Kumar S. 2004. Prospects for inferring very large phylogenies by using the neighbor-joining method.*Proc NatlAcad Sci USA* **101**:11030–11035

Teo SLH, Boustany A, Dewar H, Stokesbury MJW, Weng KC et al. (2007) Annual migrations, diving behavior, and thermal biology of Atlantic bluefin tuna, *Thunnus thynnus*, on their Gulf of Mexico breeding grounds. *Mar Biol* **151**: 1-18.

Therkildsen NO, Nielsen EE, Swain DP, Pedersen JS (2010) Large effective population size and temporal genetic stability in Atlantic cod (*Gadus morhua*) in the southern Gulf of St Lawrence. *Can J Fish Aquat Sci* **67**: 1585-1595

Tseng MC, Jean CT, Smith PJ, Hung YH (2012). Interspecific and intraspecific genetic diversity of *Thunnus* species. INTECH Open Access Publisher.

Turner S (1999) Review of the bluefin tuna release and recapture information from the ICCAT tagging data base. *ICCAT Col Vol Sci Pap* **XLIX:** 360- 362.

Turner TF, Wares JP, Gold JR (2002) Genetic effective size is three orders of magnitude smaller than adult census size in an abundant, estuarine-dependent marine fish (*Sciaenops ocellatus*). *Genetics* **162**: 1329-1339.

Ueyanagi S (1969) Observations on the distribution of tuna larvae in the indo-pacific Ocean with emphasis on the delineation of the spawning areas of albacore, *Thunnus alalunga*. *Bull Far Seas Fish Res Lab* **2**: 177-256.

Ueyanagi S (1971) Larval distribution of tunas and billfishes in the Atlantic Ocean. *FAO Fisheries report* **71**: 297-305

Viñas J, Santiago J, Pla C (1999) Genetic characterization and Atlantic Mediterranean stock structure of Albacore, *Thunnus alalunga*. *ICCAT Col Vol Sci Pap* **49**: 188-191.

Viñas J & Tudela S (2009) A Validated Methodology for Genetic Identification of Tuna Species (Genus *Thunnus*). *PLoS ONE*, **4**: e7606.

Viñas J, Alvarado Bremer JR, Pla C (2004) Inter-oceanic genetic differentiation among albacore (*Thunnus alalunga*) populations. *Mar Biol* **145**: 225-232.

Viñas J, Gordoa A, Fernández-Cebrián R, Pla C, Vahdet U et al. (2011) Facts and uncertainties about the genetic population structure of Atlantic bluefin tuna (*Thunnus thynnus*) in the Mediterranean. Implications for fishery management. Reviews in *Fish Biol Fish* **21**: 527-541.

Vucetich JA, Waite TA, Nunney L (1997) Fluctuating population size and the ratio of effective to census population size (Ne/N). *Evolution* **51**: 2017-2021.

Waldman JR, Richards RA, Schill WB, Wirgin I y Fabrizio MC (1997) An empirical comparison of stock identification techniques applied to striped bass. *Trans Am Fish Soc* **126**: 369-385

Waldman JR (1999) The importance of comparative studies in stock analysis. *Fish Res* **43**: 237-246.

Walli A, Teo SLH, Boustany A, Farwell CJ, Williams T et al. (2009) Seasonal Movements, Aggregations and Diving Behavior of Atlantic Bluefin Tuna (*Thunnus thynnus*) Revealed with Archival Tags. *PLoS ONE* **4**: e6151.

Wang J (2005) Estimation of effective population sizes from data on genetic markers. *Philos T Roy Soc B* **360**: 1395-1409.

Wang IJ (2011) Choosing appropriate genetic markers and analytical methods for testing landscape genetic. *Mol Ecol* **20**: 2480–2482.

Waples RS (1989) A generalized approach for estimating effective population size from temporal changes in allele frequency. *Genetics* **121**: 379-391.

Waples, RS (1998). Separating the wheat from the chaff: patterns of genetic differentiation in high gene flow species. *J Hered* **98**: 438-450.

Waples RS, Gaggiotti O (2006) Invited review: What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Mol Ecol* **15**: 1419-1439.

Waples RS, Naish KA (2009) Genetic and evolutionary considerations in fishery management: Research needs for the future. In: The Future of Fisheries Science in North America. *Fish Fish Ser* **31** (ed. by R. Beamish & B. Rothschild): 427-451. Springer, Netherlands.

Waples RS, Yokota M (2007) Temporal estimates of effective population size in species with overlapping generations. *Genetics* **175**: 219-233.

Ward RD (2000) Genetics in fisheries management. *Hydrobiologia* **420**: 191-201.

Weir BS, Cockerham CC (1984) Estimating F statistics for the analysis of population structure. *Evolution* **38**: 1358-1370.

Whiteley AR, Hastings K, Wenburg JK, Frissell CA, Martin JC, Allendorf FW (2010) Genetic variation and effective population size in isolated populations of coastal cutthroat trout. *Conserv Genet* **11**: 1929-1943.

Whitlock MC, Barton NH (1997) The effective size of a subdivided population. *Genetics* **146**: 427-441.

Yeh SY, Hui CF, Treng TD, Kuo CL (1996) Indian Ocean albacore stock structure studies by morphometric and DNA sequence methods. In: Anganuzzi A.A., Stobberup K.A., Webb N.J.e. (eds) *Proceedings of the sixth expert consultation on Indian ocean tunas.* IPTP, Colombo, Sri Lanka, 25-29 September 1995: 258-263

Yeh SY, Treng TD, Hui CF, Penney AJ (2007) Mitochondrial DNA sequence analysis on albacore, *Thunnus alalunga*, meat samples collected from the waters off western South Africa and the eastern Indian Ocean. *Col Vol Sci Pap*. International Commission for the Conservation of Atlantic Tunas **46**: 152-159.

Zarraonaindia I (2011) Estudio genético de la estructura poblacional de la anchoa europea (*Engraulis encrasicolus*). PhD thesis. University of the Basque Country (UPV/EHU), Genetics, Physical Anthropology & Animal Physiology Department.

ANIMAL GENETICS Immunogenetics, Molecular Genetics and Functional Genomics

doi: 10.1111/age.12051

Single nucleotide polymorphism discovery in albacore and Atlantic bluefin tuna provides insights into worldwide population structure

A. Albaina*, M. Iriondo*, I. Velado*, U. Laconcha*,¶, I. Zarraonaindia*, H. Arrizabalaga¶, M. A. Pardo†, M. Lutcavage‡, W. S. Grant§ and A. Estonba*

*Genetika, Antropologia Fisikoa eta Animalien Fisiologia Saila, Zientzia eta Teknologia Fakultatea, Euskal Herriko Unibertsitatea (UPV/EHU), P.O. Box 48940, Leioa, Spain. †AZTI Tecnalia, Food Research Unit, Parque Tecnológico de Bizkaia, Astondo Bidea, Edif. 609, Derio (Bizkaia), Spain. ‡Environmental Conservation Department, Marine Fisheries Institute, University of Massachusetts Amherst, P.O. Box 3188, Gloucester, MA, 01931, USA. §Commercial Fisheries Division, Alaska Department of Fish and Game, Anchorage, AK, 99518, USA. ¶AZTI Tecnalia, Herrera Kaia Portualdea z/g, P.O. Box 20110, Pasaia, Gipuzkoa, Spain.

Summary

The optimal management of the commercially important, but mostly over-exploited, pelagic tunas, albacore (*Thunnus alalunga* Bonn., 1788) and Atlantic bluefin tuna (BFT; *Thunnus thynnus* L., 1758), requires a better understanding of population structure than has been provided by previous molecular methods. Despite numerous studies of both species, their population structures remain controversial. This study reports the development of single nucleotide polymorphisms (SNPs) in albacore and BFT and the application of these SNPs to survey genetic variability across the geographic ranges of these tunas. A total of 616 SNPs were discovered in 35 albacore tuna by comparing sequences of 54 nuclear DNA fragments. A panel of 53 SNPs yielded F_{ST} values ranging from 0.0 to 0.050 between samples after genotyping 460 albacore collected throughout the distribution of this species. No significant heterogeneity was detected within oceans, but between-ocean comparisons (Atlantic, Pacific and Indian oceans along with Mediterranean Sea) were significant. Additionally, a 17-SNP panel was developed in Atlantic BFT by cross-species amplification in 107 fish. This limited number of SNPs discriminated between samples from the two major spawning areas of Atlantic BFT (F_{ST} = 0.116). The SNP markers developed in this study can be used to genotype large numbers of fish without the need for standardizing alleles among laboratories.

Keywords fisheries management, population genetics, single nucleotide polymorphism (SNP), single nucleotide polymorphism discovery, *Thunnus alalunga*, *Thunnus thynnus*

Introduction

Molecular genetics has led to considerable progress in understanding the ecologies of marine species by providing new insights into the demographic and evolutionary dynamics of wild populations (Hauser & Carvalho 2008). Genetic markers are widely used to identify stocks, to estimate mixed stocks in a fishery, to monitor genetic diversity within populations and to measure connectivity between populations, among many other applications (e.g., Nielsen *et al.* 2009; Waples & Naish 2009). These studies have overturned the classic notion that large marine populations are genetically homogeneous with limited local adaptation by showing extensive genetic population structure in many marine species (reviewed in Hauser & Carvalho 2008). Moreover, genetic structuring has been reported even across small spatial scales (e.g., Knutsen *et al.* 2003, 2007, 2011; Jørgensen *et al.* 2005). However, data are still rare for the vast majority of highly exploited species, even though a large number of studies on genetic population structure of marine fish have been published in the past decades. Genetic studies are needed to improve the management of species for which stock structure and migration patterns are still unclear. An improper management of these fishery resources can lead to the extirpation of small independent stocks or to the under-utilization of large stocks.

This study focuses on two widely distributed pelagic tunas: albacore (*Thunnus alalunga* Bonn., 1788) and Atlantic bluefin tuna (BFT; *Thunnus thynnus* L., 1758). Albacore is one of the smallest tunas and BFT one of the largest in the genus Thunnus. Although albacore is a widely distributed species, inhabiting both temperate and tropical pelagic waters of all oceans, the distribution of Atlantic BFT is

Address for correspondence

A. Estonba, Genetika, Antropologia Fisikoa eta Animalien Fisiologia Saila, Zientzia eta Teknologia Fakultatea, Euskal Herriko Unibertsitatea (UPV/EHU), P.O. Box 48940, Leioa, Spain.
E-mail: andone.estonba@ehu.es

Accepted for publication 18 March 2013

© 2013 The Authors, Animal Genetics © 2013 Stichting International Foundation for Animal Genetics

RESEARCH ARTICLE

New Nuclear SNP Markers Unravel the Genetic Structure and Effective Population Size of Albacore Tuna (*Thunnus alalunga*)

Urtzi Laconcha[1,2], Mikel Iriondo[2], Haritz Arrizabalaga[1]*, Carmen Manzano[2], Pablo Markaide[2], Iratxe Montes[2], Iratxe Zarraonaindia[2], Igor Velado[2], Eider Bilbao[3], Nicolas Goñi[1], Josu Santiago[4], Andrés Domingo[5], Saadet Karakulak[6], Işık Oray[6], Andone Estonba[2]*

1 AZTI, Marine Research Division, Herrera Kaia Portualdea z/g, Pasaia (Gipuzkoa), Spain, **2** Department of Genetics, Physical Anthropology and Animal Physiology, University of the Basque Country (UPV/EHU), Sarriena auzoa z/g, Leioa (Bizkaia), Spain, **3** Cell Biology & Environmental Toxicology Research Group, Research Centre for Experimental Marine Biology & Biotechnology (PIE) & Department of Zoology & Animal Cell Biology, University of the Basque Country (UPV/EHU), Sarriena auzoa z/g, Leioa (Bizkaia), Spain, **4** AZTI, Marine Research Division, Txatxarramendi Ugartea z/g, Sukarrieta (Bizkaia), Spain, **5** Laboratory of Pelagic Resources, Dirección Nacional de Recursos Acuáticos (DINARA), Constituyente 1497, Montevideo, Uruguay, **6** Fisheries Faculty of Istanbul University, Laleli, Istanbul, Turkey

* harri@azti.es (HA); andone.estonba@ehu.es (AE)

OPEN ACCESS

Citation: Laconcha U, Iriondo M, Arrizabalaga H, Manzano C, Markaide P, Montes I, et al. (2015) New Nuclear SNP Markers Unravel the Genetic Structure and Effective Population Size of Albacore Tuna (*Thunnus alalunga*). PLoS ONE 10(6): e0128247. doi:10.1371/journal.pone.0128247

Academic Editor: Roberta Cimmaruta, Tuscia University, ITALY

Received: December 4, 2014

Accepted: April 23, 2015

Published: June 19, 2015

Copyright: © 2015 Laconcha et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This research was supported by the projects ATM2010 Hegaluze (351BI20090047) and TUNASNIP (SPE10UN92) funded by the Basque Government. Urtzi Laconcha's work was supported by a PhD grant by the Fundación Centros Tecnológicos Iñaki Goenaga. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

In the present study we have investigated the population genetic structure of albacore (*Thunnus alalunga*, Bonnaterre 1788) and assessed the loss of genetic diversity, likely due to overfishing, of albacore population in the North Atlantic Ocean. For this purpose, 1,331 individuals from 26 worldwide locations were analyzed by genotyping 75 novel nuclear SNPs. Our results indicated the existence of four genetically homogeneous populations delimited within the Mediterranean Sea, the Atlantic Ocean, the Indian Ocean and the Pacific Ocean. Current definition of stocks allows the sustainable management of albacore since no stock includes more than one genetic entity. In addition, *short-* and *long-term* effective population sizes were estimated for the North Atlantic Ocean albacore population, and results showed no historical decline for this population. Therefore, the genetic diversity and, consequently, the adaptive potential of this population have not been significantly affected by overfishing.

Introduction

Albacore tuna (*Thunnus alalunga*, Bonnaterre 1788) is distributed in the Atlantic, Pacific and Indian Oceans and in the Mediterranean Sea, extending from 50–55°N to 40–45°S [1]. This species is the fourth most important one of the *Thunnus* genus with regard to captures [2]. This fact reflects the high commercial value of the albacore and its related products, which makes this species likely to be exploited beyond its maximum sustainable yield [1]. Migrations on this species has been studied for several decades through tag-recapture experiments showing low rate of albacore migration between hemispheres [3], and no transoceanic [3,4,5]