UNIVERSIDADE DO ALGARVE

FACULDADE DE CIÊNCIAS E TECNOLOGIA

Study of genetic gradients among populations of Atlantic anchovy (*Engraulis encrasicolus* L.) located along marine ecotones

Gonçalo Jorge Franco Silva

TESE PARA A OBTENÇÃO DO GRAU DE DOUTOR

NO RAMO DE CIÊNCIAS BIOLÓGICAS

ESPECIALIDADE DE BIOLOGIA EVOLUTIVA

ORIENTADOR|SUPERVISOR

Profª Doutora Rita Castilho

ANO

2014

UNIVERSIDADE DO ALGARVE

FACULDADE DE CIÊNCIAS E TECNOLOGIA

Study of genetic gradients among populations of Atlantic anchovy (*Engraulis encrasicolus* L.) located along marine ecotones

Gonçalo Jorge Franco Silva

TESE PARA A OBTENÇÃO DO GRAU DE DOUTOR

NO RAMO DE CIÊNCIAS BIOLÓGICAS

ESPECIALIDADE DE BIOLOGIA EVOLUTIVA

ORIENTADOR|SUPERVISOR

Profª Doutora Rita Castilho

ANO

2014

Aos meus pais

e aos meus avós

A realização desta tese foi possível graças ao financiamento da Fundação para a Ciência e Tecnologia (FCT), com co-financiamento do Fundo Social Europeu, através da bolsa de doutoramento SFRH/BD/36600/2007 e do financiamento estratégico concedido ao Centro de Ciências do Mar (CCMAR).

This thesis was funded by the Portuguese Science Foundation (FCT) though the grant SFRH/BD/36600/2007 and the Strategic funding from FCT to the Center for Marine Sciences (CCMAR).

Declaração de autoria de trabalho:

Declaro ser o autor deste trabalho, que é original e inédito. Autores e trabalhos consultados estão devidamente citados no texto e constam da listagem de referências incluída. Esta publicação não poderá ser reproduzida de forma parcial ou integral, nem poderá ser copiada de alguma maneira, incluindo electronicamente, mecanicamente, fotocopiada, gravada ou digitalizada, sem autorização escrita do autor.

Gonçalo Jorge Franco Silva

Author Edition

ISBN: 978-989-20-4468-2

AGRADECIMENTOS / ACKNOWLEDGMENTS

Este doutoramento só foi possível graças ao contributo e colaboração entre diversas instituições, investigadores, colaboradores ocasionais, família e amigos. Sem eles a conclusão desta tese não teria sido possível e por isso estou eternamente agradecido a todos eles.

Em primeiro lugar queria agradecer à Professora Rita Castilho que aceitou a orientação deste doutoramento e que me ensinou bastante ao longo destes anos. Para além da inquestionável contribuição científica na produção dos artigos e da tese em geral, a Professora Rita Castilho fez-me perceber a importância do rigor na escrita científica, da apresentação gráfica e do cuidado com os pormenores no sucesso dos manuscritos, "obrigando-me" a melhorar significativamente nestas áreas.

À Fundação para a Ciência e Tecnologia por ter financiado a minha bolsa de doutoramento (SFRH/BD/36600/2007) e ao projecto MarinERA-NET-project/189570 "Marine phylogeographic structuring during climate change: the signature of leading and rear edge of range shifting populations" da European Science Foundation.

To the members of the jury for reading my thesis and to participate in the defense.

I would like to highlight the importance of many colleagues in collecting and shipping samples, namely Else Torstensen (IMR, Norway), Miroslaw Wyszynski (SFI, Poland), Ciara O'Donnel (Ireland), Sandrine Vaz, Frank Coppin, Yves Verin, Didier Leroy and Bernard Liorzou (IFREMER, France), Iratxe Zarraonaindia (UPV/EHU, Spain), Radhouan Ben-Hamadou, Karim Erzini, André Dias, Anabela Brito, Pedro Veiga, Henrique Cabral, Ana Faria, Pedro Morais, Mafalda Rangel, Catarina Oliveira, Wilson Pinto, Ana Tereza Gonçalves, Miguel Ruano, João Afonso, Sónia Massa, Cristina Paulino and Aníbal Guerra (Portugal), Jorge Tornero, Joana Duarte and José Alarcón (IEO, Spain), Urbi (Spain), Maria Tereza Santamaria (Canary Islands), Alain Hervé (Senegal), Luís Fonseca (Guinea-Bissau), Hederick Dankwa (WRI, Ghana), Rob Leslie and Stewart Grant (South Africa), Stefania Marcato (Italy), Voula Karachle (Greece), Remzi Kavcıoğlu (Turkey), Michel Bariche (Libano), Danail Pavlov (Bulgary), Morad Khaled (Yemen), James Vasslides and Roland Hagan (RUMFS, USA) and Edward Wiley and Andrew Bentley (KNHM, USA), Stacy Rowe (WHOI, USA), Peter Chase, Kathy Hieb, Laura Carney (USA), Jorge Pablo Castelo (Brazil), Daniel Ruzzante (Canada), Kevin Rowling (Australia), Rob Tesker (New Zealand). TO ALL THEM, THANK YOU SO MUCH!

I would like to acknowledge to people who made relevant contributions to this work in particular to the co-authors of the scientific articles published or in the process of publication John Horne, Fernando Lima, Paulo Martel, Regina Cunha and Rita Castilho; to the anonymous referees, Martin Limborg and Christine Maggs who made constructive comments on drafts of the manuscripts; to those who contributed with fruitful discussions João Neiva, Joana Macedo, Radhouan Ben-Hamadou, Cláudia Patrão, Katty Nicastro and Gerardo Zardi; to people who contributed with technical help Nelson Coelho, Marta Valente, N. O. Therkildsen, Ana Ramos, Celine Madeira, Cláudia Patrão, Cátia Marques, Ana Balau, Carla Sousa-Santos, Aschwin Engelen, Luci Jesus and Sara Teixeira, and Alex Yu for sharing microsatellites; to Doctor Rafael Zardoya (MNCN, Madrid, Spain) and Doctor William Stewart Grant (Alaska Department of Fish and Game, USA) for their collaboration.

A todos os meus "lab mates" que passaram e que me acompanharam durante estes anos, um agradecimento mesmo mesmo especial porque sem vocês não sei o que teria sido de mim, que aturaram o meu mau feitio (principalmente de manhã antes do café), que estiveram sempre ao meu lado e que jamais esquecerei. À Xana Ramos (um agradecimento infinito para a sherifa, és demais), Cláudia Patrão (és uma mulher daquelas, uma grande referência para mim e ficarás para sempre no meu coração sua lesma gosmenta que me aturaste e que estiveste sempre ao meu lado), Nelso (o menino "má lindo" e o rei do 2.51, a tua paciência para comigo quase que é inesgotável), à Tânia Aires (que nos aturamos há mais de 15 anos, "amo você"), ao meu himáguo "Neves" por tudo, à Celine Madeira um agradecimento muito especial pela ajuda numa fase bastante crítica, à Licínia pelo teu sorriso, à Joana Costa (esse bicho do mato que eu gosto muito), à "Ritinha do papá", à "Quitina", à Carla Lourenço, à "Pati Tortilla", ao "Davidé", Mirjiam, Onno, Aschwin, Assunción, Regina, Buga, Catarina Mota, Pablo Manent (Ah si?), Filipe Alberto (o caçteiro), Tânia Pereira (a murcona mais pagodeira a sul), Sónia Massa, Maria João, Marta Valente e Ana Amaral (obrigado pelas tuas Anamarelices). Ao pessoal do lab 2.13 que se habituou às minhas "pedinchices".

A todos os meus amigos da adolescência. Aos meus amigos e colegas da Universidade (são tantos que não dá para individualizar). Aos professores que me marcaram no meu percurso académico, a profª da escola primária Maria Isabel, o profº de filosofia Alberto Antunes, a profª fisico-química Adelina Machado e ao meu orientador de tese de licenciatura e amigo Manuel Mota. Aos meus irmãos da luta, "La famiglia Bica/Meco" Dany, Eskilo, Ric, Tico, Tigas, Andreia, Lara, Vânia, já está! Até que enfim! Aos manos Tavares (família de sempre e para sempre). À minha família Farense, à Catarina, à Rute, à Boavida, ao Diogo Paulo, ao Jorge, à Li, ao Bruno, ao Veiga, ao Denis, ao Francisco Pires e ao Francisco Fernandes; à Gisela; ao Dinarte e à Macedo; aos meus "papás" Wilson e Catarina pelas vezes que me alimentaram e me deram de beber! Ao pessoal espalhado por esse mundo fora que estiveram sempre comigo nos bons e nos maus momentos. A todos os meus amigos que não referi ou que me esqueci (mil desculpas) e que passaram pela minha vida nestes últimos anos, o meu muito obrigado.

Por último não podia deixar de agradecer à minha família que é um enorme exemplo para mim, que esteve sempre comigo nos momentos mais difíceis, que apoiaram sempre as minhas decisões e a quem eu queria dedicar este importante capítulo da minha vida: aos meus avós Celeste, Júlia, António e Jorge, aos meus tios Artur e Filipe, ao meu "puto" Ricardo, ao Chuwinga o meu grande companheiro e aos meus pais que são os melhores do mundo! Obrigado, vocês são tudo para mim!

To all you I am indebt! Thank you so much! Obrigado do fundo do coração!

TABLE OF CONTENTS

STUDY OF THE GENETIC GRADIENTS AMONG POPULATIONS OF ATLANTIC ANCHOVY (*ENGRAULIS ENCRASICOLUS* L.) LOCATED ALONG MARINE ECOTONES

ABSTRACT

The Quaternary climate oscillations had a major role in shaping the genetic architecture of living species. In the marine realm, the apparent lack of physical barriers to dispersal allows organisms to track optimum physiological conditions by displacing their distribution ranges. The European anchovy *Engraulis encrasicolus* is a small pelagic fish that has a broad distribution range in the Atlantic Ocean. Despite the high ability for dispersal, this species exhibit an unusual population structure and two mitochondrial clades clinally distributed along the eastern Atlantic and the Mediterranean. In the present thesis, we investigated North Atlantic anchovy's response to climate cycles at the leading edges of the distribution range. These small pelagic fishes massively followed suitable thermal conditions cyclically over the Pleistocene and therefore were able to preserve high levels of genetic diversity. We further explored the variation of the mitochondrial clades of the European anchovy and found that the anti-tropically distributed clade is under positive selection, suggesting that temperature is shaping the contemporary distribution of mtDNA clade frequencies. The Old World Anchovies (OWA) complex, of which the European anchovy is part, has taxa distributed in the Pacific and Atlantic oceans. This complex originated at 3.16 Ma in the Indo-Pacific during the late Pliocene and split in two groups, one that remained in the Pacific Ocean and one that colonized the Atlantic Ocean during the Pleistocene (0.62 Ma). The genetic patterns among the OWA indicate no genetic diferentiation between putative species from the Atlantic Ocean, and low levels of ongoing geneflow between Atlantic and Pacific anchovies. Within the Pacific Ocean, two well supported mitochondrial clades reveal ancient trans-Equatorial migrations, while nuclear loci support contemporary admixture.

KEYWORDS: Engraulidae, *Engraulis encrasicolus*, Old World Anchovies, natural selection, speciation, phylogeography, biogeography, dispersal, migration, leading-edge colonization.

ESTUDO DOS GRADIENTES GENÉTICOS ENTRE POPULAÇÕES DE BIQUEIRÃO (*ENGRAULIS ENCRASICOLUS* L.) AO LONGO DE ECÓTONOS MARINHOS

RESUMO

As oscilações climáticas do Quaternário tiveram um papel fundamental na modelação da arquitectura genética das espécies vivas. No meio marinho a aparente ausência de barreiras físicas à dispersão permite que os organismos desloquem as suas áreas de distribuição de forma a acompanhar o seu óptimo ambiental. O biqueirão *Engraulis encrasicolus* é um pequeno peixe pelágico com ampla distribuição no Oceano Atlântico. Apesar de ter uma elevada capacidade de dispersão, esta espécie tem padrões pouco usuais para uma espécie com estas características apresentando um elevado grau de estrutura populacional e duas linhagens mitocondriais que se distribuem clinalmente ao longo do Atlântico Este e do Mediterrâneo. Na presente tese, investigámos a resposta aos ciclos climáticos dos biqueirões do Atlântico Norte, em particular das populações com distribuição geográfica marginal. Estes pequenos peixes pelágicos acompanharam ciclicamente as condições térmicas que lhe são mais favoráveis ao longo do Pleistoceno e, desta forma, preservaram os elevados níveis de diversidade genética. Explorámos também a variação dos grupos mitocondriais do biqueirão e descobrimos que o clade com distribuição anti-tropical está sob selecção positiva, sugerindo que a temperatura modela a distribuição contemporânea da frequência dos clades mitocondriais.

O complexo de espécies designado por "Old World Anchovies", do qual o biqueirão faz parte, é composto por seis taxa distribuídos nos oceanos Pacífico e Atlântico. Este complexo teve origem há 3.16 Ma no Indo-Pacífico no fim do Plioceno e dividiu-se em dois grupos, um que permaneceu no Pacífico e outro que colonizou o Oceano Atlântico durante o Pleistoceno (0.62 Ma). Os padrões genéticos das "old world anchovies", indicam que não há diferenciação genética entre as espécies putativas do Oceano Atlântico, mas que os níveis de migração contemporânea são baixos entre os biqueirões do Atlântico e do Pacífico. No Oceano Pacífico, os dois clades mitocondriais revelam migrações trans-Equatoriais anteriores ao último máximo glaciar, enquanto que os loci nucleares suportam migrações contemporâneas.

ESTUDO DOS GRADIENTES GENÉTICOS ENTRE POPULAÇÕES DE BIQUEIRÃO (*ENGRAULIS ENCRASICOLUS* L.) AO LONGO DE ECÓTONOS MARINHOS

RESUMO ALARGADO

As oscilações climáticas do Quaternário tiveram um papel fundamental na modelação da arquitectura genética das espécies vivas. Durante os períodos glaciares, os organismos de climas temperados contraíram a sua área de distribuição, permanecendo em refúgios onde encontraram as condições fisiológicas mínimas à sua sobrevivência, enquanto que expandiram a distribuição para latitudes mais elevadas durante períodos inter-glaciares. No meio marinho, a aparente ausência de barreiras físicas à dispersão permite que os organismos desloquem as suas áreas de distribuição de forma a acompanhar o seu óptimo ambiental, em contraste com o modelo de contracção-expansão dos organimos terrestres. As fluctuações de abundância e as variações na área de distribuição das espécies têm impacto na sua diversidade genética, ficando registado no seu ADN.

O biqueirão ou anchova Europeia *Engraulis encrasicolus* é um pequeno peixe pelágico com ampla distribuição no Oceano Atlântico, Mar Báltico, Mar Mediterrâneo e Mar Negro. Esta espécie caracteriza-se por ter um ciclo de vida curto, por ocupar um nível médio na cadeia trófica, por ser sensível às oscilações climáticas e por ter populações grandes. Apesar de ter uma elevada capacidade de dispersão esta espécie tem padrões pouco usuais para uma espécie com elevada mobilidade apresentando estrutura populacional e duas linhagens mitocondriais que se distribuem clinalmente ao longo do Atlântico Este e do Mar Mediterrâneo.

O objectivo desta tese foi investigar os padrões genéticos contemporâneos e históricos no biqueirão e identificar os processos que contribuem para a modelação da distribuição dos padrões genéticos. Para este fim utilizámos marcadores moleculares mitocondriais, o citocromo *b*, e marcadores nucleares hipervariáveis para analisar amostras colhidas ao longo da área de distribuição. Investigámos a resposta aos ciclos climáticos dos biqueirões do Atlântico Norte, em particular das populações com distribuição geográfica marginal, isto é, populações no extremo Norte da distribuição. Estes pequenos peixes pelágicos acompanharam ciclicamente ao longo do Pleistoceno as condições térmicas que lhe são mais favoráveis e, desta forma, preservaram elevados níveis de diversidade genética, que de outra forma teriam sido erodidos

principalmente por efeitos de deriva genética. As alterações climáticas promoveram deslocações latitudinais na distribuição do biqueirão e deram a oportunidade a que esta espécie se dispersasse também para o Atlântico Oeste. A migração trans-Atlântica ocorreu provavelmente através da Corrente Norte-Equatorial, entre África Oeste e o nordeste do continente Sul-Americano, expandindo-se para norte até aos 42º graus de latitude.

Estudos anteriores sugeriram que o padrão clinal observado no ADN mitocondrial do biqueirão se deveu a divergência em isolamento dos grupos genéticos provocado pelos ciclos climáticos do Pleistoceno e que a distribuição actual resulta de um contacto secundário. Coligimos evidência que suporta a hipótese de o clade com distribuição anti-tropical possui um codão no gene mitocondrial do citocromo *b* que se encontra sob selecção positiva. Embora não tenhamos detectado uma alteração na conformação da proteína, os testes de selecção foram todos significativamente positivos. Explorámos a possibilidade de a variação da frequência dos grupos mitocondriais do biqueirão estar associada a diferentes variáveis ambientais, temperatura, salinidade, utilização aparente de oxigénio e a concentração de nutrientes como os fosfatos, os nitratos e os silicatos. Detectámos uma correlação elevada entre a frequência dos grupos mitocondriais e a temperatura da água, em que esta contribui com 57% para a variabilidade dos grupos ao longo da área de distribuição. A temperatura é o factor que melhor prediz a distribuição contemporânea da frequência dos dois grupos mitocondriais do biqueirão. Esta possibilidade contradiz a hipótese anterior e obriga a uma re-interpretação dos resultados previamente obtidos para o grupo B, uma vez que a variação genética não é neutral. A selecção mitocondrial em peixes ectotérmicos é particularmente importante, uma vez que promove alterações na cadeia respiratória celular e, consequentemente, maior oxigenação e/ ou maior quantidade de energia transmitida aos tecidos musculares. A estrutura populacional complexa que esta espécie apresenta pode exactamente reflectir o resultado de uma melhor adaptação de uma parte dos indivíduos a águas mais frias, fazendo com que estes peixes tenham maior preferência por uma dispersão local, evitando as águas tropicais

O complexo de espécies designado por “Old World Anchovies”, do qual o biqueirão faz parte, é composto por seis espécies putativas (*E. encrasicolus, E. eurystole, E. capensis, E. albidus, E. japonicus* and *E. australis*) que compõem este complexo foram descritas com base no suposto isolamento geográfico entre as

diferentes áreas de distribuição, já que os caracteres morfológicos se sobrepõem e as divergências genéticas são muito baixas. Evidências anteriores apontam para uma origem deste complexo durante o Mioceno/ Plioceno, resultante de uma colonização do Pacífico Este para o Pacífico Oeste. No Oceano Pacífico foram descritas as espécies *E. japonicus* e *E. australis*, cuja distribuição ocupa zonas costeiras do noroeste e sudoeste desta bacia hidrográfica, respectivamente. Estudos anteriores inferiram que a colonização da bacia Atlântica deverá ter ocorrido devido a pelo menos duas migrações, uma vez que as linhagens mitocondriais das "old world anchovies" Atlânticas não são monofiléticas. Foi ainda proposto que estas colonizações deveriam ter ocorrido através do Índico Sul, onde as águas são temperadas, já que se assumia que os biqueirões não existiam em águas tropicais. Na bacia Atlântica foram previamente descritas as espécies anchova Europeia *E. encrasicolus* e a anchova do Cabo *E. capensis* no Atlântico Este, a anchova prateada *E. eurystole* no Atlântico Oeste. Mais recentemente foi descrita uma outra espécie, a anchova branca *E. albidus*, com distribuição muito restricta em lagoas coasteiras do noroeste Mediterrâneo. Os nossos resultados confirmam que a origem deste complexo foi há cerca de 3.16 Ma no Indo-Pacífico no fim do Plioceno e que o complexo se dividiu em dois grupos: um que ficou no Pacífico e outro que colonizou o Oceano Atlântico durante o Pleistoceno (0.62 Ma). Contudo, a tolerância a águas quentes (confirmada pela captura de indivíduos na Guiné-Bissau, no Ghana e em Israel), a ausência de plataforma continental no Índico Sul e a partilha de haplótipos entre indivíduos do Japão e da África do Sul sugerem que a plataforma continental do Sul da Ásia, do Médio Oriente e de África Oriental foi rota de migração mais provável entre os oceanos Pacífico e Atlântico. Os padrões genéticos das "old world anchovies", indicam que não há diferenciação genética entre as espécies putativas do Oceano Atlântico, e que os níveis de migração contemporânea são baixos entre os biqueirões do Atlântico e do Pacífico. A parafilia entre os dois clades Atlânticos que tanto tem confundido os autores em trabalhos anteriores, dever-se-à ao efeito da selecção no clade B. No Oceano Pacífico, os dois clades mitocondriais revelam migrações trans-Equatoriais anteriores ao último máximo glaciar, enquanto que os loci nucleares suportam migrações contemporâneas.

PALAVRAS-CHAVE: Engraulidae, *Engraulis encrasicolus*, Anchovas do Mundo Antigo, selecção natural, filogeografia, biogeografia, migração, dispersão, especiação, frente de colonização

STRUCTURE OF THE THESIS

This thesis is organized in five chapters. The first chapter consists of an introduction to the subject and presents the general aims of the thesis. Chapters II to IV are autonomous studies that address the proposed goals. These chapters constitute scientific papers, in press or submitted, and therefore can be read separately.

Chapter II. Genetic patterns of leading-edge populations in the North Atlantic anchovies.

Chapter III. Selection in the mitochondrial DNA of the European anchovy *E. encrasicolus* and environmental correlates as possible factors contributing to maintain mitochondrial clades.

Chapter IV. Evolution of the Old World Anchovies *Engraulis spp.* by assessing the reproductive isolation among putative species within the group and analyzing the connectivity between the Atlantic and Pacific anchovies. We explored a mitochondrial phylogeny of the Engraulidae family to date main lineage splitting events to determine the age of the species complex and propose a biogeographical scenario.

Chapter V. Inclusive synthesis of the overall contribution and future perspectives.

In front of Chapters II to IV there is the information on the co-authors involved in the publication and the status of publication.

CHAPTER I ● GENERAL INTRODUCTION

Species evolution relies on the ability to adapt to environmental changes, based on the genetic variation that confers plasticity either in decadal or millennia periods (Lande & Shannon, 1996). Therefore, the measurement of the genetic variation is pivotal to understand the processes that shape species distributions, one of the major goals of evolutionary biology. Patterns of genetic diversity and differentiation result from the interplay between random and deterministic processes (Hedrick, 2005), acting over multiple time-scales (e.g. Hewitt, 1996, 2000). Mutation, recombination, genetic drift, gene-flow and selection combined with historical and contemporary climatic events shape both within and between populations genetic architecture (Hedrick, 2005). The effect of these evolutionary forces depends on species population size, life-history traits, how populations are spatially arranged and how genes evolve to adapt to external pressures. Understanding the relative role of each driver is thus important to understand how species evolve and how they respond to adversities since that in extreme cases may lead to speciation processes (Coyne & Orr, 2004). The molecular study of these genetic patterns may provide evidences of which evolutionary forces are driving species evolution, and this knowledge may allow the prediction of species responses to climate change when defining conservation measures (Moritz, 1994).

CLIMATIC OSCILLATIONS

Living organisms inhabit almost everywhere on Earth, from deep-sea thermal vents to the highest mountain peaks, and survive to the most extreme environmental regimes. However, no single species lives in all of these places, having a restricted range imposed by their specific physiological tolerances (Brown & Lomolino, 1998). Species distribution ranges and diversity are limited by environmental features, which were not constant over Earth's history. Two major forces mostly influence the Planet Earth, the heat that comes from its core stored at the time of formation of the planet and the energy that is irradiated from the Sun (Brown & Lomolino, 1998). The latter is dissipated through the mantle and the crust, and promotes large-scale convection process driving tectonic movements, volcanic eruptions and earthquakes. The energy irradiated from the Sun is converted into heat that warms the planet surface. Differences in temperature and density promote air and water circulation, causing predominant winds and oceanic currents. However, Earth's movements and

orientation are not constant. The relative position and distance to the Sun oscillates continuously and alters the amount of solar radiation that reaches the Earth's surface. As the Earth spins around its axis and orbits around the Sun, three distinct gravitational processes (orbital eccentricity, obliquity and precession) produce quasi-periodic variations on Earth's movements that result in cyclical climatic oscillations known as Croll-Milankovitch climate cycles (Figure 1.1; Hays *et al.*, 1976). The combined effect of these three deviations in Earth's movements occurs around each 100ky, leading to periods of extreme climate (Figure 1.2).

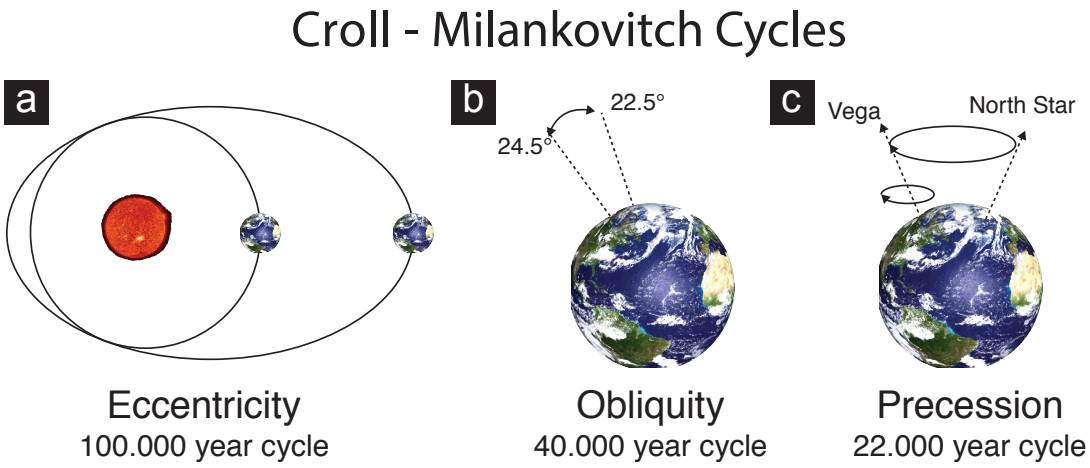


Figure 1.1 - Collective effects of changes in the orbit movements of Planet Earth (Croll-Milankovitch cycles).

In addition to Earth's orbital changes, geological events also have had major impacts on Earth's climate. For instance, before the formation of the Panama Isthmus, Pacific surface waters flowed into the Atlantic Ocean, and the mixing of these waters balanced the two oceans' salinity. The formation of the Isthmus of Panama occurred during the Pliocene epoch (5.3 - 2.6 Ma; Lessios *et al.*, 1999) divided the Atlantic and the Pacific oceans and forced a rearrangement of global ocean circulation (Figure 1.3).

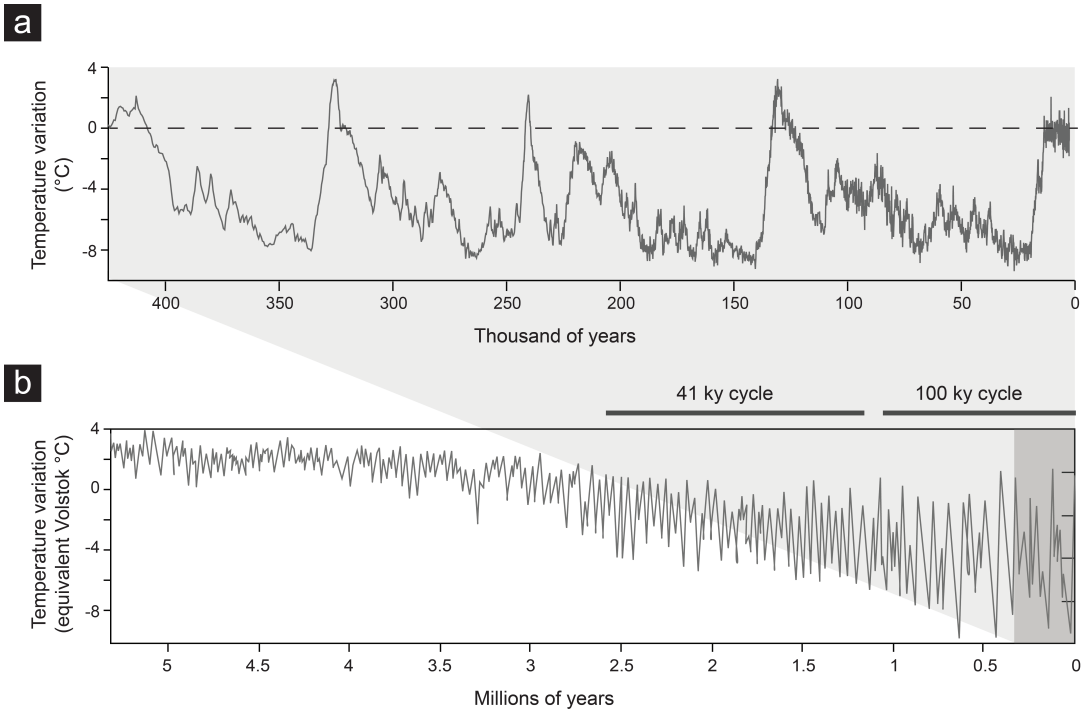


Figure 1.2 - Antarctic temperature variation; a) Reconstructed temperature from the Vostok ice core for the past 420ky (Petit *et al.*, 1999); b) reconstructed temperature by combining measurements from 57 globally distributed deep-sea sediment cores for the past 5 MY (Lisiecki & Raymo, 2005).

After the closure of the Central American seaway the planet cooled down, the Arctic ice cap was formed and the Quaternary period was initiated (2.6 Ma- present). Over the Quaternary, Croll-Milankovitch cycles contributed to Earth's cyclical cooling and warming, promoting ice sheets advances and retreats.

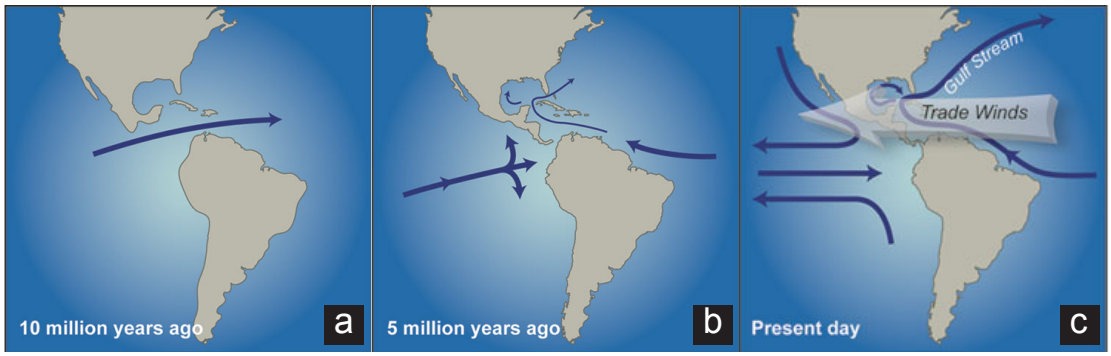


Figure 1.3 - Panama Isthmus formation; a) 10 MY ago the Atlantic and Pacific oceans were connected; b) 5MY ago the Isthmus of Panama was on progressive formation restricting water flow between the Pacific and the Atlantic oceans; c) present-day geologic configuration of the Isthmus of Panama; blue arrows indicate ocean currents (Source: adapted from http://www.whoi.edu/oceanus/feature/how-the-isthmus-of-panama-put-ice-in-the-arctic).

The Last Glacial Maximum (LGM) was the last and most severe glacial period over the Pleistocene (18 - 21 ka), with the ice sheet covering much of North America, northern Europe and Asia. In Europe, the ice sheet extended southwards to the English Channel, covering most of the North and Baltic seas, although some brackish lakes remained within the ice sheet (CLIMAP Project Members, 1984; Maggs *et al.*, 2008; Kettle *et al.*, 2011). Sea surface temperature dropped 10ºC in the North Atlantic and isotherms were compressed (Figure 1.4). Global patterns of wind, humidity and ocean currents affected primary productivity and, consequently, the organisms that live at the ocean's surface. As a result, sea level dropped 120-140 m and salinity increased, affecting the circulation patterns by changing the strength and location of sea currents (Lambeck *et al.*, 2002; Huddart & Stott, 2010), as the southward shift of the Gulf Stream (Otto-Bliesner *et al.*, 2006).

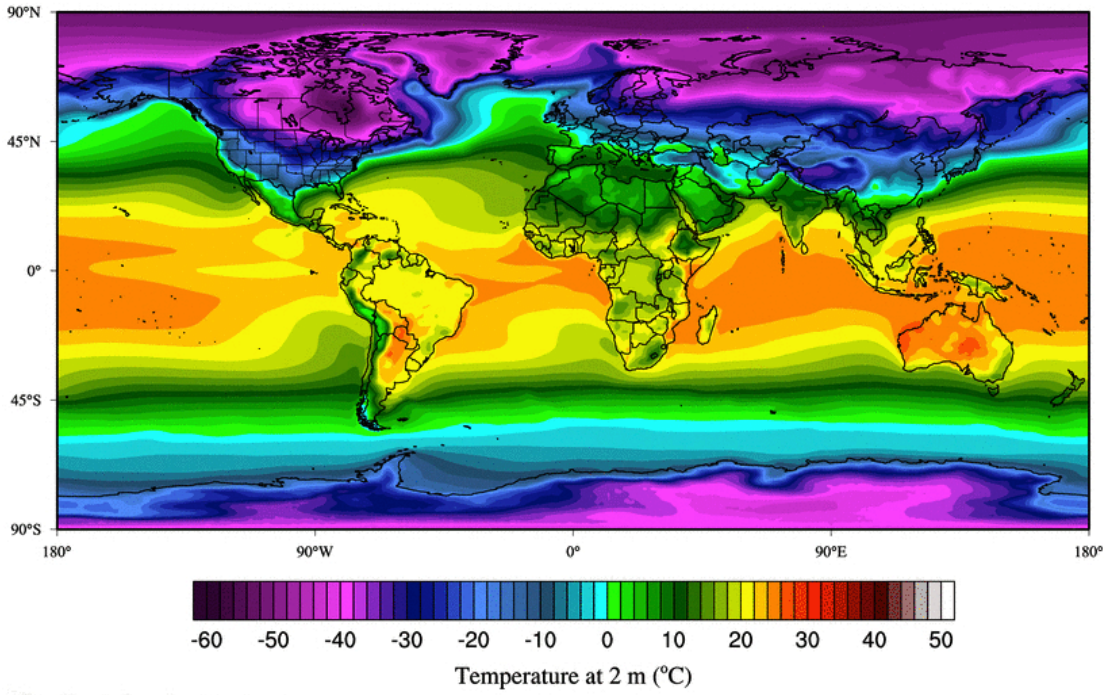


Figure 1.4 - Community Climate System Model version 4 (CCSM4) reconstruction of sea surface temperature in January (◦C) at the Last Glacial Maximum (21 ky). (Source: http://cci-reanalyzer.org/animations/scycle/CCSM4-LGM_T2_scycle.gif)

After the LGM, temperatures increased, ice sheets retreated to higher latitudes and modern circulation patterns were established (both oceanographic and atmospheric). Earth climate system entered in an interglacial period, the Holocene (12 ka - present). Although minor shifts in temperatures occurred during this period (Crowley, 2000; Walker, 2004), environmental conditions remained more or less globally stable. The

modern North Atlantic biota was established when the ice sheet retreated completely in the Baltic Sea around 8.5 ka (Berglund *et al.*, 2005).

PRESENT-DAY OCEANOGRAPHY

The distributions and dynamics of marine organisms are strongly influenced by present-day physical oceanographic processes and geological features. Marine organisms respond to environmental fluctuations with range shifts and variation in population size, either in seasonal, decadal or millennia periods. Current oceanographic patterns were roughly established after the closure of the Isthmus of Panama (Figure 1.5), despite variations in the intensity of winds and currents, temperatures and ice sheets volumes.

In the Atlantic Ocean, two large gyres drive the surface water circulation: the North Atlantic gyre and the South Atlantic gyre (Figure 1.5). In the first, westerly winds blow from west to east enhancing the Gulf Stream, while trade winds blow from east to west enhancing equatorial currents. More localized phenomena, such as fluctuations in the atmospheric pressure at the sea level originate the North Atlantic Oscillation (NAO; Wallace & Gutzler, 1981), that regulates the strength and direction of westerly winds in this area and consequently surface currents. In the South Atlantic gyre, cold upwelled water from the Antarctic Ocean flows northwards along the African coast to the tropics where it is transported to the south American coast by the South Equatorial Current (Mariano, 2013). The Indian Ocean holds one of the major oceanic gyres, where the warm currents mostly flow from east to west in the northern part, while cold currents flow from southwest to east in the southern part. The Pacific Ocean is also characterized by two large gyres, the North Pacific gyre and the South Pacific gyre. The North Pacific gyre has a clockwise circular pattern and transfers the upwelled cold water in the northeastern Pacific to the western Pacific enhanced by trade winds, while the South Pacific gyre transports water masses with low levels of nutrients in counter clockwise.

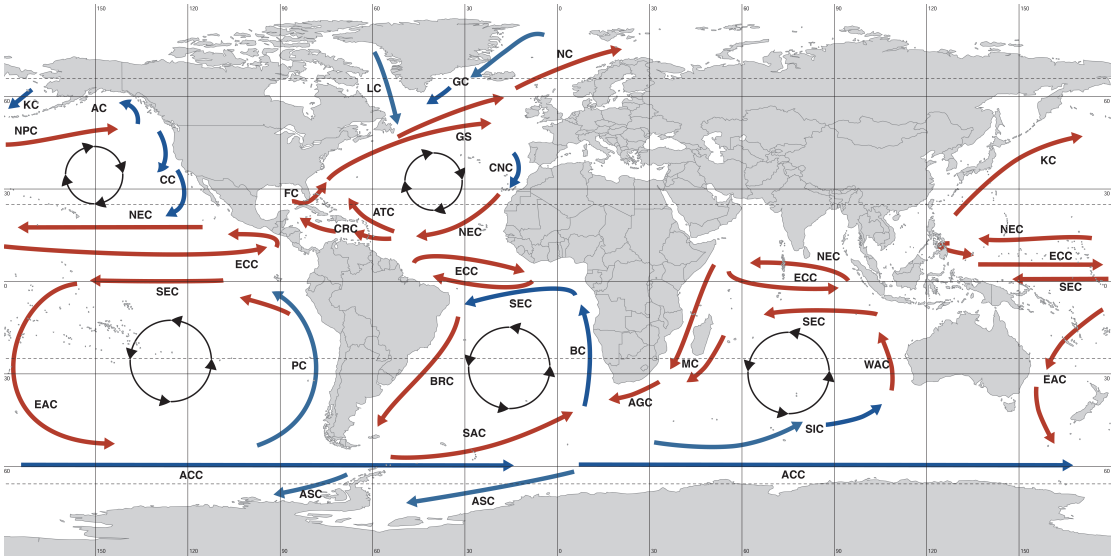


Figure 1.5 - World sea surface currents; red - warm currents; blue - cold currents; arrows point direction of the current; black circles - major oceanic gyres; KC - Kamchatka current; AC - Alaska current; NPC - North Pacific current; CC - California current; NEC - North Equatorial current; ECC - Equatorial counter current; SEC - South Equatorial current; EAC - Eastern Australia current; PC - Peru current; ACC - Antarctic Circumpolar current; ASC - Antarctic Sub-circumpolar current; LC - Labrador current; GC - Greenland current; NC - Norwegian current; GS - Gulf Stream; CNC - Canaries current; ATC - Antilles current; FC - Florida current; CRC - Caribbean current; BRC - Brazil current; SAC - South Atlantic current; BC - Benguela current; AGC - Agulhas current; MC - Mozambique current; SIC - South Indian current; WAC - Western Australia current; KC - Kuroshio current.

Surface currents and predominant winds also influence the spatial patterns of sea surface temperature (SST) and salinity. Marine fishes are ectothermic and their metabolism is regulated by water temperature, despite some exceptions. Therefore, their distribution ranges are bounded by their physiological tolerance to particular temperature limits. SST changes on inter-annual and longer timescales are influenced by the combination of atmospheric and oceanic processes. As observed in figure 1.6, SST presents its minimum in the polar seas (-2ºC) and increases towards the tropics (in the Persian Gulf can be as warm as 36ºC) in a latitudinal gradient. Regional variations in SST can cause periodic anomalies such as the Atlantic multi-decadal Oscillation (AMO) or the east-central Pacific *El Niño-Southern Oscillation* (ENSO). These cycles are triggered by differences in temperature between the ocean and the atmosphere and cause breaks on primary productivity influencing the stability and structure of food webs. These regional phenomena thus have dramatic consequences in the marine ecosystems (Doney *et al.*, 2012), imposing strong fluctuations in the distribution abundances and of marine organisms (Rose, 2005; Alheit *et al.*, 2012).

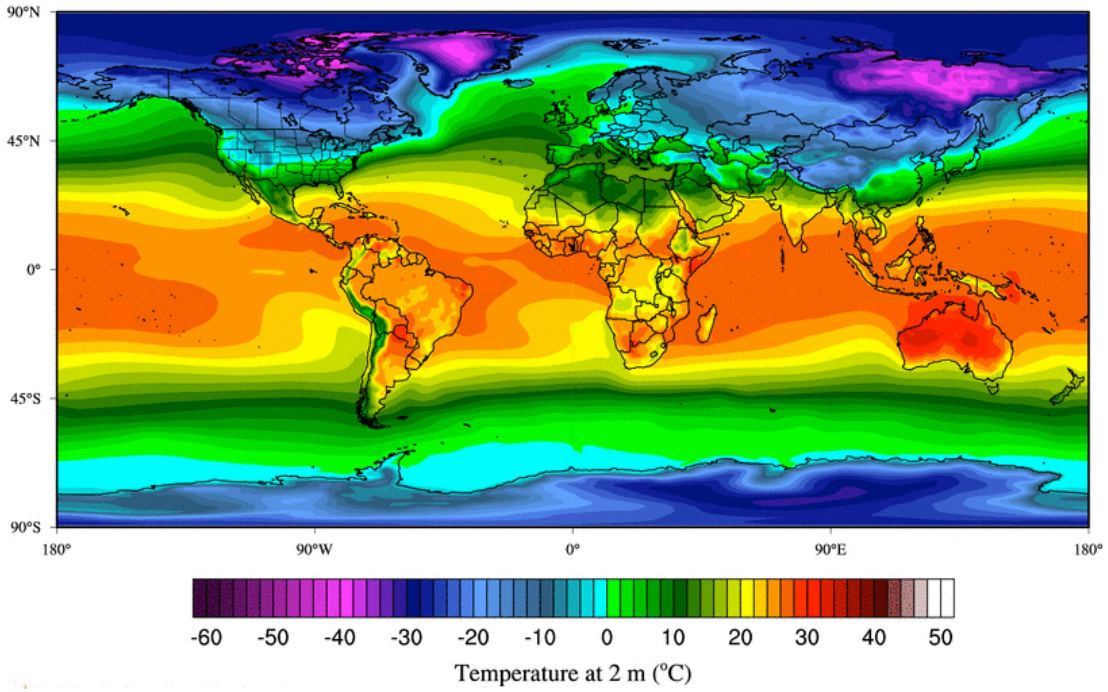


Figure 1.6 - Global ocean sea surface temperature ERA-Interim reanalysis: January mean temperature from 1979-2000.
(Source: Data/image obtained using Climate Reanalyzer™, http://cci-reanalyzer.org, Climate Change Institute, University of Maine, Orono, Maine, USA)

Salinity, as well as temperature, plays an important role on global ocean circulation. Salinity regulates the density of the water and influences the sinking and rising of water masses. Average salinity in the Atlantic is higher than in the Pacific and Indian oceans, while inland seas present different composition as compared to oceanic waters (Figure 1.7). On average, ocean salinity is 35 psu, but this value ranges from 3-5 psu in riverine environments, and the Baltic and Black seas, to more than 39 psu in the Mediterranean Sea.

The interplay between environmental factors promotes different biomes with distinct physical properties in different geographical regions. The transitional areas between those habitats are denominated ecotones (Livingston, 1903), environmental stress zones that may be gradual across a broad area forming clines (e.g. temperature gradient in the eastern Atlantic) or may be sharp (e.g. salinity in the Black sea vs. Mediterranean sea). Species whose distribution ranges encompass two or more biomes must be adapted to contrasting environmental regimes, and may exhibit

phenotypic and genomic divergence. Furthermore, ecotones undertake a key role in the understanding of species adaptation mechanisms to environmental features (Kark & van Rensburg, 2006).

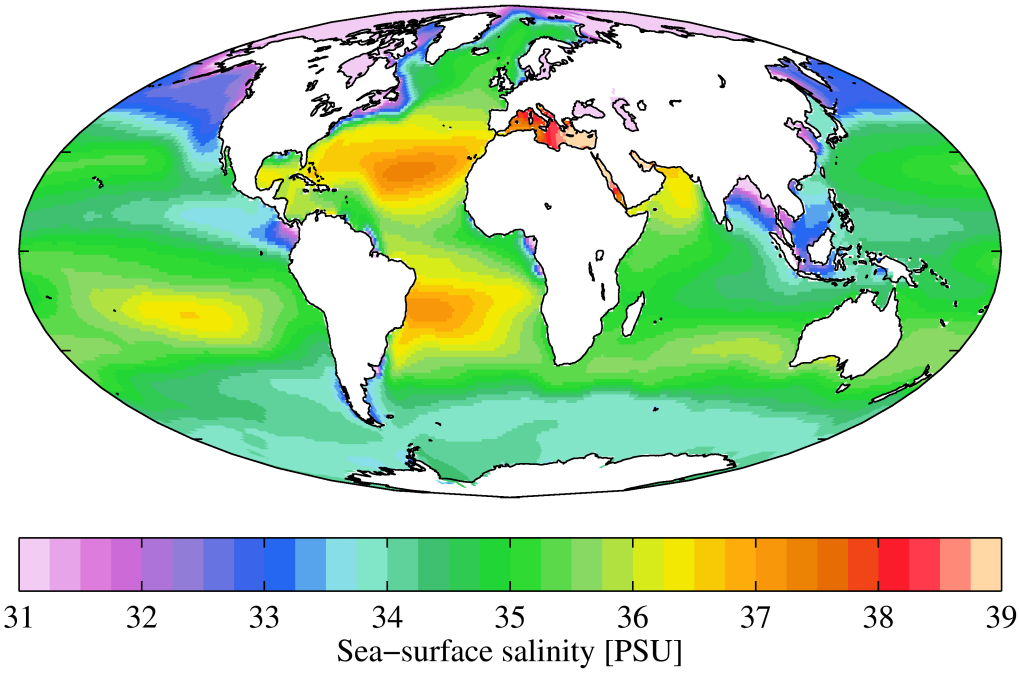


Figure 1.7 - Annual mean sea surface salinity from the World Ocean Atlas 2009. Salinity here is in practical salinity units (PSU).
(Source: http://www.nodc.noaa.gov/OC5/WOA09/pr_woa09.html)

PHYLOGEOGRAPHY - HISTORICAL EVALUATION OF GENETIC VARIATION

Phylogeography was first used by Avise *et al.* (1987) and is the use of phylogenetic methods in population studies in the context of geography. Phylogeography is the study of the historical processes that may be responsible for the contemporary geographic distribution of individuals, under the light of the patterns associated with gene genealogies (Avise, 2000; Hewitt, 2000). Moreover, phylogeography is mostly based on the interpretation of genetic patterns through time and space, supported by fossil records, biogeographic events and climatologic simulations (Avise, 2000; Hewitt, 2000). Nevertheless, depending on the molecular marker, genetic imprints can be framed on time, although it may be difficult to disentangle the contribution of contemporary signatures and historical polymorphism (Avise, 2000). Mitochondrial

DNA (mtDNA) has been commonly used in the last 35 years to infer populations divergence and past range dynamics due to its unique properties, namely: 1) abundant and easy to work with; 2) haploid; 3) uniparental inheritance; 4) lack of recombination which allows to follow genealogic lineages; 5) mutation rate is higher than nuclear genes and slower than microsatellites; and 6) selectively neutral (Dowling *et al.*, 2008). Furthermore, mutation rates of mtDNA frame the evolutionary time of the genetic imprints within the late Pleistocene, contributing to the knowledge of species response to Quaternary climate change (Avise, 2000). Theoretically, mtDNA represents and ideal marker to address phylogeographical questions, but anomalies to the assumptions have been discovered (Dowling *et al.*, 2008 and references therein), including selection, mito-nuclear gene interactions, introgression (Ballard & Whitlock, 2004) and even recombination (Kraytsberg *et al.*, 2004). To circumvent potential problems, scientists have tried to increase the power of statistical analysis and add more data to detect possible deviations to neutrality. Phylogeography became more integrative (e.g. Bayesian analysis, ecological modelling) and comparative (e.g. super trees, genomics), benefitting from computation power and newly massive sequence techniques (e.g. 454, Illumina). Also, the development of the coalescent theory (Kingman, 1982) allowed testing model-based methods (e.g. Knowles & Maddison, 2002), estimates of gene flow and time since divergence (Beaumont & Rannala, 2004), impossible to address using traditional descriptive approaches.

PHYLOGEOGRAPHY OF MARINE FISH IN THE ATLANTIC OCEAN

In the marine realm as in terrestrial biotas, organisms respond to climate changes with demographic fluctuations, range shifting and adaptation to newly established environmental conditions and habitats (Avise, 2000; Hewitt, 2000). While in terrestrial organisms the effect of the Pleistocene climate cycles is well known (Hewitt, 2000), in the marine realm is more challenging to interpret. The marine environment is theoretically continuous with the apparent lack of physical barriers to dispersal and gene flow. Also, there are less fossil records (e.g. Provan & Bennett, 2008) and organisms are more difficult to track (Bohonak, 1999). Nevertheless, phylogeographic studies in the Atlantic Ocean have revealed geographically distinct

lineages arrangements, glacial refugia (Maggs *et al.*, 2008; Kettle *et al.*, 2011), phylogenetic breaks (Patarnello *et al.*, 2007), barriers to dispersal (Floeter *et al.*, 2008), range shifts (Maggs *et al.*, 2008; Kettle *et al.*, 2011), including trans-Atlantic colonisations (Wares & Cunningham, 2001; Luiz-Júnior *et al.*, 2004), and allowed inferring the role of different evolutionary forces driving genetic patterns (e.g. Riginos *et al.*, 2004; Engelhaupt *et al.*, 2009; Teacher *et al.*, 2012). In some species, historical signatures were erased by the homogenizing effects of gene-flow (e.g. *Petromyzon marinus*; Almada *et al.*, 2008; e.g. *Clupea harengus*; Gaggiotti *et al.*, 2009), but others exhibit remarkable population structure and/ or retain the genetic imprints of past demographic shifts. Expected genetic differentiation has been found in intertidal fish (e.g. common goby *Pomatoschistus microps*; Gysels *et al.*, 2004), in reef-associated fish (e.g. corkwing wrasse *Symphodus melops*; Robalo *et al.*, 2012) and in demersal fish (e.g. thornback ray *Raja clavata*; Chevolot *et al.*, 2006). Surprisingly, marked genetic differentiation was found even in species with high dispersal ability such as the schooling pelagics European anchovy *Engraulis encrasicolus* (Magoulas *et al.*, 2006) and European sprat *Sprattus sprattus* (Debes *et al.*, 2008; Limborg *et al.*, 2012).

RANGE SHIFTS AND GLACIAL REFUGIA

Pleistocene climate oscillations enhanced species distribution displacements, as their physiological tolerances were challenged. In terrestrial environments, species often contracted their distribution ranges to southern refugia during glacial periods and expanded northwards to their climatic optima (Hewitt, 2004). Therefore, range shifts influenced the genetic characteristics of populations, with genetic diversity being preserved in refugial areas, whereas peripheral populations are extirpated and suffer strong bottlenecks at each climatic event (McInerny *et al.*, 2009). Populations inhabiting previously glaciated areas typically show lower genetic diversities and more shared haplotypes than populations from southern refugial latitudes (Maggs *et al.*, 2008). This trend translates not only the periphery of the distribution range where populations are sparse, less abundant and at the geographic limit of the physiological tolerances (McInerny *et al.*, 2009), but also a recent colonization that occurred after the LGM (Maggs *et al.*, 2008). Nevertheless, genetic imprints of strong climatic events that impose population shifts and severe bottlenecks, as the case of the LGM,

may erase signatures of previous population oscillations (Grant *et al.*, 2012).

In the marine realm, species responses to climate change may not follow the terrestrial contraction-expansion refugia models, as most species are able to track suitable isotherms and adjust their distribution (Figure 1.8 a, c and d). As described in figure 1.8 (b and d), strict contraction to refugial areas may represent a small portion of range shifts, as the majority of species have pelagic phases and can disperse either actively, or passively through currents. Species with low dispersal abilities (e.g. sessile species with no pelagic larval stages; terrestrial species), lacking suitable habitat southwards (Figure 1.8 b) or those trapped within the ice sheet in inhabitable areas may have persisted in refugia (Figure 1.8 d). Nevertheless, some organisms may have not been able to find suitable habitats and became extinct (Figure 1.8 e). Species that are capable to disperse, at least during larval pelagic phases, are most likely to have tracked suitable habitats along the shores (Figure 1.8 a, c, d and f). Areas where species were able to persist continuously during both glacial and interglacial periods (presently denominated as refugia), are expected to harbour stable and old populations and thus represent diversity hotspots (Galarza *et al.*, 2009), since there was a preservation of genetic diversity over the Pleistocene (Maggs *et al.*, 2008). On the other hand at the leading- and rear-edges, genetic diversity is mostly attributed to colonisation events departing from biodiversity hotspots, since these populations were cyclically extirpated (Hewitt, 1999). Species with large population sizes and high dispersal ability are expected to migrate massively and preserve their diversity (Figure 1.8 f; Andre *et al.*, 2011). Temperate species with large distribution ranges (e.g. transequatorial) and consequently with large thermal tolerance ranges (e.g. European anchovy *E. encrasicolus*) may have contracted at the leading-edge during cooling periods (Zarraonaindia *et al.*, 2012), but no range shifts in the rear edge are expected during warming periods (Figure 1.8 c).

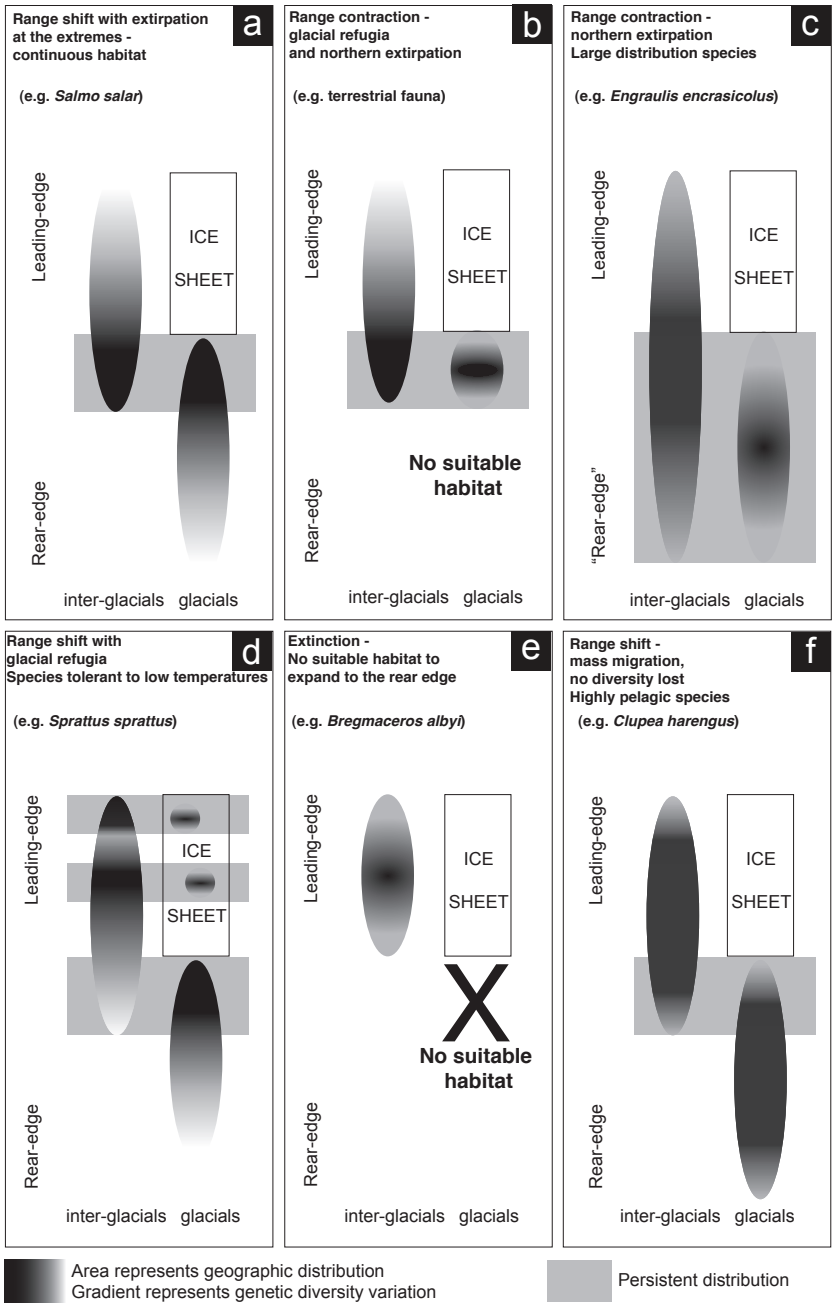


Figure 1.8 - Range shifts of temperate marine organisms imposed by glacial and interglacial periods in the Pleistocene; a) species able to track suitable habitats and genetic diversity was maintained were distribution was persistent both in glacial and inter-glacial periods; b) species not able to track suitable habitats and genetic diversity was maintained were distribution was persistent both in glacial and inter-glacial periods; c) species whose distribution range contracted at the leading-edge and genetic diversity was maintained were distribution was persistent both in glacial and inter-glacial periods; d) species able to track suitable habitats and genetic diversity was maintained were distribution was persistent both in glacial and inter-glacial periods, including refugia within the glacial ice-sheet; e) species that were not able to survive to Pleistocene shifts and became extinct; f) species with large population sizes and high dispersal ability were capable to move massively and thus preserving diversity.

Several authors pointed some potential marine refugia in the North Atlantic, namely: 1) Iceland and Faroe islands; 2) the Norway Deep; 3) the Gdansk Deep; 4) southwestern Ireland; 5) the Hurd Deep in the English Channel; 6) northwestern Iberian Peninsula; 7) Mediterranean Sea; 8) Azores, Canary Islands and northwest Africa; 9) New Foundland and the Canadian Grand Banks; and 10) from Carolinas to Florida and the Gulf of Mexico (Figure 1.9).

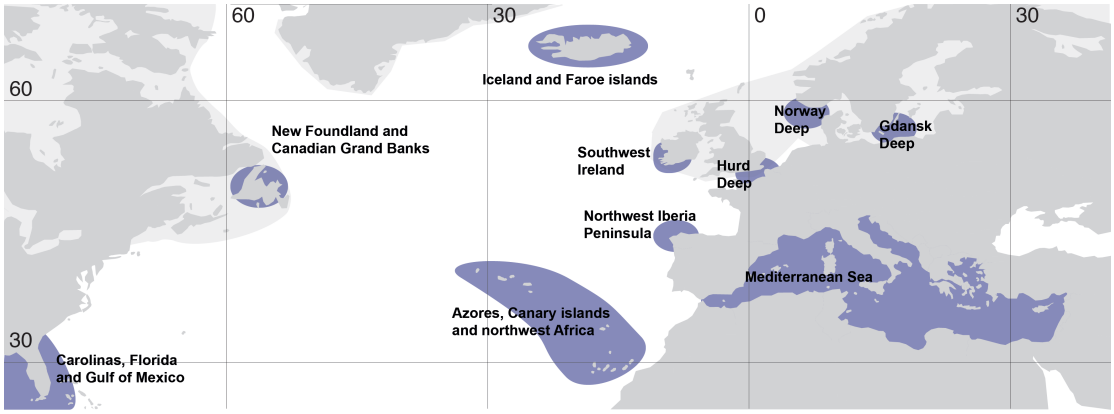


Figure 1.9 – a) Present-day Atlantic Ocean and Mediterranean Sea showing putative Last Glacial Maximum refugia (purple shading); light grey shading represents the extent of terrestrial ice sheets at the Last Glacial Maximum (18–20 ka) (source: Bigg *et al.*, 2008; Maggs *et al.*, 2008; Gaggiotti *et al.*, 2009; Kettle *et al.*, 2011).

Pleistocene climatic oscillations also provided the opportunity for species to find suitable environmental conditions to cross the Atlantic. Species with amphi-Atlantic distributions imply recent or historical trans-Atlantic migration(s) within periods of favourable eco-physiologic conditions. The origin of a given species with amphi-Atlantic distribution is based on the larger geographic area that the species occupy in one of the sides of the Atlantic Ocean vs. its narrower distribution range on the other side (Briggs, 1974; Briggs, 1995). Most of organisms with amphi-Atlantic distributions have their origin on the western Atlantic (Briggs, 1974), but some may have migrated from westwards (Joyeux *et al.*, 2003). Trans-Atlantic dispersals may occur through four main routes: 1) along the continental platform in the northern extreme (Vermeij, 2005); 2) from the Caribbean to north-east Atlantic through the Gulf Stream; 3) from northern Brazil to the Gulf of Guinea by the Equatorial Counter Current; 4) from southern Africa to southern Brazil through the Southern Equatorial Current (see Figure 1.5; Joyeux *et al.*, 2001). Trans-Atlantic colonisations may have been favoured by the increasing of the intensity of the prevailing westerlies during

interglacial periods, and by the speeding up of trade winds during cooling episodes (Janecek & Rea, 1985). Dispersal across the North Atlantic was likely limited to interglacial periods, when warming allowed stepping-stone colonisations through the North Atlantic islands, Greenland and Canada (Vermeij, 2005). For instance, the three-spined stickleback *Gasterosteus aculeatus* (Makinen & Merila, 2008) and the near shore fish *Pholis gunnellus* (Hickerson & Cunningham, 2006) may have benefited from the Gulf Stream and westerlies enhancements during interglacial periods in west-to-east colonisations. The European herring (*Clupea harengus*) distribution and genetic patterns suggest a westwards post-LGM migration through the continental platform (Bekkevold *et al.*, 2005). Most of equatorial trans-Atlantic migrations occurred eastwards (e.g. rock hind *Epinephelus adscensionis*; Joyeux *et al.*, 2001; Carlin *et al.*, 2003), but four fish species were reported to have migrated in the opposite direction, namely *Acanthurus monroviae* and *Parablenius pilicornis* (Luiz-Júnior *et al.*, 2004), *Epinephelus marginatus* (Joyeux *et al.*, 2001) and *Aulostomus strigosus* (Bowen *et al.*, 2001). East-to-west migrants may have reached the western Atlantic through the South Equatorial Current (Joyeux *et al.*, 2001). The effect of climatic events on ocean currents may play an important role in trans-Atlantic colonisations, by enhancing periods of dispersal alternating with isolation (Philander, 1986; Joyeux *et al.*, 2001; Carlin *et al.*, 2003).

PHYLOGEOGRAPHIC BREAKS AND BARRIERS TO DISPERSAL

Barriers to dispersal, either historical or contemporary, may constrain or prevent species to track suitable habitats during range shifts. Comparative phylogeographic studies revealed concordant phylogeographic breaks within species distributions (Patarnello *et al.*, 2007; Floeter *et al.*, 2008; Maggs *et al.*, 2008) or concordant distributions with the same range boundaries in several organisms (Spalding *et al.*, 2007). In the Atlantic Ocean and Mediterranean Sea several barriers to dispersal have been identified (Figure 1.10). These include land barriers that promoted allopatric speciation, such as the closure of the Isthmus of Panama (Lessios *et al.*, 1999; Leigh *et al.*, 2013) or the separation of the Mediterranean and Red seas after the closure of the ancient Tethys Sea (Roëgl, 1998). For example, the sister species *Cetengraulis edentulus* and *C. mysticetus* evolved in isolation after the closure of the Isthmus of Panama (Grant *et al.*, 2010). Softer barriers such as the Amazon discharge, the Mid-

Atlantic Ridge, the Benguela current and the Cape Horn restricted dispersal and promoted the development of new species in allopatry and/or parapatry (Grant *et al.*, 2005; Floeter *et al.*, 2008). The Peruvian anchovy (*Engraulis ringens*) and the Argentine anchovy (*Engraulis anchoita*) are likely to have diverged during the mid Pliocene onset of global cooling that divided a former continuous distribution around the tip of South America (Grant *et al.*, 2005). Phylogeographic breaks also revealed recent divergence within species related to post-glacial range shifts in the Baltic Sea and the North Sea (Olsen *et al.*, 2004), the English Channel (Jolly *et al.*, 2005) or the Bay of Biscay (Larmuseau *et al.*, 2009). In the Mediterranean sea, genetic discontinuities are mostly attributed to retentive currents in the Almeria-Oran front (Patarnello *et al.*, 2007) or to different oceanographic features in the central Mediterranean (Sá-Pinto *et al.*, 2012 and references therein). In the western Atlantic, phylogeographic breaks were detected in Cape Hatteras where the Labrador Current and the Gulf Stream have met periodically during the Quaternary (Weinberg *et al.*, 2003) and in Cape Canaveral, that divides the Atlantic ocean from the Gulf of Mexico (reviewed in Avise, 1992). However, soft barriers to dispersal are not effective for all organisms, showing that the strength of environmental factors impeding connectivity is permeable to some species and do not depend on life-history traits (Wares & Cunningham, 2001; Patarnello *et al.*, 2007; Galarza *et al.*, 2009).

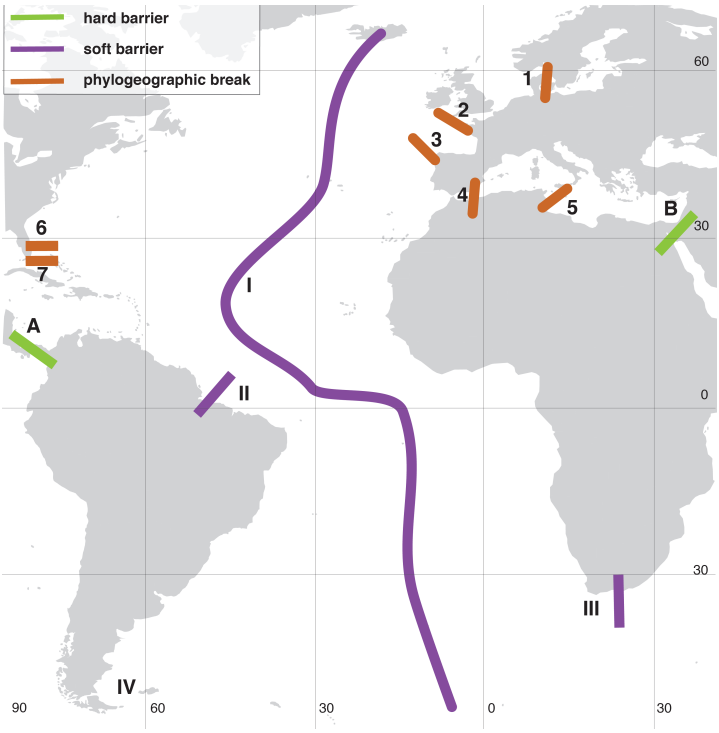


Figure 1.10 - Barriers to dispersal in the Atlantic ocean and Mediterranean sea; hard barriers - A) Isthmus of Panama, B) Mediterranean-Red sea land bridge; soft barriers - I) Mid-Atlantic ridge, II) Amazon discharge, III) Benguela current, IV) Cape Horn; phylogeographic breaks - 1) Baltic sea -North sea, 2) English Channel, 3) Bay of Biscay, 4) Almeria-Oran front, 5) central Mediterranean, 6) Cape Hatteras, 7) Cape Canaveral.

ADAPTATION

As climate changes, organisms shift their distributions and adapt to newly established conditions. Several marine organisms in the North Atlantic and Mediterranean Sea provide prime examples of adaptation and adaptive divergence that influenced either phenotypic or genotypic patterns. Sympatric ecomorphs of the anadromous fishes three-spined sticklebacks (*Gasterosteus aculeatus*; Taylor & McPhail, 2000) and Arctic charr (*Salvelinus alpinus*; Wilson *et al.*, 2004) are specialized to limnetic or benthic habitats within the same lake. The white anchovy *Engraulis albidus* cryptically diverged from the European anchovy *E. encrasicolus*, possibly as a result to be better adapted to estuarine/ inshore waters (Borsa *et al.*, 2004).

Natural selection in the marine environment is often related to thermal and saline gradients or to resistance to human-mediated pollutants (Galtier *et al.*, 2009 and references therein). The Arctic charr *S. alpinus* mtDNA genome has introgressed into the temperate brook charr *S. fontinalis*, providing a selective advantage to colder waters (Doiron *et al.*, 2002). The Atlantic cod *Gadus morhua* showed selective sweeps and exhibits latitudinal clines on nuclear temperature-associated genes in response to ocean temperature (Bradbury *et al.*, 2010). The killifish *Fundulus heteroclitus* has the ability to upregulate heat shock proteins expression in response to local thermal differences (Schulte *et al.*, 2000; Fangue *et al.*, 2006). Moreover, the killifish is also capable to adapt to short-term exposures of dioxin-like toxic compounds (Nacci *et al.*, 1999; Cohen, 2002). The European flounder *Platichthys flesus* exhibited a strong sinal of selection, although no selective agents were identified (Hemmer-Hansen *et al.*, 2007). Patterns of divergence in nuclear genic markers of the three-spined sticklebacks *G. aculeatus* were correlated with the salinity gradient along the Baltic Sea (DeFaveri & Merilä, 2013). Adaptation to salinity gradients also promoted genetic structure in the European herring *Clupea harengus* associated to both nuclear (Lamichhaney *et al.*, 2012) and mtDNA markers (Bekkevold *et al.*, 2005; Gaggiotti *et al.*, 2009; Teacher *et al.*, 2012). Considering that the most of phylogeographic studies were based on single mtDNA markers and these were rarely tested for selection, phylogeographic patterns should be interpreted cautiously.

THE STUDIED SPECIES

ANCHOVIES

Anchovies assume a major role in most of the marine ecosystems around the globe, being extremely abundant and one of the main food sources for other pelagic fish, birds and marine mammals (Cury *et al.*, 2000). Anchovies (family Engraulidae, order Clupeiformes) are distributed worldwide, from temperate to tropical regions (Whitehead *et al.*, 1988). Most species are marine, although many tolerate low salinity during some stages of their life cycles, while few are exclusively adapted to freshwater environments in South America and Southeast Asia (Whitehead *et al.*, 1988). Most species school for protection and feed on plankton. This family comprises about 139 species distributed in 16 genera (Whitehead *et al.*, 1988). It is divided in two major clades, the subfamilies Coilinae and Engraulinae. The former is restricted to the Indo-Pacific and includes the genera *Coilia*, *Lycothrissa*, *Setipinna*, *Thryssa* and *Paupengraulis*. Engraulinae inhabits both the Indo-Pacific and Atlantic oceans and includes the genera *Stolephorus, Encrasicholina, Anchoa, Anchovia, Anchoviella, Cetengraulis, Jurengraulis, Lycengraulis, Pterengraulis, Amazonsprattus* and *Engraulis* (Figure 1.11; Whitehead *et al.*, 1988; Li & Ortí, 2007; Lavoué *et al.*, 2010; Bloom & Lovejoy, 2012).

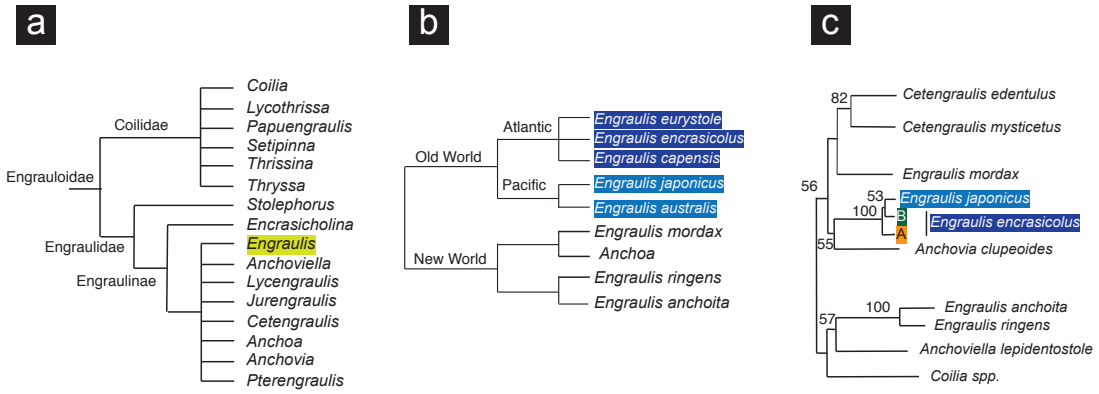


Figure 1.11 - Phylogeny of the Engraulidae family; a) phylogenetic hypothesis based on morphological characters with *Engraulis* genus highlighted in yellow shade; supra-specific names as in Grande and Nelson (1985); b) phylogenetic hypotheses for *Engraulis* species and closely related genera (Whitehead *et al.*, 1988); c) neighbour-joining bootstrap tree of partial mtDNA cytochrome *b* gene (Grant *et al.*, 2010); dark blue and light blue shading indicate Atlantic and Pacific species, respectively; *E. encrasicolus* mitochondrial phylogroups A and B are represented in orange and green.

The genus *Engraulis* is globally distributed and inhabits marine coastal waters, estuaries and lagoons of the major temperate and tropical sea basins, from 60ºN to 43ºS (Reid, 1967; Whitehead *et al.*, 1988). The genus is polyphyletic and comprises eight to nine described species, divided in two main groups: New World (NWA) and Old World (OWA) anchovies (Whitehead *et al.*, 1988; Borsa *et al.*, 2004). The NWA include the eastern Pacific Ocean species, the Californian anchovy *Engraulis mordax* Girard, 1856 and the Peruvian anchoveta *Engraulis ringens* Jenyns, 1842, and also the Argentine anchovy *Engraulis anchoita* Hubbs & Marini, 1935 in the southwestern Atlantic Ocean. The OWA comprehend the Japanese anchovy *Engraulis japonicus* Temminck & Schlegel, 1846 and the Australian anchovy *Engraulis australis* Shaw, 1790 in the western Pacific, and the Atlantic and Mediterranean European anchovy *Engraulis encrasicolus* Linnaeus, 1758, the Cape anchovy *Engraulis capensis* Gilchrist, 1913, and the silver anchovy *Engraulis eurystole* Swain & Meek, 1884 (Figure 1.11). Recently, a new species was described in the Mediterranean Sea, the white anchovy *Engraulis albidus* Borsa, Collet & Durand, 2004.

OLD WORLD ANCHOVIES

The OWA are coastal pelagic fish distributed throughout the Pacific and Atlantic oceans, including the Mediterranean, the Baltic and the Black seas (Figure 1.12). These fishes inhabit offshore areas above the continental platform, as well as inshore environments such as estuaries, inlets and bays, being tolerant to low salinities (Whitehead *et al.*, 1988). Also, they feed on phytoplankton and zooplankton, and school for protection from larger fishes, dolphins and birds. There are several conflicting aspects in the taxonomy of the OWA. Morphological differences among putative species are "slight and overlap" (Whitehead *et al.*, 1988). OWA are characterized by having a small to moderate size (up to 20 cm standard length), oval shape in cross-section and little compressed. Their back is blue/green; flanks are silver and lacks the lateral line. OWA have one short dorsal fin near the mid-point of the body, two pelvic and one anal fin, and lack the pre- and post-pelvic scutes, that are present in most other species of Engraulidae. The OWA complex is characterized by the snout projected beyond the tip of the lower jaw and the articulation of the lower jaw is well behind the eye and most species can be identified based on the 2nd supra maxilla shape and length, and by the number of gillrakers (Whitehead *et al.*, 1988).

Despite overlapping characteristics, taxa were identified as separate species according to their geographic distribution ranges, disregarding possible gene flow between individuals from those areas. Nonetheless, Whitehead *et al.* (1988) has postulated that "carefully revisionary work" should be performed, since this group could constitute one single species, supported by Grant *et al.* (2006) work that found shallow genetic divergences between taxa.

Evidences indicate that OWA have originated in the northwestern Pacific Ocean and colonised the Atlantic Ocean through the temperate south Indian Ocean, with stepping-stones in Australia and South Africa (Grant & Bowen, 2006). The paraphyletic relationship between the two clades of the European anchovy implies two dispersal events from the Pacific to the Atlantic Ocean, with about 1.5 Ma and 1.2 Ma of separation, respectively (Grant & Bowen, 2006). Within the Atlantic Ocean, the European anchovy *E. encrasicolus* (or the most recent common ancestor - MRCA) has split in the silver anchovy E. *eurystole*, the Cape anchovy *Engraulis capensis* and in the white anchovy *E. albidus* (Whitehead *et al.*, 1988; Borsa *et al.*, 2004). The European anchovy back dispersed to South Africa, experiencing several extinction-colonisation cycles driven by climatic shifts (Grant & Bowen, 2006). Within the Pacific Ocean, two phylogroups were found in the western Pacific species: one endemic to Australia and the other is present both in Australia and Japan, revealing recent colonizations from Japan to Australia (Grant & Bowen, 2006). Nevertheless, *E. japonicus* and *E. australis* diverged 105 - 420 ky ago (Liu *et al.*, 2006), probably resulting from trans-Equatorial dispersals during episodes of global cooling (Grant & Bowen, 2006).

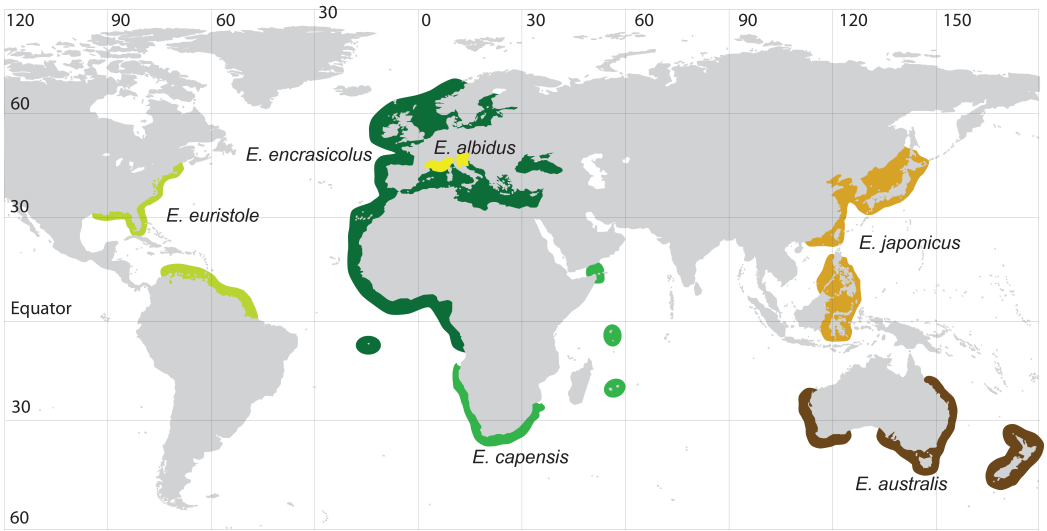


Figure 1.12 - Distribution range of Old World Anchovies.

In the next sections we provide detailed information on the biology, ecology and genetic variation of the European anchovy *Engraulis encrasicolus*, the type species of the family Engraulidae and the focus of this thesis. We also compiled the distribution range of each putative species, based on Whitehead *et al.* (1988) and Borsa *et al.* (2004) (see Annexe I).

THE MODEL SPECIES

EUROPEAN ANCHOVY *ENGRAULIS ENCRASICOLUS* (LINNAEUS, 1758)

BIOLOGY AND ECOLOGY OF THE EUROPEAN ANCHOVY

Figure 1.13 - European anchovy *Engraulis encrasicolus* (Linnaeus, 1758).

The European anchovy (Figure 1.13) is distributed from Norway to Angola, throughout the Mediterranean, Black and Azov Seas (Whitehead *et al.*, 1988) and the Baltic sea (Draganik & Wyszynski, 2004; see Figure 1.11). However, one shared haplotype was found between the Mediterranean Sea and South Africa (Grant *et al.*, 2005), indicating that *E. encrasicolus* distribution may not be reproductively isolated from *E. capensis*. The European anchovy inhabits coastal waters above the continental platform, from the surface down to 400m depth (Whitehead *et al.*, 1988), tolerating a wide range of temperatures (0^+ - 30 ºC) and salinities (5 - 41 psu). Therefore it can be found in estuaries, coastal lagoons and inlets, although is more abundant in upwelling zones. It schools for protection, for feeding, to migrate and to spawn (Blaxter & Hunter, 1982).

Although morphological differences among OWA are "slight and overlap" (Whitehead *et al.*, 1988), some variances within the European anchovy distribution range were previously reported. Based on morphology, Fage (1911, 1920) divided the European anchovy into two races, the Atlantic (subdivided into northern and southern groups) and the Mediterranean (subdivided into western and eastern groups). Within

the Black Sea, Alexandrov (1927) identified two subspecies, *E. encrasicolus ponticus* in western Black Sea and *E. encrasicolus maeoticus* in the Sea of Azov and eastern Black sea. Latter, Pousanov (1936) combined the Black Sea and the Mediterranean in a single species and considered the Azov anchovy *E. maeoticus* as a different taxa. In the 1990's, Bembo *et al.* (1996) described two different "stocks" in the Adriatic Sea based on size and morphology, where "black and big" anchovies were found in southern Adriatic, contrastingly to "silver and thin" anchovies in northern Adriatic.

Fish migrations are mainly associated to spawning and feeding behaviours. Small pelagic fish have high ability for dispersal, but are difficult to track by conventional methods, such as physical tags or satellite tracking, due to small dimensions and large population sizes. In the European anchovy, regional movements were depicted from artisanal fishing (Cunningham, 1895), trawling and acoustic surveys (Chashchin, 1996) and modelling (Mullon *et al.*, 2003). Anchovies may disperse passively through currents during egg and larvae stages, or actively during juvenile and adult stages. The passive dispersal occurs in a regional scale where eggs and larvae may be transported as far as currents allow (Mullon *et al.*, 2003; Parada *et al.*, 2003). Anchovies move seasonally between inshore areas to spawn during the summer and offshore areas during the winter to feed (Hutchings *et al.*, 1998). At the northern extreme of the distribution range, anchovies migrate southwards to less severe environments during the winter and expand to higher latitudes during spring (Alheit *et al.*, 2012). Genetic similarities between fish from South Africa and the Mediterranean (Grant & Bowen, 2006) reveal long distance dispersal, unlikely to have occurred in one single generation.

Anchovies have a short generation time, living mostly up to three years, however four to five years old fish were recorded in the Black Sea (Samsun *et al.*, 2004). Since anchovy populations have few age classes, recruitment and mortality are determinant in setting demographic structure (Borja *et al.*, 1998). Anchovies spawning season is extremely variable and depends on the geographical location. In the northern hemisphere anchovies spawn between late spring/early summer when temperatures range from 11.5 ºC - 16.5 ºC (Lluch-Belda *et al.*, 1991). In the southern hemisphere, spawning takes place during the austral summer, between December and March, when temperatures range from 16 ºC - 20 ºC (Richardson *et al.*, 1998; Twatwa *et al.*, 2005). Adult anchovies are highly plastic in reproductive tactics, having the ability to change spawning characteristics, depending on environmental conditions

(Millán, 1999). First maturation occurs after the first year (Parada *et al.*, 2003), with females maturing later than males (Sinovcic & Zorica, 2006). The European anchovy is a multiple spawner (Blaxter & Hunter, 1982; Melo, 1994), often leaving 10 batches of offspring within a season (Hunter & Leong, 1981), but no parental care is provided. It spawns at night after the diel migration above the thermocline (Palomera, 1991). Eggs are layed in the upper 50m of the water column, where float from one to four days until hatch, depending on water temperature (Regner, 1996). Spawning near the surface and living and feeding beneath the thermocline prevent cannibalism (Tudela & Palomera, 1997; Plounevez & Champalbert, 1999, 2000). The structured environment maintains the eggs in the upper layers with high levels of oxygen, vital to their development (Ekau & Verheye, 2005). When the environmental conditions are not ideal to spawn, oocytes are absorbed and no spawning occurs. Larvae development takes as much as 60 days from the egg phase to juvenile (Palomera *et al.*, 1988), mostly dependent on food availability (Palomera *et al.*, 2007).

After hatching, larvae may feed from the yolk sac for two days, preventing starvation (Durovic *et al.*, 2012). After this period, larvae need to catch their preys actively preventing "the point of no return" (Blaxter & Hempel, 1963). Larvae under 12 mm mainly feed on copepodites and nauplii (Tudela *et al.*, 2002). Adults feed more intensively during late winter/ early spring to store energy reserves for gonad maturation and spawning. These planktonic fish feed on zooplankton (copepods and crustacean larvae; Raab *et al.*, 2011) and phytoplankton (Plounevez & Champalbert, 1999). Their feeding habits are mainly diurnal below the thermocline, even though some crepuscular and nocturnal activities have been documented (Plounevez & Champalbert, 1999).

External causes of fish mortality mostly include predation, fishing pressures and sudden environmental changes that promote a declining of the populations with serious repercussions at the higher levels of the trophic chain (Pauly *et al.*, 1998). Top predators (Cury *et al.*, 2000) eat almost 55% of all small pelagic fish in the Benguela Current, as it is often observed during the migration of sardines and anchovies in South Africa. Anchovies have been one of the most exploited marine resources, mostly captured by purse seine or mid-water trawl fleets, heavily contributing to population fluctuations. Anchovies form large shoals of high densities, which make them an easy target and particularly vulnerable to overexploitation (Azzali *et al.*, 2002). In the last 25 years, an average of 500,000 t were captured annually (FAO,

2013), which contributed to the collapse of some regional stocks. In 1987, anchovy population crashed in the Adriatic Sea (Azzali *et al.*, 2002), recovering slowly since then. From 2005 to 2009, the European Union banned the anchovy fishery in the Bay of Biscay due to reduced adult biomass and low recruitment (European Comission, 2013).

Although predation and fisheries are two of the most important non-natural causes for mortality, unsuitable environmental conditions may have dramatic impact on populations' survival. Physical factors that are important to anchovy's growth and development include nutrient cycles, climate, temperature, enrichment processes (upwelling and mixing), concentration processes (convergence, frontal formation and water column stability) and processes favouring retention within appropriate habitat (Agostini & Bakun, 2002). However, the importance of these factors in each development stage is not fully understood. Sudden climate shifts during the earlier stages of the development (spawning and recruitment) can cause large population fluctuations and turnovers (Motos *et al.*, 1996; Borja *et al.*, 1998; Allain *et al.*, 2003). Consequently, the annual recruitment success determines the demographic structure of the populations, mostly composed by three or four cohorts.

GENETIC VARIATION OF THE EUROPEAN ANCHOVY *ENGRAULIS ENCRASICOLUS*

POPULATION STRUCTURE AND GENETIC DIFFERENTIATION

The European anchovy shows an unexpected degree of genetic differentiation, sharp discontinuities between sea basins and high levels of genetic diversity. Although differences were found between molecular markers, allozymes, mtDNA, microsatellites, SNP's and nuclear introns loosely show congruent genetic patterns among geographic regions. Despite different "stocks" were identified within the same geographic regions (e.g. Bay of Biscay, Adriatic Sea; Zarraonaindia *et al.*, 2009), nine main populations within restricted geographic regions have been identified: 1) Norway to the Bay of Biscay, 2) Galicia to South Portugal, 3) Gulf of Cadiz, Alboran Sea, southwards to Canaries and Senegal, 4) western Mediterranean, 5) northern Adriatic Sea, 6) southern Adriatic and Ionian seas, 7) Aegean sea, 8) Black sea and 9) South Africa. Among sea basins, both smooth clines and marked phylogeographic breaks between population boundaries were identified (Magoulas *et al.*, 2006; Zarraonaindia *et al.*, 2012). Nevertheless, previous studies on genetic variation were

mostly restricted to particular geographic areas and focused their scope on population structure either to identify fishing units or to uncover the historical demography of the species.

In the northeastern Atlantic, concordant patterns among different genetic markers showed that anchovies from the Bay of Biscay were genetically closer to the northwestern Mediterranean rather than to Atlantic adjacent areas, indicating an historical Mediterranean origin (Magoulas *et al.*, 2006; Sanz *et al.*, 2008; Zarraonaindia *et al.*, 2009; Borrell *et al.*, 2012; Zarraonaindia *et al.*, 2012; Viñas *et al.*, 2013). Thus, anchovies from the Bay of Biscay may have been the front wave of the northeastern Atlantic colonisation after the LGM (Zarraonaindia *et al.*, 2012). Populations northwards from the English Channel exhibit lower genetic diversities than southern populations, likely reflecting post-glacial colonisation (Zarraonaindia *et al.*, 2012). Nevertheless, shorter climate cycles such as the decadal North Atlantic Oscillation and the Atlantic multi-decadal Oscillation may have also influenced genetic diversity and range shifts of European anchovies (Rose, 2005; Alheit *et al.*, 2012; Petitgas *et al.*, 2012). Anchovies southwards Bay of Biscay to Senegal are genetically similar and thus, share the same genetic background (Magoulas *et al.*, 2006; Bouchenak-Khelladi *et al.*, 2008; Sanz *et al.*, 2008; Zarraonaindia *et al.*, 2012; Viñas *et al.*, 2013). Anchovies inhabiting the Alboran sea (western Mediterranean sea) are genetically closer to Atlantic than to Mediterranean populations (Magoulas *et al.*, 2006; Bouchenak-Khelladi *et al.*, 2008; Sanz *et al.*, 2008; Zarraonaindia *et al.*, 2012; Viñas *et al.*, 2013). Thus, the Almeria-Oran oceanfront may constitute a barrier to dispersal in the European anchovy (Magoulas *et al.*, 2006; Patarnello *et al.*, 2007).

Within the Mediterranean Sea, the complex geomorphology and oceanographic processes contributed to a strong genetic differentiation between geographic regions (Magoulas *et al.*, 2006; Bouchenak-Khelladi *et al.*, 2008; Sanz *et al.*, 2008; Borrell *et al.*, 2012; Zarraonaindia *et al.*, 2012; Viñas *et al.*, 2013). In the northern Mediterranean, significant differences among areas in allozyme, mtDNA and nuclear allele frequencies, body shape and otolith shape suggested several reproductively isolated populations (Kristoffersen & Magoulas, 2008). In the northwestern Mediterranean, no differences were found between locations (Sanz *et al.*, 2008). Although no stock boundaries were delimited due to seasonal variations, significant differences in allelic distribution were found between the Adriatic,

Tyrrhenian, Ionian and Aegean seas and the Sicilian strait (Bembo *et al.*, 1996). Spanakis *et al.* (1989) also recognized that Aegean and Ionian Sea populations were not panmictic suggesting restricted gene flow between these areas (Spanakis *et al.*, 1989). Within the Adriatic Sea, north and south anchovies showed different allelic composition, most probably corresponding to different morphotypes (Bembo *et al.*, 1996), and showed a degree of heterogeneity to adjacent waters (Bembo *et al.*, 1995). Anchovies from the Aegean Sea differed significantly from other locations (Bembo *et al.*, 1995). The restricted gene flow between the Aegean Sea and the Black Sea anchovies imposed by the Bosphorus strait, delimited different genetic populations (Grant, 2005; Ivanova & Dobrovolov, 2006; Magoulas *et al.*, 2006). Within the Black sea, including the Azov Sea, the two subspecies *E. encrasicolus maeoticus* and *E. encrasicolus ponticus* previously described based on morphology, were recognized based on immunological analyses, biochemical and genetic methods (Kalnin *et al.*, 1984; Kalnina & Kalnin, 1984; Kalnin & Kalnina, 1985).

Finally, anchovies from South Africa were found to recently derive from Mediterranean and northeastern Atlantic populations (Grant & Bowen, 2006). Despite mtDNA of South African anchovies is closer to northwestern Mediterranean fish, nuclear alleles were more related to anchovies from Canary Islands, Gulf of Cadiz and Alboran Sea (Bouchenak-Khelladi *et al.*, 2008). Moreover, typical alleles from *E. albidus* were found in the southern Africa sample, suggesting past introgression (Bouchenak-Khelladi *et al.*, 2008). Low mtDNA diversities in South African anchovies indicated that these populations might have experienced several extinction-recolonisation cycles (Grant & Bowen, 2006; Zarraonaindia *et al.*, 2012).

MTDNA CLADES A & B - VARIATION AND ORIGIN

The European anchovy mtDNA forms two groups (clades A and B) that vary abruptly in frequency between the eastern and western sea basins within the Mediterranean Sea (Figure 1.14; Magoulas *et al.*, 1996). These authors postulated that clade A was formed in isolation in the Black Sea during the Pliocene or the early Pleistocene and invaded the Mediterranean by a massive outflow after the LGM. Nevertheless, Grant (2005) after reanalysing Magoulas *et al.* (1996) data found that each clade had experienced different demographic histories and proposed that clade A was unlikely

to be in isolation in the Black Sea due to the harshness of climatic conditions. Instead, clade A must have invaded the Mediterranean from the Atlantic during the Weichselian Ice age (50-25ky), while clade B has a more complex genealogy. This clade must have invaded the Mediterranean earlier during the Eemian, the last interglacial period (450-350ky).

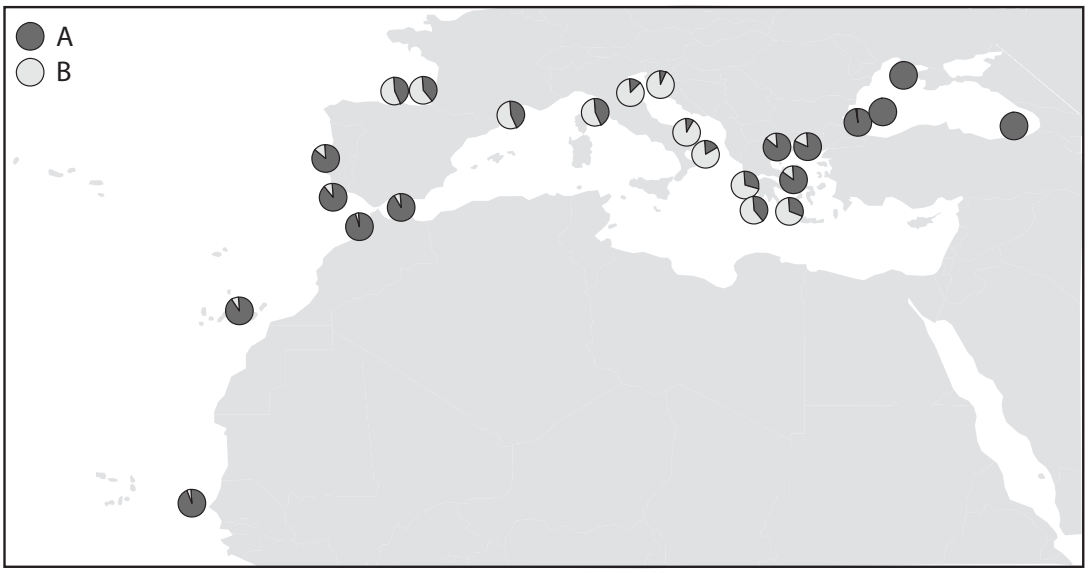


Figure 1.14 - mtDNA clade frequencies presented in Magoulas *et al.* (2006); dark grey - clade A; light grey - clade B).

Low diversities in the Black Sea indicated that this sea basin was colonised by a founder effect (Grant, 2005). Although divergence in isolation followed by secondary contact has been the most consensual explanation for the origin of the mtDNA clades, other hypothesis (including sex-biased dispersal, nuclear allelic convergence, incomplete mtDNA lineage sorting, adaptive introgression, demographic disparities, gamete incompatibility or adaptive selection) have not been excluded (Grant, 2005; Magoulas *et al.*, 2006; Kristoffersen & Magoulas, 2008). Clade A occurs at higher frequencies in the Black and Aegean seas and in the eastern Atlantic, while clade B occurs at higher frequencies in the Adriatic Sea and northwards to the English Channel (Magoulas *et al.*, 2006; Zarraonaindia *et al.*, 2012). Anchovies from the northwestern Mediterranean, Ligurian and Ionian seas, the Patraikos Gulf in the southwestern Aegean sea and the Bay of Biscay exhibited intermediate frequencies from both clades (Magoulas *et al.*, 2006; Zarraonaindia *et al.*, 2012). Magoulas *et al.*

(2006) confirmed that the Mediterranean was most probably seeded from an Atlantic refugium and the Black Sea colonisation was the product of a founder effect, while clade B reflected a more complex genealogy due to long-lasting presence in the Mediterranean. Deep divergences between clades reinforced the evolution in isolation, followed by a secondary contact, while shallow divergences within clades were attributed to the complex Mediterranean Sea geography and oceanographic processes.

Intron length polymorphism was analysed to evaluate the reproductive isolation between the two mtDNA clades, but no differences were found, indicating that the two lineages interbreed (Kristoffersen & Magoulas, 2008). Nevertheless, to correctly estimate historical and contemporary patterns of genetic differentiation, population structure, demography and gene flow, it is essential to include sampling locations from most sea basins and geographic regions with particular oceanographic features. The lack of samples from southern and eastern Mediterranean, as well as Central Africa, precluded previous authors to support some of their inferences (e.g. glacial refugia; Grant, 2005; Magoulas *et al.*, 2006).

OBJECTIVES

The major objective of this thesis is to investigate the historical and present processes that shape the genetic patterns of a small pelagic fish, the European anchovy *Engraulis encrasicolus*. To this purpose, both nuclear and mitochondrial molecular markers were used throughout its distribution range and at more regional spatial scales to 1) examine the genetic patterns of population structure, genetic diversity, genetic differentiation and historical demography of leading-edge populations; 2) test the potential effect of positive selection in the mitochondrial DNA; 3) uncover the phylogenetic relationships among the OWA group and to propose a biogeographic scenario of the evolution of the species complex.

REFERENCES

Agostini, V.N. & Bakun, A. (2002) "Ocean triads" in the Mediterranean Sea: physical mechanisms potentially structuring reproductive habitat suitability (example application to European anchovy, *Engraulis encrasicolus*). *Fisheries Oceanography*, **11**, 129–142.

Alexandrov, A.I. (1927) Anchovy from Azov- Black Sea basin and it origin and taxonomy. *Proc. Kerch Fish. Stn.*, 1/2 and 3:43.

Alheit, J., Pohlmann, T., Casini, M., Greve, W., Hinrichs, R., Mathis, M., O'Driscoll, K., Vorberg, R. & Wagner, C. (2012) Climate variability drives anchovies and sardines into the North and Baltic Seas. *Progress In Oceanography*, **96**, 128—139.

Allain, G., Petitgas, P., Grellier, P. & Lazure, P. (2003) The selection process from larval to juvenile stages of anchovy (*Engraulis encrasicolus*) in the Bay of Biscay investigated by Lagrangian simulations and comparative otolith growth. *Fisheries Oceanography*, **12**, 407—418.

Almada, V.C., Pereira, A.M., Robalo, J.I., Fonseca, J.P., Levy, A., Maia, C. & Valente, A. (2008) Mitochondrial DNA fails to reveal genetic structure in sea-lampreys along European shores. *Molecular Phylogenetics and Evolution*, **46**, 391—396.

Andre, C., Larsson, L.C., Laikre, L., Bekkevold, D., Brigham, J., Carvalho, G.R., Dahlgren, T.G., Hutchinson, W.F., Mariani, S., Mudde, K., Ruzzante, D.E. & Ryman, N. (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.

Avise, J.C. (1992) Molecular population structure and the biogeographic history of a regional fauna: a case history with lessons for conservation biology. *Oikos*, **62**, 62—76.

Avise, J.C. (2000) *Phylogeography: the history and formation of species*. Harvard University Press, Cambridge, Massachussets, USA.

Avise, J.C., Arnold, J., Ball, R.M., Bermingham, E., Lamb, T., Neigel, J.E., Reeb, C.A. & Saunders, N. (1987) Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annual Reviews in Ecology and Systematics*, **18**, 489—522.

Azzali, M., De Felice, A., Luna, M., Cosimi, G. & Parmiggiani, F. (2002) The state of the Adriatic Sea centered on the small pelagic fish populations. *Marine Ecology*, **23**, 78—91.

Ballard, J.W.O. & Whitlock, M.C. (2004) The incomplete natural history of mitochondria. *Molecular Ecology*, **13**, 729–744.

Beaumont, M. & Rannala, B. (2004) The Bayesian revolution in genetics. *Nature Reviews Genetics*, **5**, 251—261.

Bekkevold, D., Andre, C., Dahlgren, T.G., Clausen, L.A.W., Torstensen, E., Mosegaard, H., Carvalho, G.R., Christensen, T.B., Norlinder, E. & Ruzzante, D.E. (2005) Environmental correlates of population differentiation in Atlantic herring. *Evolution*, **59**, 2656—2668.

Bembo, D., Carvalho, G.R., Cingolani, N. & Pitcher, T. (1996) Electrophoretic analysis of stock structure in Northern Mediterranean anchovies, *Engraulis encrasicolus*. *ICES Journal of Marine Science*, **53**, 115—128.

Bembo, D.G., Carvalho, G.R., Snow, M., Cingolani, N. & Pitcher, T. (1995) Stock discrimination among European anchovies *Engraulis encrasicolus*, by means of PCR-amplified mitochondrial DNA analysis. *Fisheries Bulletin*, **94**, 31–40.

Berglund, B.E., Sandgren, P., Barnekow, L., Hannon, G., Jiang, H., Skog, G. & Yu, S.-Y. (2005) Early Holocene history of the Baltic Sea, as reflected in coastal sediments in Blekinge, southeastern Sweden. *Quaternary International*, **130**, 111—139.

Bigg, G.R., Cunningham, C.W., Ottersen, G., Pogson, G.H., Wadley, M.R. & Williamson, P. (2008) Ice-age survival of Atlantic cod: agreement between palaeoecology models and genetics. *Proceedings of the Royal Society B - Biological Sciences*, **275**, 163–U13.

Blaxter, J.H.S. & Hempel, G. (1963) The influence of egg size on herring larvae (*Clupea harengus* L.). *ICES Journal of Marine Science*, **28**, 211—240.

Blaxter, J.H.S. & Hunter, J.R. (1982) The biology of the clupeoid fishes. *Advances in Marine Biology*, **20**, 3—223.

Bloom, D.D. & Lovejoy, N.R. (2012) Molecular phylogenetics reveals a pattern of biome conservatism in New World anchovies (family Engraulidae). *Journal of Evolutionary Biology*, **25**, 701—715.

Bohonak, A.J. (1999) Dispersal, gene flow, and population structure. *Quarterly Review of Biology*, **74**, 21–45.

Borja, A., Uriarte, A., Egaña, J., Motos, L. & Valencia, V. (1998) Relationships between anchovy (*Engraulis encrasicolus*) recruitment and environment in the Bay of Biscay (1967-1996) *Fisheries Oceanography*, **7**, 375—380.

Borrell, Y.J., Piñera, J.A., Sánchez Prado, J.A. & Blanco, G. (2012) Mitochondrial DNA and microsatellite genetic differentiation in the European anchovy *Engraulis encrasicolus* L. *ICES Journal of Marine Science*, **69**, 1357–1371.

Borsa, P., Collet, A. & Durand, J.D. (2004) Nuclear-DNA markers confirm the presence of two anchovy species in the Mediterranean. *Comptes Rendus Biologies*, **327**, 1113–1123.

Bouchenak-Khelladi, Y., Durand, J.D., Magoulas, A. & Borsa, P. (2008) Geographic structure of European anchovy: a nuclear-DNA study. *Journal of Sea Research*, **59**, 269–278.

Bowen, B.W., Bass, A.L., Rocha, L.A., Grant, W.S. & Robertson, D.R. (2001) Phylogeography of the trumpetfishes (*Aulostomus*): ring species complex on a global scale. *Evolution*, **55**, 1029–1039.

Bradbury, I.R., Hubert, S., Higgins, B., Borza, T., Bowman, S., Paterson, I.G., Snelgrove, P.V.R., Morris, C.J., Gregory, R.S., Hardie, D.C., Hutchings, J.A., Ruzzante, D.E., Taggart, C.T. & Bentzen, P. (2010) Parallel adaptive evolution of Atlantic cod on both sides of the Atlantic Ocean in response to temperature. *Proceedings of the Royal Society B*, **277**, 3725—3734.

Briggs, J.C. (1974) *Marine Zoogeography*. McGraw-Hill New York.

Briggs, J.C. (1995) *Global Biogeography*. Elsevier Science, Amsterdam, Netherlands.

Brown, J.H. & Lomolino, M.V. (1998) *Biogeography*, 2nd edn. Sinauer Associates, Inc. Publishers, Sunderland, MA.

Carlin, J.L., Robertson, D.R. & Bowen, B.W. (2003) Ancient divergences and recent connections in two tropical Atlantic reef fishes *Epinephelus adscensionis* and *Rypticus saponaceous* (Percoidei: Serranidae). *Marine Biology*, **143**, 1057–1069.

Chashchin, A.K. (1996) The Black Sea populations of anchovy. *Scientia Marina*, **60**, 219—225.

Chevolot, M., Hoarau, G., Rijnsdorp, A.D., Stam, W.T. & Olsen, J.L. (2006) Phylogeography and population structure of thornback rays (*Raja clavata* L., Rajidae). *Molecular Ecology*, **15**, 3693—3705.

CLIMAP Project Members (1984) The last interglacial ocean. *Quaternary Research*, **21**, 123—224.

Cohen, S. (2002) Strong positive selection and habitat-specific amino acid substitution patterns in MHC from an estuarine fish under intense pollution stress. *Mol Biol Evol*, **19**, 1870—1880.

Coyne, J.A. & Orr, H.A. (2004) *Speciation*. Sinauer Associates, Incorporated Publishers, Sunderland, MA, USA.

Crowley, T.J. (2000) Causes of climate change over the past 1000 years. *Science*, **289**, 270—277.

Cunningham, J.T. (1895) The migration of the anchovy. *Journal of the Marine Biological Association of the United Kingdom*, **3**, 300-303.

Cury, P., Bakun, A., Crawford, R.J.M., Jarre, A., Quinones, R.A., Shannon, L.J. & Verheye, H.M. (2000) Small pelagics in upwelling systems: patterns of interaction and structural changes in "wasp-waist" ecosystems. *Ices Journal of Marine Science*, **57**, 603—618.

Debes, P.V., Zachos, F.E. & Hanel, R. (2008) Mitochondrial phylogeography of the European sprat (*Sprattus sprattus* L., Clupeidae) reveals isolated climatically vulnerable populations in the Mediterranean Sea and range expansion in the northeast Atlantic. *Molecular Ecology*, **17**, 3873–3888.

DeFaveri, J. & Merilä, J. (2013) Evidence for adaptive phenotypic differentiation in Baltic Sea sticklebacks. *Journal of Evolutionary Biology*, **26**, 1700—1715.

Doiron, S., Bernatchez, L. & Blier, P.U. (2002) A comparative mitogenomic analysis of the potential adaptive value of Arctic charr mtDNA introgression in brook charr populations (*Salvelinus fontinalis* Mitchill). *Mol Biol Evol*, **19**, 1902—1909.

Doney, S.C., Ruckelshaus, M., Emmett Duffy, J., Barry, J.P., Chan, F., English, C.A., Galindo, H.M., Grebmeier, J.M., Hollowed, A.B., Knowlton, N., Polovina, J., Rabalais, N.N., Sydeman, W.J. & Talley, L.D. (2012) Climate change impacts on marine ecosystems. *Annual Review of Marine Science*, **4**, 11—37.

Dowling, D.K., Friberg, U. & Lindell, J. (2008) Evolutionary implications of non-neutral mitochondrial genetic variation. *Trends in Ecology & Evolution*, **23**, 546—554.

Draganik, B. & Wyszynski, M. (2004) The European anchovy (*Engraulis encrasicolus* [L.]) in the Baltic Sea. *Bulletin of the Sea Fisheries Institute*, **162**, 53-58.

Durovic, M., Pesic, A., Regner, S., Joksimovic, A., Mandic, M., Kasalica, O., Ikica, Z. & Krpo-Cetkovic, J. (2012) Daily otolith increments and growth rate of juvenile anchovy, *Engraulis encrasicolus* (Linnaeus, 1758), in the southeastern Adriatic Sea. *Acta Adriatica*, **53**, 331—340.

Ekau, W. & Verheye, H.M. (2005) Influence of oceanographic fronts and low oxygen on the distribution of ichthyoplankton in the Benguela and southern Angola currents. *African Journal of Marine Science*, **27**, 629—639.

Engelhaupt, D., Rus Hoelzel, A., Nicholson, C., Frantzis, A., Mesnick, S., Gero, S., Whitehead, H.A.L., Rendell, L., Miller, P., De Stefanis, R., Cañadas, A.N.A., Airoldi, S. & Mignucci-Giannoni, A.A. (2009) Female philopatry in coastal basins and male dispersion across the North Atlantic in a highly mobile

marine species, the sperm whale (*Physeter macrocephalus*). *Molecular Ecology*, **18**, 4193—4205.

European Comission (2013) *Commission proposes catch possibilities for anchovy in the Bay of Biscay for 2013/14* Available at: http://ec.europa.eu/information_society/newsroom/cf/mare/itemdetail.cfm?item_id=11347&lang=pt (accessed 01-12-2013)

Fage, L. (1911) Recherches sur la biologie de l'anchois *Engraulis encrasicolus* (L.) *Ann. Inst. Oceanography Monaco*, **2**, 1—40.

Fage, L. (1920) Danish oceanographic expedition 1908-1910 to the Mediterranean and adjacent seas. In: (ed. J. Schmidt), p. 137, Copenhagen, Denmark.

Fangue, N.A., Hofmeister, M. & Schulte, P.M. (2006) Intraspecific variation in thermal tolerance and heat shock protein gene expression in common killifish, *Fundulus heteroclitus*. *J Exp Biol*, **209**, 2859—2872.

FAO (2013) *Fact sheet Engraulis encrasicolus L.* Available at: http://www.fao.org/fishery/species/2106 (accessed 01-12-2013)

Floeter, S.R., Rocha, L.A., Robertson, D.R., Joyeux, J.C., Smith-Vaniz, W.F., Wirtz, P., Edwards, A.J., Barreiros, J.P., Ferreira, C.E.L., Gasparini, J.L., Brito, A., Falcón, J.M., Bowen, B.W. & Bernardi, G. (2008) Atlantic reef fish biogeography and evolution. *Journal of Biogeography*, **35**, 22—47.

Gaggiotti, O.E., Bekkevold, D., Jorgensen, H.B.H., Foll, M., Carvalho, G.R., Andre, C. & Ruzzante, D.E. (2009) Disentangling the effects of evolutionary, demographic, and environmental factors influencing genetic structure of natural populations: Atlantic herring as a case study. *Evolution*, **63**, 2939–2951.

Galarza, J.A., Carreras-Carbonell, J., Macpherson, E., Pascual, M., Roques, S., Turner, G.F. & Rico, C. (2009) The influence of oceanographic fronts and early-life-history traits on connectivity among littoral fish species. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 1473—1478.

Galtier, N., Nabholz, B., Giémin, S. & Hurst, G.D.D. (2009) Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Molecular Ecology*, **18**, 4541–4550.

Grande, L. & Nelson, G. (1985) Interrelationships of fossil and recent anchovies (Teleostei: Engrauloidea) and description of a new species from the Miocene of Cyprus. *American Museum Novitates*, 1—16.

Grant, W.S. (2005) A second look at mitochondrial DNA variability in European anchovy (*Engraulis encrasicolus*): assessing models of population structure and the Black Sea isolation hypothesis. *Genetica*, **125**, 293–309.

Grant, W.S. & Bowen, B.W. (2006) Living in a tilted world: climate change and geography limit speciation in Old World anchovies (*Engraulis*; Engraulidae). *Biological Journal of the Linnean Society*, **88**, 673–689.

Grant, W.S., Leslie, R.W. & Bowen, B.W. (2005) Molecular genetic assessment of bipolarity in the anchovy genus *Engraulis*. *Journal of Fish Biology*, **67**, 1242–1265.

Grant, W.S., Lecomte, F. & Bowen, B.W. (2010) Biogeographical contingency and the evolution of tropical anchovies (genus *Cetengraulis*) from temperate anchovies (genus *Engraulis*). *Journal of Biogeography*, **37**, 1352–1362.

Grant, W.S., Liu, M., Gao, T. & Yanagimoto, T. (2012) Limits of Bayesian skyline plot analysis of mtDNA sequences to infer historical demographies in Pacific

herring (and other species). *Molecular Phylogenetics and Evolution*, **65**, 203–212.

Gysels, E.S., Hellemans, B., Pampoulie, C. & Volckaert, F.A.M. (2004) Phylogeography of the common goby, *Pomatoschistus microps*, with particular emphasis on the colonization of the Mediterranean and the North Sea. *Molecular Ecology*, **13**, 403—417.

Hays, J.D., Imbrie, J. & Shackleton, N.J. (1976) Variations in the Earth's orbit: pacemaker of the ice ages. *Science*, **194**, 1121—1132.

Hedrick, P.W. (2005) *Genetics of populations*, 3rd edition edn. Jones and Bartlett Publishers, Sudbury, MA.

Hemmer-Hansen, J., Nielsen, E.E., Frydenberg, J. & Loeschcke, V. (2007) Adaptive divergence in a high gene flow environment: Hsc70 variation in the European flounder (*Platichthys flesus* L.). *Heredity*, **99**, 592—600.

Hewitt, G.M. (1996) Some genetic consequencies of ice ages, and their role in divergence and speciation. *Biological Journal of Linnean Society*, **58**, 247—276.

Hewitt, G.M. (1999) Molecular genetics re-colonization of European biota. *Biological Journal of the Linnean Society*, **68**, 87–112.

Hewitt, G.M. (2000) The genetic legacy of the Quaternary ages. *Nature*, **405**, 907–913.

Hewitt, G.M. (2004) Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, **359**, 183–195.

Hickerson, M.J. & Cunningham, C.W. (2006) Nearshore fish (*Pholis gunnellus*) persists across the North Atlantic through multiple glacial episodes. *Molecular Ecology*, **15**, 4095–4107.

Huddart, D. & Stott, T. (2010) *Earth environments: past, present and future*. John Wiley & Sons, Ltd, Oxford.

Hunter, J.R. & Leong, R. (1981) The spawning energetics of female northern anchovy, *Engraulis mordax*. *Fishery Bulletin*, **79**, 215—230.

Hutchings, L., Barange, M., Bloomer, S.F., Boyd, A.J., Crawford, R.J.M., Huggett, J.A., Kerstan, M., Korrubel, J.L., De Oliveira, J.A.A., Painting, S.J., Richardson, A.J., Shannon, L.J., Schulein, F.H., Van der Lingen, C.D. & Verheye, H.M. (1998) Multiple factors affecting South African anchovy recruitment in the spawning, transport and nursery areas. *South African Journal of Marine Science*, **19**, 211—225.

Ivanova, P. & Dobrovolov, I. (2006) Population genetic structure on European anchovy (*Engraulis encrasicolus*, Linnaeus, 1758) (Osteichthyes: Engraulidae) from Mediterranean basin and Atlantic Ocean. *Acta Adriatica*, **47**, 13—22.

Janecek, T.R. & Rea, D.K. (1985) Quaternary fluctuations in the northern hemisphere trade winds and westerlies. *Quaternary Research*, **24**, 150—163.

Jolly, M.T., Jollivet, D., Gentil, F., Thiebaut, E. & Viard, F. (2005) Sharp genetic break between Atlantic and English Channel populations of the polychaete *Pectinaria koreni*, along the north coast of France. *Heredity*, **94**, 23—32.

Joyeux, J.C., Floeter, S.R., Ferreira, C.E.L. & Gasparini, J.L. (2001) Biogeography of tropical reef fishes: the South Atlantic puzzle. *Journal of Biogeography*, **28**, 831–841.

Joyeux, J.C., Floeter, S.R., Ferreira, C.E.L. & Gasparini, J.L. (2003) Solving the South Atlantic puzzle requires more data, not more speculation. *Journal of Biogeography*, **30**, 1461–1463.

Kalnin, V.V. & Kalnina, O.V. (1985) Introgressive race hybridization and population structure of the Black Sea anchovy. In: *Genetic differentiation and reproductive relationship between the Black Sea race and Azov Sea race of European anchovy* (ed. Genetics), p. 1352—1360

Kalnin, V.V., Kalnina, O.V. & Dashova, M.B. (1984) Biochemical polymorphism. In: *Genetic differentiation and reproductive relationship between the Black Sea race and Azov Sea race of European anchovy* (ed. Genetics), p. 303—308

Kalnina, O.V. & Kalnin, V.V. (1984) Genetic differences between the Black Sea race and the Azov Sea race and their intrarace heterogeny. In: *Genetic differentiation and reproductive relationship between the Black Sea race and Azov Sea race of European anchovy* (ed. Genetics), p. 309—313

Kark, S. & van Rensburg, B.J. (2006) Ecotones: marginal or central areas of transition. *Israel Journal of Ecology & Evolution*, **52**, 29—53.

Kettle, A.J., Morales-Muñiz, A., Roselló-Izquierdo, E., Heinrich, D. & Vøllestad, L.A. (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.

Knowles, L.L. & Maddison, W.P. (2002) Statistical Phylogeography. *Molecular Ecology*, **11**, 2623—2635.

Kraytsberg, Y., Schwartz, M., Brown, T.A., Ebralidse, K., Kunz, W.S., Clayton, D.A., Vissing, J. & Khrapko, K. (2004) Recombination of human mitochondrial DNA. *Science*, **304**, 981.

Kristoffersen, J.B. & Magoulas, A. (2008) Population structure of anchovy *Engraulis encrasicolus* L. in the Mediterranean Sea inferred from multiple methods. *Fisheries Research*, **91**, 187–195.

Lambeck, K., Esat, T.M. & Potter, E.K. (2002) Links between climate and sea levels for the past three million years. *Nature* **419**, 199–206.

Lamichhaney, S., Barrio, A.M., Rafati, N., Sundström, G., Rubin, C.-J., Gilbert, E.R., Berglund, J., Wetterbom, A., Laikre, L., Webster, M.T., Grabherr, M., Ryman, N. & Andersson, L. (2012) Population-scale sequencing reveals genetic differentiation due to local adaptation in Atlantic herring. *Proceedings of the National Academy of Sciences*, **109**, 19345—19350.

Lande, R. & Shannon, S. (1996) The role of genetic variation in adaptation and population persistence in a changing environment. *Evolution*, **50**, 434—437.

Larmuseau, M.H.D., Van Houdt, J.K.J., Guelinckx, J., Hellemans, B. & Volckaert, F.A.M. (2009) Distributional and demographic consequences of Pleistocene climate fluctuations for a marine demersal fish in the north-eastern Atlantic. *Journal of Biogeography*, **36**, 1138—1151.

Lavoué, S., Miya, M. & Nishida, M. (2010) Mitochondrial phylogenomics of anchovies (family Engraulidae) and recurrent origins of pronounced miniaturization in the order Clupeiformes. *Molecular Phylogenetics and Evolution*, **56**, 480—485.

Leigh, E.G., O'Dea, A. & Vermeij, G.J. (2013) Historical biogeography of the Isthmus of Panama. *Biological Reviews*,

Lessios, H.A., Kessing, B.D., Robertson, D.R. & Paulay, G. (1999) Phylogeography of the pantropical sea urchin *Eucidaris* in relation to land barriers and ocean currents. *Evolution*, **53**, 806—817.

Li, C. & Ortí, G. (2007) Molecular phylogeny of Clupeiformes (Actinopterygii) inferred from nuclear and mitochondrial DNA sequences. *Molecular Phylogenetics and Evolution*, **44**, 386—398.

Limborg, M., Hanel, R., Debes, P., Ring, A., André, C., Tsigenopoulos, C. & Bekkevold, D. (2012) Imprints from genetic drift and mutation imply relative divergence times across marine transition zones in a pan-European small pelagic fish (*Sprattus sprattus*). *Heredity*, **109**, 96–107.

Lisiecki, L.E. & Raymo, M.E. (2005) A Pliocene-Pleistocene stack of 57 globally distributed benthic δ18O records. *Paleoceanography*, **20**, PA1003.

Liu, J.X., Gao, T.X., Zhuang, Z.M., Jin, X.S., Yokogawa, K. & Zhang, Y.P. (2006) Late Pleistocene divergence and subsequent population expansion of two closely related fish species, Japanese anchovy (*Engraulis japonicus*) and Australian anchovy (*Engraulis australis*). *Molecular Phylogenetics and Evolution*, **40**, 712—723.

Livingston, B.E. (1903) The distribution of the upland plant societies of Kent County, Michigan. *Botanical Gazette*, **35**, 36—55.

Lluch-Belda, D., Lluch-Cota, D., Hernandez-Vazquez, S., Salinas-Zavala, C. & Schwartzlose, R.A. (1991) Sardine and anchovy spawning as related to temperature and upwelling in the California current system. In: *CalCOFI*

Luiz-Júnior, O.J., Floeter, S.R., Gasparini, J.L., Ferreira, C.E.L. & Wirtz, P. (2004) The occurrence of *Acanthurus monroviae* (Perciformes: Acanthuridae) in the south-western Atlantic, with comments on other eastern Atlantic reef fishes occurring in Brazil. *Journal of Fish Biology*, **65**, 1173–1179.

Maggs, C.A., Castilho, R., Foltz, D., Henzler, C., Jolly, M.T., Kelly, J., Olsen, J., Perez, K.E., Stam, W., Vainola, R., Viard, F. & Wares, J. (2008) Evaluating signatures of glacial refugia for North Atlantic benthic marine taxa. *Ecology*, **89**, S108–S122.

Magoulas, A., Tsimenides, N. & Zouros, E. (1996) Mitochondrial DNA phylogeny and the reconstruction of the population history of a species: the case of the European anchovy (*Engraulis encrasicolus*). *Molecular Biology and Evolution*, **13**, 178–190.

Magoulas, A., Castilho, R., Caetano, S., Marcato, S. & Patarnello, T. (2006) Mitochondrial DNA reveals a mosaic pattern of phylogeographical structure in Atlantic and Mediterranean populations of anchovy (*Engraulis encrasicolus*). *Molecular Phylogenetics and Evolution*, **3**, 734–746.

Makinen, H.S. & Merila, J. (2008) Mitochondrial DNA phylogeography of the three-spined stickleback (*Gasterosteus aculeatus*) in Europe - evidence for multiple glacial refugia. *Molecular Phylogenetics and Evolution*, **46**, 167—182.

Mariano, A.J. (2013) *Ocean surface currents*. Available at: http://oceancurrents.rsmas.miami.edu (accessed 29th of November 2013).

McInerny, G.J., Turner, J.R.G., Wong, H.Y., Travis, J.M.J. & Benton, T.G. (2009) How range shifts induced by climate change affect neutral evolution. *Proceedings of the Royal Society B: Biological Sciences*, **276**, 1527–1534.

Melo, Y.C. (1994) Spawning frequency of the anchovy *Engraulis capensis*. *South African Journal of Marine Science*, **14**, 321—331.

Millán, M. (1999) Reproductive characteristics and condition status of anchovy *Engraulis encrasicolus* L. from the Bay of Cadiz (SW Spain). *Fisheries Research*, **41**, 73—86.

Moritz, C. (1994) Applications of mitochondrial DNA analysis in conservation: a critical review. *Molecular Ecology*, **3**, 401—411.

Motos, L., Uriarte, A. & Valencia, V. (1996) The spawning environment of the Bay of Biscay anchovy (*Engraulis encrasicolus* L). *Scientia Marina*, **60**, 117—140.

Mullon, C., Freon, P., Parada, C., Van Der Lingen, C. & Huggett, J. (2003) From particles to individuals: modelling the early stages of anchovy (*Engraulis capensis/ encrasicolus*) in the southern Benguela. *Fisheries Oceanography*, **12**, 396—406.

Nacci, D., Coiro, L., Champlin, D., Jayaraman, S., McKinney, R., Gleason, T.R., Munns Jr, W.R., Specker, J.L. & Cooper, K.R. (1999) Adaptations of wild populations of the estuarine fish *Fundulus heteroclitus* to persistent environmental contaminants. *Marine Biology*, **134**, 9—17.

Olsen, J.L., Stam, W.T., Coyer, J.A., Reusch, T.B.H., Billingham, M., Boström, C., Calvert, E., Christie, H., Granger, S., La Lumière, R., Milchakova, N., Oudot-Le Secq, M.P., Procaccini, G., Sanjabi, B., Serrão, E., Veldsink, J., Widdicombe, S. & Wyllie-Echeverria, S. (2004) North Atlantic phylogeography and large-scale population differentiation of the seagrass *Zostera marina* L. *Molecular Ecology*, **13**, 1923—1941.

Otto-Bliesner, B.L., Brady, E.C., Clauzet, G., Tomas, R., Levis, S. & Kothavala, Z. (2006) Last Glacial Maximum and Holocene Climate in CCSM3. *Journal of Climate*, **19**, 2526—2544.

Palomera, I. (1991) Vertical distribution of eggs and larvae of *Engraulis encrasicolus* in stratified waters of the western Mediterranean. *Marine Biology*, **111**, 37—44.

Palomera, I., Moralesnin, B. & Lleonart, J. (1988) Larval growth of anchovy, *Engraulis encrasicolus*, in the western Mediterranean Sea. *Marine Biology*, **99**, 283–291.

Palomera, I., Olivar, M.P., Salat, J., Sabates, A., Coll, M., Garcia, A. & Morales-Nin, B. (2007) Small pelagic fish in the NW Mediterranean Sea: an ecological review. *Progress in Oceanography*, **74**, 377—396.

Parada, C., van der Lingen, C.D., Mullon, C. & Penven, P. (2003) Modelling the effect of buoyancy on the transport of anchovy (*Engraulis capensis*) eggs from spawning to nursery grounds in the southern Benguela: an IBM approach. *Fisheries Oceanography*, **12**, 170–184.

Patarnello, T., Volckaert, F.A.M.J. & Castilho, R. (2007) Pillars of Hercules: is the Atlantic-Mediterranean transition a phylogeographical break? *Molecular Ecology*, **16**, 4426–4444.

Pauly, D., Christensen, V., Dalsgaard, J., Froese, R. & Torres, F. (1998) Fishing down marine food webs. *Science*, **279**, 860—863.

Petit, J.R., Jouzel, J., Raynaud, D., Barkov, N.I., Barnola, J.M., Basile, I., Bender, M., Chappellaz, J., Davis, M., Delaygue, G., Delmotte, M., Kotlyakov, V.M., Legrand, M., Lipenkov, V.Y., Lorius, C., Pepin, L., Ritz, C., Saltzman, E. & Stievenard, M. (1999) Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica. *Nature*, **399**, 429—436.

Petitgas, P., Alheit, J., Peck, M., Raab, K., Irigoien, X., Huret, M., van der Kooij, J., Pohlmann, T., Wagner, C., Zarraonaindia, I. & Dickey-Collas, M. (2012) Anchovy population expansion in the North Sea. *Marine Ecology Progress Series*, **444**, 1–13.

Philander, S.G.H. (1986) Unusual conditions in the tropical Atlantic Ocean in 1984. *Nature*, **322**, 236—238.

Plounevez, S. & Champalbert, G. (1999) Feeding behaviour and trophic environment of *Engraulis encrasicolus* (L.) in the Bay of Biscay. *Estuarine Coastal and Shelf Science*, **49**, 177—191.

Plounevez, S. & Champalbert, G. (2000) Diet, feeding behaviour and trophic activity of the anchovy (*Engraulis encrasicolus* L.) in the Gulf of Lion (Mediterranean Sea). *Oceanologica Acta*, **23**, 175—192.

Pousanov, I.I.P., I.I. 1936. , (1936) Anchovy scientific-commercial monography. In: *Notes of Gorkij Univ.*, p. 101, Gorkij.

Provan, J. & Bennett, K.D. (2008) Phylogeographic insights into cryptic glacial refugia. *Trends in Ecology & Evolution*, **23**, 564—571.

Raab, K., Nagelkerke, L.A.J., Boerée, C., Rijnsdorp, A.D., Temming, A. & Dickey-Collas, M. (2011) Anchovy *Engraulis encrasicolus* diet in the North and Baltic Seas. *Journal of Sea Research*, **65**, 131—140.

Regner, S. (1996) Effects of environmental changes on early stages and reproduction of anchovy in the Adriatic Sea. *Scientia Marina*, **60**, 167—177.

Reid, J. (1967) Oceanic environments of the genus *Engraulis* around the World. In: *California Cooperative Oceanic Fisheries Investigations* (ed. S.I.O. Oceanography), p. 29—33, La Jolla, California.

Richardson, A.J., Mitchell-Innes, B.A., Fowler, J.L., Bloomer, S.F., Verheye, H.M., Field, J.G., Hutchings, L. & Painting, S.J. (1998) The effect of sea temperature and food availability on the spawning success of Cape anchovy *Engraulis capensis* in the southern Benguela. *South African Journal of Marine Science*, **19**, 275—290.

Riginos, C., Hickerson, M.J., Henzler, C.M. & Cunningham, C.W. (2004) Differential patterns of male and female mtDNA exchange across the Atlantic Ocean in the blue mussel, *Mytilus edulis*. *Evolution*, **58**, 2438–2451.

Robalo, J.I., Castilho, R., Francisco, S.M., Almada, F., Knutsen, H., Jorde, P.E., Pereira, A.M. & Almada, V.C. (2012) Northern refugia and recent expansion in the North Sea: the case of the wrasse *Symphodus melops* (Linnaeus, 1758). *Ecology and Evolution*, **2**, 153—164.

Roëgl, F. (1998) Palaeogeographic considerations for Mediterranean and Paratethys seaways (Oligocene to Miocene). *Annalen des Naturhistorischen Museums in Wien*, **99**, 279–310.

Rose, G.A. (2005) On distributional responses of North Atlantic fish to climate change. *ICES Journal of Marine Science*, **62**, 1360–1374.

Sá-Pinto, A., Branco, M.S., Alexandrino, P.B., Fontaine, M.C. & Baird, S.J.E. (2012) Barriers to gene flow in the marine environment: insights from two common intertidal limpet species of the Atlantic and Mediterranean. *Plos One*, **7**, e50330.

Samsun, O., Samsun, N. & Karamollaoúlu, A. (2004) Age, growth and mortality rates of the European anchovy (*Engraulis encrasicolus* L. 1758) off the Turkish Black Sea coast. *Turk. J. Vet. Anim. Sci.*, **28**, 901—910.

Sanz, N., Garcia-Marín, J.L., Viñas, J., Roldán, M. & Pla, C. (2008) Spawning groups of European anchovy: population structure and management implications. *ICES Journal of Marine Science*, **65**, 1635–1644.

Schulte, P.M., Glémet, H.C., Fiebig, A.A. & Powers, D.A. (2000) Adaptive variation in lactate dehydrogenase-B gene expression: role of a stress-responsive regulatory element. *Proceedings of the National Academy of Sciences of the United States of America*, **97**, 6597—6602.

Sinovcic, G. & Zorica, B. (2006) Reproductive cycle and minimal length at sexual maturity of *Engraulis encrasicolus* (L.) in the Zrmanja river estuary (Adriatic Sea, Croatia). *Estuarine, Coastal and Shelf Science*, **69**, 439—448.

Spalding, M.D., Fox, H.E., Allen, G.R., Davidson, N., Ferdaña, Z.A., Finlayson, M., Halpern, B.S., Jorge, M.A., Lombana, A., Lourie, S.A., Martin, K.D., McManus, E., Molnar, J., Recchia, C.A. & Robertson, J. (2007) Marine ecoregions of the world: a bioregionalization of coastal and shelf areas. *BioScience*, **57**, 573–583.

Spanakis, E., Tsimenides, N. & Zouro, E. (1989) Genetic differences between populations of sardine, *Sardina pilchardus*, and anchovy, *Engraulis encrasicolus*, in the Aegean and Ionean seas. *Journal of Fish Biology*, **35**, 417 – 437.

Taylor, E.B. & McPhail, J.D. (2000) Historical contingency and ecological determinism interact to prime speciation in sticklebacks, *Gasterosteus*. *Proceedings of the Royal Society B*, **267**, 2375—2384.

Teacher, A., Andre, C., Merila, J. & Wheat, C. (2012) Whole mitochondrial genome scan for population structure and selection in the Atlantic herring. *BMC Evolutionary Biology*, **12**, 248.

Tudela, S. & Palomera, I. (1997) Trophic ecology of the European anchovy *Engraulis encrasicolus* in the Catalan Sea (northwest Mediterranean). *Marine Ecology-Progress Series*, **160**, 121—134.

Tudela, S., Palomera, I. & Quilez, G. (2002) Feeding of anchovy *Engraulis encrasicolus* larvae in the northwest Mediterranean. *Journal of the Marine Biological Association of the United Kingdom*, **82**, 349—350.

Twatwa, N.M., van der Lingen, C.D., Drapeau, L., Moloney, C.L. & Field, J.G. (2005) Characterising and comparing the spawning habitats of anchovy *Engraulis encrasicolus* and sardine *Sardinops sagax* in the southern Benguela upwelling ecosystem. *African Journal of Marine Science*, **27**, 487—499.

Vermeij, G.J. (2005) From Europe to America: Pliocene to recent trans-Atlantic expansion of cold-water North Atlantic molluscs. *Proceedings of the Royal Society B: Biological Sciences*, **272**, 2545—2550.

Viñas, J., Sanz, N., Peñarrubia, L., Araguas, R.-M., García-Marín, J.-L., Roldán, M.-I. & Pla, C. (2013) Genetic population structure of European anchovy in the Mediterranean Sea and the northeast Atlantic Ocean using sequence analysis of the mitochondrial DNA control region. *ICES Journal of Marine Science*,

Walker, M.J.C. (2004) A lateglacial pollen record from Hallsenna Moor, near Seascale, Cumbria, NW England, with evidence for arid conditions during the Loch Lomond (Younger Dryas) stadial and early Holocene. *Proceedings of the Yorkshire Geological and Polytechnic Society*, **55**, 33—42.

Wallace, J.M. & Gutzler, D.S. (1981) Teleconnections in the geopotential height field during the northern hemisphere winter. *Monthly Weather Review*, **109**, 784—812.

Wares, J.P. & Cunningham, C.W. (2001) Phylogeography and historical ecology of the North Atlantic intertidal. *Evolution*, **55**, 2455–2469.

Weinberg, J.R., Dahlgren, T.G., Trowbridge, N. & Halanych, K.M. (2003) Genetic differences within and between species of deep-sea crabs (*Chaceon*) from the North Atlantic Ocean. *The Biological Bulletin*, **204**, 318—326.

Whitehead, P.J., Nelson, W.S. & Wongratana, T. (1988) *Clupeoid fishes of the World (Suborder Clupeoidei)*. FAO, Rome.

Wilson, A.J., Gíslason, D., Skúlason, S., Snorrason, S.S., Adams, C.E., Alexander, G., Danzmann, R.G. & Ferguson, M.M. (2004) Population genetic structure of Arctic Charr, *Salvelinus alpinus* from northwest Europe on large and small spatial scales. *Molecular Ecology*, **13**, 1129—1142.

Zarraonaindia, I., Pardo, M.A., Iriondo, M., Manzano, C. & Estonba, A. (2009) Microsatellite variability in European anchovy (*Engraulis encrasicolus*) calls for further investigation of its genetic structure and biogeography. *ICES Journal of Marine Science*, **66**, 2176–2182.

Zarraonaindia, I., Iriondo, M., Albaina, A., Pardo, M.A., Manzano, C., Grant, W.S., Irigoien, X. & Estonba, A. (2012) Multiple SNP markers reveal fine-scale population and deep phylogeographic structure in European anchovy (*Engraulis encrasicolus* L.). *PLoS ONE*, **7**, e42201.

ANNEXE I ● DISTRIBUTION OF OLD WORLD ANCHOVIES

Australian anchovy *Engraulis australis* (Shaw, 1970)

Distribution (Whitehead *et al.*, 1988)

Southwest Pacific: Australia (from Queensland at about Cape Capricorn south to southern Tasmania; entire southern coast of Australia, except for Great Australian Bight, and north to Shark Bay, western Australia), including Lord Howe Island and Norfolk island; New Zealand (most of the north island and all but the southeast coast of the south island; Figure 1.11).

Japanese anchovy *Engraulis japonicus* Temminck & Schlegel, 1846

Distribution (Whitehead *et al.*, 1988)

Northwestern Pacific (southern Sakhalin Island, Sea of Japan and Pacific coasts of Japan, and south to almost Taiwan island); rare records off Philippine coasts (Luzon, western Mindanao) and from Indonesia (Manado, Ujung, Pandang, Sulawesi) (Figure 1.11).

White anchovy *Engraulis albidus* Borsa, Collet & Durand, 2004

Distribution (Borsa *et al.*, 2004)

Northwestern Mediterranean, apparently along the shoreline of Camargue; Mauguio, Prevost, and Thau lagoons in southern France; northern Adriatic Sea (Figure 1.11).

Southern African anchovy *Engraulis capensis* Gilchrist, 1913

Distribution (Whitehead *et al.*, 1988)

Southeastern Atlantic and western Indian Ocean (Atlantic coasts of southern Africa from Angola/ Namibia border south to Cape Town, then north to about Maputo; recorded from Mauritius and Seychelles, also in upwelling area around Somalia; Figure 1.11).

Silver anchovy *Engraulis eurystole* (Swain & Meek, 1884)

Distribution (Whitehead *et al.*, 1988)

Northwestern and central Atlantic (Massachusetts south to Florida, northeastern Gulf of Mexico - to about Mississippi Sound, but not recorded elsewhere - and from Venezuelan coast southwards to Amazonas river in the northern Brazil; Figure 1.11).

CHAPTER II ● ANCHOVIES GO NORTH AND WEST WITHOUT LOSING DIVERSITY: POST-GLACIAL RANGE EXPANSIONS IN A SMALL PELAGIC FISH

Gonçalo Silva, John B. Horne and Rita Castilho*

Centre of Marine Sciences, CCMAR/CIMAR Associate Laboratory, University of Algarve, Gambelas, 8005-139 Faro, Portugal

*Correspondence: Centre of Marine Sciences, CCMAR/CIMAR Associate Laboratory, University of Algarve, Gambelas, 8005-139 Faro, Portugal.
E-mail: rcastil@ualg.pt

Accepted for publication: *Journal of Biogeography*

ABSTRACT

Aim As part of an emerging effort to understand the role played by climatic fluctuations in shaping the geographical distributions and abundances of marine organisms, we examined the genetic patterns of leading-edge populations in the European anchovy, *Engraulis encrasicolus*, and its American counterpart, the morphologically similar silver anchovy, *Engraulis eurystole*, in the North Atlantic Ocean.

Location Adults were collected from the western Atlantic, eastern Atlantic (from Norway to Ghana) and western Mediterranean.

Methods A 1045 bp fragment of the mtDNA cytochrome *b* gene was sequenced (n = 312) and 9 microsatellite loci were genotyped (n = 462) for anchovies from 13 locations across the temperate North Atlantic. Populations were surveyed for diversity and differentiation with a range of summary statistics. Multivariate discriminant analysis of principal components was employed to detect the number of genetic clusters in the data and assign individuals to populations based on their microsatellite genotypes. Historical demographic inferences – mismatch distributions and Bayesian Skyline plots – were used to observe population size changes relating to climatic oscillations.

Results Two mitochondrial clades were recovered, consistent with previous studies of *E. encrasicolus*, in which the frequency of each clade varied by latitude. Four genetic clusters corresponding loosely to large geographical regions were identified with microsatellite data. The north-western Atlantic *E. eurystole* was not reciprocally monophyletic for either mtDNA or microsatellite analyses and is probably conspecific with *E. encrasicolus*. Genetic diversity peaked in Iberian populations, but differences in genetic diversity were only statistically significant for the least diverse population, Tangier. The indications of demographic expansion were more pronounced in the southern clade and both mtDNA clades exhibited genetic diversity and expansion imprints that are likely to be older than climatic oscillations of the recent Pleistocene.

Main conclusions The highly mobile nature of anchovies has allowed them to track their optimal thermal physiological conditions during the extreme climate shifts of the Last Glacial Maximum and avoid wholesale population reductions and genetic

bottlenecks. Both north-eastern and north-western Atlantic were probably rapidly recolonized after the Last Glacial Maximum by large numbers of anchovies, such that leading-edge populations retained the genetic diversity of parent populations.

Keywords: Anchovy, climatic fluctuation, dispersal dynamics, *Engraulis*, historical biogeography, leading-edge population, phylogeography, post-glacial expansion, trans-Atlantic migration.

INTRODUCTION

A major goal of evolutionary biogeography is to understand the effects of climatic fluctuations on the geographical distribution and genetic diversity of species. Distribution ranges of marine temperate organisms in the Northern Hemisphere typically shift southwards during periods of climate cooling and northwards during climatic optima, in pursuit of optimal thermal conditions that change latitudinally. These changes in population ranges influence the genetic characteristics of populations, especially those at the periphery of the geographical range (McInerny *et al.*, 2009). By investigating genetic patterns in peripheral, leading- or rear-edge populations, we can better understand how natural populations have responded to changes in climate during their evolutionary past.

During the Last Glacial Maximum (LGM, *c*. 20 ka), sea surface temperatures in the North Atlantic fell by 10 ºC on average (CLIMAP, 1981) and ice-sheet margins descended to 45º N (Renssen & Vandenberghe, 2003) (Figure 2.1a). Sea levels dropped 120 m and altered coastal contours, creating land-bridge barriers to marine organisms in some areas (Lambeck *et al.*, 2002). Throughout the LGM, high latitude coastal areas of the North Atlantic were frequently glaciated and unavailable as pathways for dispersal and population persistence. Although some species can tolerate negative or near 0 ºC temperatures (e.g. *Clupea harengus*; McInerny *et al.*, 2009), the majority of temperate marine organisms would have been unable to cope with LGM conditions throughout most of the North Atlantic and would have experienced dramatic shifts in their distributions and severe population bottlenecks as a result of widespread local extinction in the northern parts of their ranges (Alheit *et al.*, 2012). Therefore, present-day distributions of temperate marine species in the northern areas of the North Atlantic largely represent recolonizations from refugial populations that have occurred since the LGM (Maggs *et al.*, 2008).

Anchovies (genus *Engraulis*) are found in a wide range of temperatures (2–30 ºC) and salinities (5–41 psu) (Whitehead *et al.*, 1988) and these physical parameter ranges define the ecophysiological tolerance and limitations of this species to geographical shifts. The European anchovy, *Engraulis encrasicolus* (Linnaeus, 1758), inhabits the eastern Atlantic, the Mediterranean Sea and the Black Sea (Whitehead *et al.*, 1988), while its less abundant congener, the silver anchovy, *Engraulis eurystole* (Swain &

Meek, 1884), occupies the north and central western Atlantic. Morphological differences between these two species are minor, yet the geographical isolation between both sides of the Atlantic led previous authors to maintain these taxa as different species until 'careful revisionary work' is undertaken (Whitehead *et al.*, 1988).

Even though anchovies have a large dispersal potential, genetic studies show moderate levels of genetic structure among populations in the Mediterranean Sea (e.g. Magoulas *et al.*, 2006) and in the eastern North Atlantic (Petitgas *et al.*, 2012). The geographical distribution of genetic diversity indicates that the potential for dispersal may be limited (Magoulas *et al.*, 2006) by such factors as retentive currents, complex shorelines, oceanic fronts or isolation by distance (Agostini & Bakun, 2002). In addition to the genetic structure imposed by partial isolation between locations, two deeply separated mitochondrial DNA (mtDNA) clades varying in relative frequency were identified in *E. encrasicolus*, which appear to reflect ancient isolations and recent secondary contact (Magoulas *et al.*, 2006).

Our study is part of an emerging effort to understand the role played by climatic fluctuations in shaping the geographical distributions and genetic diversity of marine organisms in the North Atlantic (e.g. Maggs *et al.*, 2008, and references therein). We provide a phylogeographical study based on a large fragment of the mitochondrial cytochrome *b* gene (cyt *b*) and nine nuclear microsatellites. The emphasis of the work is: (1) on the levels of genetic diversity of recently established northern populations compared to persistent more southern populations; and (2) on the genetic diversity and differentiation between *E. encrasicolus* and *E. eurystole*, in the eastern and western North Atlantic, respectively. We build on previous studies (Magoulas *et al.*, 2006; Bouchenak-Khelladi *et al.*, 2008; Sanz *et al.*, 2008; Borrell *et al.*, 2012; Zarraonaindia *et al.*, 2012) to provide insights on the timeframe, the source and the mode of colonization of marginal populations. We predict that recently colonized northern populations of the European anchovy are comparable in genetic diversity to populations further south, given that anchovies are highly mobile pelagic fish able to rapidly track suitable habitats in large numbers. Recolonization involving large numbers of migrants does not result in decreased genetic diversity in the new populations, as is expected when populations are originally founded by only a few colonizers (McInerny *et al.*, 2009). For the same reason, we do not predict finding

signals of sudden demographic population growth within the post-LGM timeframe, contrary to expectations of newly founded populations from low dispersal species.

MATERIAL AND METHODS

Sample collection, DNA extraction and PCR amplification

Samples of *E. encrasicolus* and *E. eurystole* were collected at a total of 13 sites in the eastern and western North Atlantic Ocean (Table 1, Figure 2.1a) and at one site in the western Mediterranean, where the *E. encrasicolus* population was previously described as being genetically close to Atlantic populations (Magoulas *et al.*, 2006). Fish were purchased at small coastal fish markets, as artisanal fisherman do not venture far, or were collected on scientific cruises (see Acknowledgements). A small portion of white muscle or fin was preserved in 96% ethanol and stored at −20 °C. Total genomic DNA was extracted by a saline method (Sambrook & Russell, 2001). A 1045 bp fragment from the mitochondrial cyt *b* and nine nuclear microsatellite loci were amplified by polymerase chain reaction (PCR). Sequences and fragment lengths (using the GeneScan-500 LIZ standard; Life Technologies Europe BV, Porto, Portugal) were obtained using an ABI 3130XL automated sequencer (Applied Biosystems, Foster City, CA, USA) see Appendix S1 in the Supporting Information for primers, amplification details and fragment characteristics). Microsatellite raw allele sizes were manually scored in STRAND 2.4.59 (Toonen & Hughes, 2001).

Genetic analysis

Cyt *b* sequences were aligned using CLUSTALX 2.0.3 with default settings, implemented in GENEIOUS 5.4 (Drummond *et al.*, 2011) and checked manually. We used COLLAPSE 1.2 (Posada, 2004) to reduce sequences to haplotypes. For cyt *b*, number of individuals (n), frequency (f), number of haplotypes (n), number of private haplotypes (n_P), and haplotype (h) and nucleotide diversities (π) were calculated in ARLEQUIN 3.5.1.2 (Excoffier & Lischer, 2010). To compare haplotype diversities, the Salicru method (Salicru *et al.*, 1993) was used for both overall diversity and pairwise locations.

In previous analyses, the European anchovy displayed a high proportion of null alleles for microsatellite loci (e.g. Zarraonaindia *et al.*, 2009). Therefore, we used MICRO-CHECKER 2.2.3 (van Oosterhout *et al.*, 2004) and FREENA (Chapuis & Estoup, 2007) to calculate the frequency of null alleles at different loci, and F_{ST} values (Weir, 1996) were re-calculated after correcting for the presence of null alleles.

Summary statistics, number of individuals (n), mean number of alleles (n_a), observed heterozygosity (H_O) and expected heterozygosity (H_E) were calculated for each location and for each locus with GENODIVE, whereas mean allelic richness (A_r) was calculated with FSTAT 2.9.3.2 (Goudet, 1995).

Genetic structure and population differentiation

To examine the relationship between mitochondrial haplotypes, a minimum spanning network was constructed with ARLEQUIN 3.5.1.2 (Excoffier & Lischer, 2010) and visualized with HAPSTAR (Teacher & Griffiths, 2011). Population pairwise genetic differentiation was estimated with G_{st_est} (Hedrick & Goodnight, 2005) and Jost's D_{est} value (Jost, 2008) following Pennings *et al.* (2011) for mtDNA and using the R package DIVERSITY for microsatellites (Keenan *et al.*, 2013).

Spatial analysis of shared alleles (SAShA) was used to detect subtle geographical structuring of mtDNA haplotypic and microsatellite allele co-occurrences (Kelly *et al.*, 2010). This spatial analysis is done by comparing the observed distance distribution (ODD) between occurrences of each allele, with a null expected distance distribution (EDD) generated from the data. To test whether the geographical pattern of genetic differentiation is caused by isolation by distance (IBD) we ran Mantel tests for pairwise matrices between geographical distances (kilometres) of the shortest marine path among locations measured in GOOGLE EARTH and genetic differentiation $[D_{est}/(1 - D_{est})]$. Mantel tests (1000 randomizations) were performed using MANTEL.XLA 1.2.4 (Briers, 2003). To determine genetic structuring and individual assignments based on the autosomal microsatellite data set, we used discriminate analysis of principal components (DAPC) a multivariate ordination method (Jombart *et al.*, 2008) implemented in the ADEGENET package (Jombart, 2008) of R 2.15.3 (R Development Core Team, 2009; http://www.r-project.org). This method does not

assume Hardy–Weinberg equilibrium or linkage disequilibrium and is more appropriate for situations where such assumptions are not met, as is often the case with anchovies (Zarraonaindia *et al.*, 2009), than conventional approaches such as STRUCTURE (Pritchard *et al.*, 2000). DAPC yields similar results to STRUCTURE (van der Meer *et al.*, 2012; Molfetti *et al*., 2013) predicting genetic clusters based on the results of principal components analysis. A user-specified number of principal components is retained to act as predictors, in this case, representing 90% of the cumulative variance. The optimal number of populations was identified as the one for which the Bayesian information criterion (BIC) showed the lowest value and after which BIC increased or decreased by the least amount. We followed Jombart's (2013) recommendation to perform a cross-validation of the robustness of cluster assignments, by splitting the data in two parts. Twenty-five per cent of the individuals from each sample were chosen as training data and the remaining 75% were hold-out data. Clustering and DAPC were re-computed based exclusively on the training data.

Historical demography

Two neutrality tests were used to assess population expansion, Fu's F_S (Fu, 1997) and R_2 (Ramos-Onsins & Rozas, 2002). Mismatch distributions, frequencies of pairwise differences between haplotypes, were estimated for each clade. Significance of F_S and R_2 was evaluated by comparing the observed value with a null distribution generated by 10,000 coalescent simulations, using the empirical population sample size and observed number of segregating sites implemented in DNASP 5.10 (Librado & Rozas, 2009). Sum of squares deviations (SSD*)* and raggedness statistics (*rg*; Harpending, 1994) significances were obtained based on 10,000 permutations. Lower R_2 and *rg* values are expected for a population growth scenario (Harpending, 1994; Ramos-Onsins & Rozas, 2002). These analyses were performed in ARLEQUIN 3.5.1.2 (Excoffier & Lischer, 2010). DNASP (Librado & Rozas, 2009) was used to obtain observed and expected distributions under the constant population model and the growth population model. Time and magnitude of inferred population expansion can be determined by mismatch analysis (Rogers & Harpending, 1992) by calculating θ_0, θ_1 and τ, where $\theta_0 = 2N_0\mu$ (N_0 = population size before expansion); $\theta_1 = 2N_1\mu$ (N_1 = population size after expansion); and $\tau = 2ut$ (u = is the mutation rate over the whole

sequence; t = time since population growth expansion). Coalescence analysis requires an estimate of generation time and mutation rate. The generation time, defined as the average age of reproductively mature individuals in the population, is one year in *E. encrasicolus* (Parada *et al.*, 2003). The only specific divergence rate available for cyt *b* of Engraulidae is 1.9% Myr^{-1} (Grant *et al.*, 2010), roughly equivalent to a substitution rate of 1% Myr^{-1}, well within what has been usually accepted in many species of bony fishes (Bowen *et al.*, 2001, and references therein). Given a mean generation time of 1 year, the substitution rate per site per generation in cyt *b* is 1×10^{-8}. However, this rate is merely a heuristic and absolute molecular dating was not inferred using this rate. The Bayesian skyline plot (BSP) analysis of population size history, using BEAST 1.7.5 software (Drummond *et al.*, 2012) was applied to cyt *b* sequences. Genealogies were combined from 10 runs of 1×10^8 steps with a burn-in of 1×10^7 (for clade A) and 10 runs of 10^7 steps with a burn-in of 1×10^6 steps (for clade B). Runs were performed under the GTR+I+Gamma model, with a strict molecular clock and a stepwise skyline model with 20 piecewise intervals. Genealogies and model parameters were sampled every 10,000 iterations and operators were optimized automatically. Effective sample size (ESS) for each parameter exceeded 200. The trajectories were plotted by TRACER 1.5 (Rambaut & Drummond, 2007). Mismatch distributions and BSP trajectories were performed to evaluate expected signatures of demographic expansion between subsets of the data, judged to represent a coherent demographic history based on the geographical settings and the main genetic mtDNA structure.

RESULTS

Mitochondrial DNA

A total of 312 individuals were analysed (accession numbers JQ716609–JQ716731, JQ716748–JQ716756 and JX683020–JX683113). Sequences were polymorphic at 216 sites (119 parsimony-informative) defining 210 haplotypes, of which 97 were singletons and 184 (87.6%) were private (i.e. unique to a single locality). Haplotype diversity (h) was high, ranging from 0.909 to 0.980 in northern locations (Bay of Biscay to Norway and USA) and from 0.423 (Tangier) to 1.000 in southern locations (Table 2.1). Differences in haplotype diversity among the 13 locations were only

significant when Tangier was included (χ^2 = 39.54, d.f. = 12, P > 0.01; range of z-values = 0.0–1.7; global P = 0.00009). All haplotype diversity pairwise comparisons with Tangier were also significant. When Tangier was removed from the dataset, global haplotype diversity values were not significantly different (χ^2 = 8.01, d.f. = 11, range of z-values = −5.63 to 5.65, P = 0.71), nor were any of the remaining pairwise comparisons. Nucleotide diversity (π) was low, ranging from 0.3% (Tangier) to 1.5% (Bay of Biscay) (Table 2.1). Locations north of Tangier displayed high nucleotide diversities while Senegal, Guinea-Bissau and Ghana had lower diversity values (Table 2.1). Diversities in the sample from the Alboran Sea (h = 1.000, π = 0.007) were similar to those for north-eastern Atlantic samples. On the whole, diversity measures did not decrease towards the marginal northern locations (see Appendix S2).

Table 2.1 Sample locations, sample abbreviations, collection dates, sample sizes and summary statistics for a 1045 bp sequence fragment of the mtDNA cytochrome *b* and eight nuclear microsatellites of the anchovies *Engraulis encrasicolus* and *E. eurystole*.

Location	Code	Long	Lat	Mitochondrial cytochrome *b*						Microsatellites (8 loci)						
				n	n_h	n_P	n_P/n_h	*h*	π	*n*	*Aavg*	$A_r = 9$	$A_r = 18$	*Effnum*	H_O	H_E
Norway	NO	10.6	59.0	24	17	5	0.29	0.953	0.009	40	13.1	7.6	10.5	7.0	0.714	0.850
Poland	PL	16.5	54.6	9	7	3	0.43	0.917	0.014	9	8.4	8.4	-	6.1	0.736	0.864
English Channel	EC	0.1	50.8	27	18	6	0.33	0.963	0.010	45	13.0	6.8	9.6	6.6	0.729	0.837
Bay of Biscay	BB	-2.9	43.5	23	19	11	0.58	0.980	0.015	45	16.0	8.4	11.6	6.9	0.762	0.837
Portugal – North	PN	-8.8	40.7	25	24	20	0.83	0.997	0.013	45	17.3	8.1	10.9	8.4	0.825	0.882
Portugal – South	PS	-8.4	37.1	29	27	22	0.81	0.995	0.011	43	17.9	8.0	12.3	8.6	0.777	0.880
Málaga	ML	-4.3	36.6	31	31	27	0.87	1.000	0.007	46	16.6	9.5	11.3	8.3	0.758	0.880
Tangier	TG	-5.9	35.9	38	10	8	0.80	0.423	0.003	46	14.0	6.1	9.3	5.3	0.686	0.800
Canary Islands	CA	-15.0	28.3	24	23	16	0.70	0.996	0.007	42	17.8	8.9	11.8	9.9	0.792	0.888
Senegal	SN	-17.6	14.8	25	25	22	0.88	1.000	0.006	37	12.5	7.3	8.5	5.0	0.693	0.797
Guinea-Bissau	GU	-14.2	9.7	20	20	20	1.00	1.000	0.006	19	13.1	7.8	12.8	6.8	0.653	0.868
Ghana	GH	0.0	5.6	25	25	25	1.00	1.000	0.006	27	15.5	8.4	13.6	8.7	0.766	0.883
USA*	US	-66.1	41.5	12	9	8	0.89	0.909	0.004	18	13.5	9.5	13.5	8.0	0.793	0.894
Total				312	210	208	0.99	0.985	0.014	462	14.6	25.1	32.0	6.7	0.745	0.858

Long, longitude; Lat, latitude; n, number of individuals; n_h, number of haplotypes; n_P, number of private haplotypes; n_P/n_h, proportion of private haplotypes; h, haplotype diverstity; π, nucleotide diversity; *Aavg*, average number of alleles, *Ar*, allelic richness; *Effnum*, effective number of alleles; H_O, observed mean heterozygozity; H_E, expected mean heterozygozity; * putative *Engraulis eurystole*.

Haplotypes grouped into two previously described clades (Magoulas *et al.*, 1996) with frequencies shifting clinally. Clade A haplotypes were more frequent in the southern samples while clade B predominated in samples north of the English Channel. North-eastern Atlantic locations exhibited a higher proportion of shared haplotypes, in contrast with more southern locations (Figure 2.1a). Both clades were

present in all locations, with the exception of West Africa (Senegal, Guinea-Bissau and Ghana) and USA, where clade B is absent (Figure 2.1a). Strong clade-frequency shifts occurred between the English Channel and the Bay of Biscay. Net sequence divergence between clades was 1.87% (SE 0.37%).

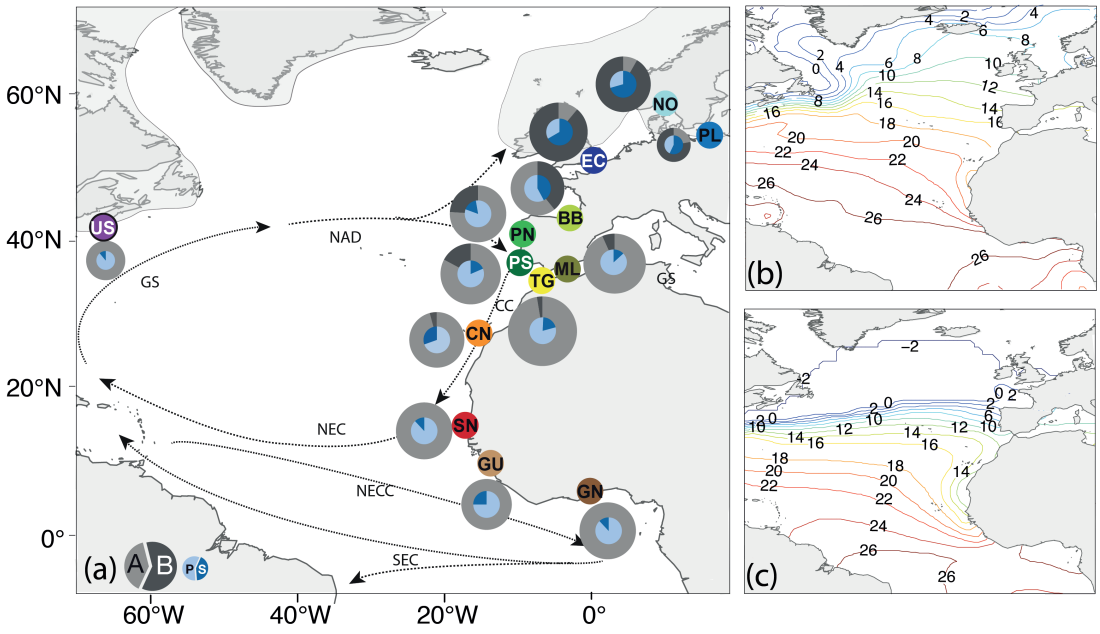


Figure 2.1 (a) Present-day Atlantic Ocean and western Mediterranean Sea showing *Engraulis encrasicolus* and *E. eurystole* sample locations. Light grey shading represents the extent of terrestrial ice sheets at the Last Glacial Maximum (LGM; 18–20 ka). Sample locations examined for mtDNA cytochrome *b* (cyt *b*) are indicated by coloured circles. Grey circles represent mtDNA cyt *b* clades A (light grey) and B (dark grey) proportions. Blue circles represent mtDNA cyt *b* private (light blue) and shared (dark blue) haplotype proportions. Black dashed lines represent main oceanic currents, and arrows show directionality (GS, Gulf Stream; NAD, North Atlantic Drift; CC, Canaries Current; NEC, North Equatorial Current; NECC, North Equatorial Counter Current; SEC, South Equatorial Current). Sample abbreviations are defined in Table 2.1. (b) Climatological present-day January mean sea surface temperature fields (SST, °C). (c) Climatological mean sea surface temperature fields (SST, °C) simulation results for the LGM based on seasonal estimates of CLIMAP (1981).

In the haplotype network, clades A and B were separated by 14 mutations (Figure 2.2). The two clades exhibited different haplotype patterns: clade A was characterized by multiple star-like radiations with relatively shallow genetic divergences; clade B lacked distinct star patterns and exhibited many unsampled or extinct haplotypes. USA haplotypes were separated by two to five mutations from the eastern Atlantic

haplotypes, with one haplotype shared with the Canary Islands. Furthermore, western Atlantic haplotypes were not reciprocally monophyletic with respect to European lineages.

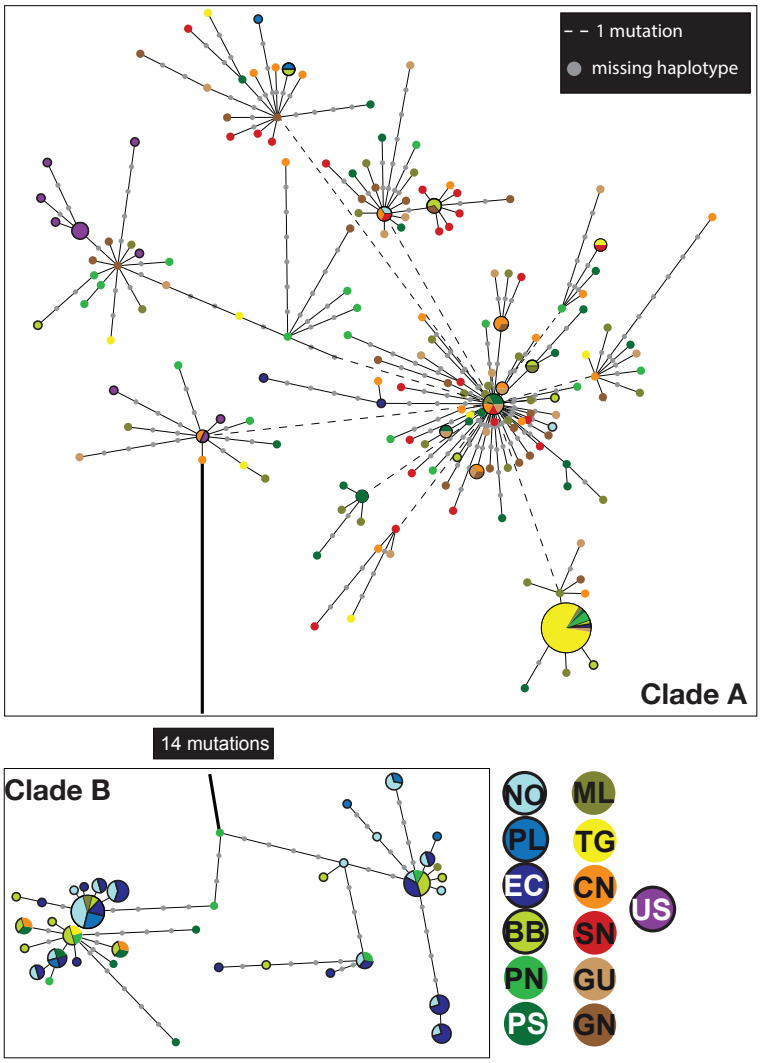


Figure 2.2 Mitochondrial DNA cytochrome *b* haplotype network of *Engraulis encrasicolus* and *E. eurystole* constructed with a median-joining algorithm. Haplotypes are coloured according to Figure 2.1 labels and sample abbreviations are defined in Table 2.1. A black outline of haplotypes indicates their origin from recently colonized areas (USA and the north of Europe, English Channel and Bay of Biscay).

SAShA analyses rejected population panmixia (ODD = 785 km, EDD = 3295 km, $P < 0.001$), because the mean geographical dispersal of alleles was smaller than expected. Pairwise differentiation was much greater when measured using D_{est} than

G_{st_est} (Figure 2.3). Additionally, adjacent locations had consistently lower D_{est} values. G_{st_est} values were mostly between 0 and 0.05 with the notable exception of Tangier and USA. IBD was significant, when considering all samples ($r^2 = 0.298$, $P < 0.001$) and also when excluding the USA ($r^2 = 0.222$, $P < 0.001$).

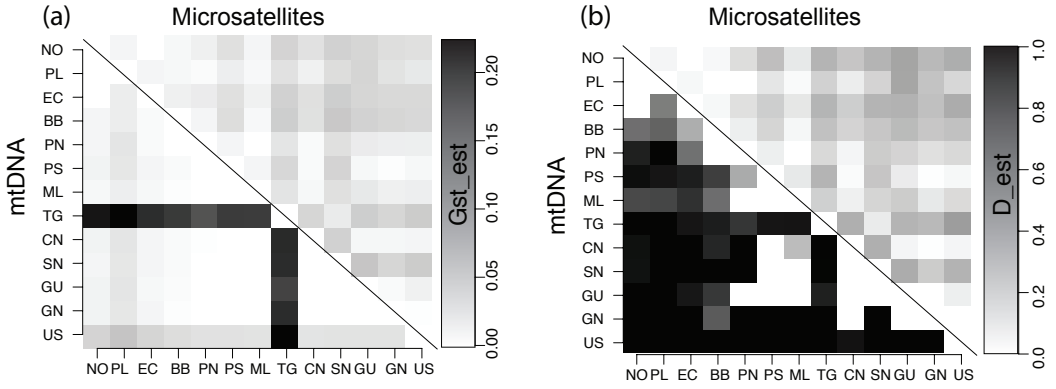


Figure 2.3 Pairwise-location G_{st_est} (a) and D_{est} (b) values for microsatellites (above diagonal) of and mtDNA sequences (below diagonal) of *Engraulis encrasicolus* and *E. eurystole*. Site abbreviations defined in Table 2.1.

Clade A displayed a typical unimodal distribution (Figure 2.4a), closely matching the expectations of the growth population model, whilst clade B was clearly bimodal (Figure 2.4b). However, SSD and raggedness probability values did not allow rejection of the sudden expansion model for both clades and Fu's F_S values were negative and significant which might also indicate population expansion. Bayesian skyline plots revealed remarkable population size stability throughout most of the evolutionary history of the lineages (Figure 2.4c, d).

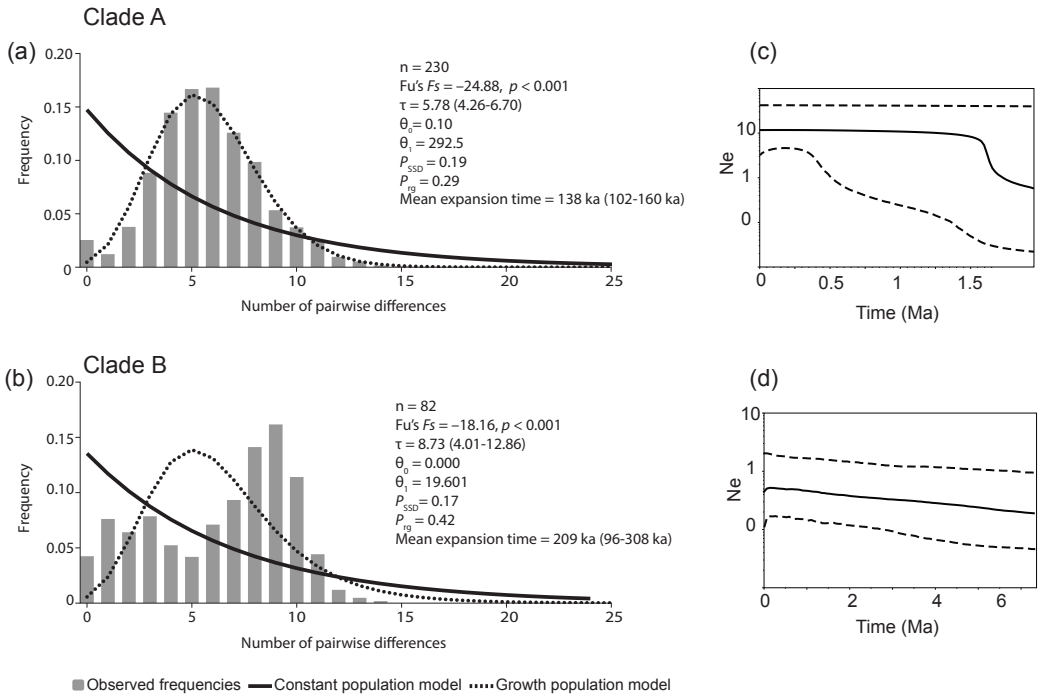


Figure 2.4 Mismatch distribution of *Engraulis encrasicolus* and *E. eurystole* in the two panels on the left (a and b), indicating number of individuals in the analysis (*n*), Fu's F_S test of selective neutrality and population expansion, evolutionary expansion age in mutational units (τ), effective population size before (θ_0) and after (θ_1) population expansion and mean expansion time in units of thousand years (ka). Note that the range in expansion age corresponds to the 95% confidence interval of (τ). P_{SSD} represents the significance of sum of squares deviations and P_{rg} the significance of the raggedness statistics. Bayesian skyline plots are shown in the two panels to the right (c and d). The *y*-axis (note logarithmic scale) indicates effective population size estimates multiplied by generation time. Operational time-scales (*x*-axis) are based on per site mutation rates and a 1-year generation interval. Solid lines are median estimates of effective population size (N_e), dotted lines are the 95% posterior density limits. Time at the origin represents the present day.

Microsatellite DNA

Multilocus genotypes from 462 anchovies were obtained from 13 locations. The number of alleles per locus varied from 14 (locus 135) to 152 (locus 41.2) over all locations (Appendix S3, Table 2). Mean allelic richness, standardized for comparison across a minimum common sample size of nine individuals, ranged from 6.1 (Tangier) to 9.5 (Malaga and USA). Expected heterozygosity (H_E) varied between 0.797 (Senegal) and 0.894 (USA) and the observed heterozygosity (H_O) varied between 0.653 (Guinea-Bissau) and 0.825 (Portugal north) (Table 2.1). North Atlantic

locations south of the English Channel have higher levels of allelic richness than northern European locations. Allelic diversities in the Mediterranean sample (A_{avg} = 16.6) were similar to those in nearby Atlantic locations (Portugal north, Portugal south, Tangier and Canaries (A_{avg} = 14.0–17.9). Average expected heterozygosities showed no geographical pattern (Appendix S2). Locus 41.2 was identified as possibly having null alleles and/or large allele dropout, as it presented high overall G_{IS} (Table 2.2). Null allele uncorrected and corrected estimated per locus F_{ST} values are very similar (Table 2.2), but because of software limitations those comparisons could not be performed for locus 41.2. Therefore, we removed this locus from further analysis to ensure that erroneous estimations were not introduced by the hypervariable nature of the locus.

SAShA suggests that anchovies are not panmictic (ODD = 2870 km, EDD = 3094 km, $P < 0.001$). Also, IBD was significant only when the western Atlantic was included (with USA: $r^2 = 0.053$, $P = 0.019$ and without USA: $r^2 = 0.058$, $P = 0.210$). No patterns for genetic differentiation of microsatellite loci (assessed both by D_{est} and G_{st_est}) were evident (Figure 2.3). Confidence intervals (data not shown) all excluded zero, indicating significant pairwise differentiation.

Table 2.2 Summary statistics across nine microsatellite loci of the anchovies *Engraulis encrasicolus* and *E. eurystole*. *Na*, total number of alleles; *Ar*, average number of alleles across locations; *Effnum*, effective number of alleles; H_O, observed heterozygosity; H_E, expected heterozygosity within populations; H_t, total heterozygosity; H'_t, corrected total heterozygosity; G_{IS}, inbreeding coefficient; Null allele$_F$, null allele frequency by FREENA; Null allele$_M$, null allele frequency by MICRO-CHECKER; F_{ST}, global per locus F_{ST} of Weir (1996); F_{ST_ENA}, global per locus F_{ST} of Weir (1996) using the ENA correction described in Chapuis & Estoup (2007).

Locus	*Na*	*Ar* = 9	*Effnum*	H_O	H_E	H_t	H'_t	G_{IS}	Null allele$_F$	Null allele$_M$	F_{ST}	$F_{ST\ ENA}$
L10	72	11.4	7.2	0.842	0.877	0.922	0.925	0.040	0.013	†	0.047	0.046
L135	14	7.6	5.8	0.756	0.845	0.862	0.864	**0.105**	0.036	0.068	0.024	0.024
L407	55	10.1	6.5	0.828	0.863	0.911	0.915	0.040	0.016	†	0.059	0.058
L452	48	12.3	11.0	0.893	0.926	0.952	0.954	0.036	0.011	†	0.029	0.028
L508	26	10.2	6.6	0.601	0.871	0.922	0.927	0.309	0.142	0.032	0.068	0.064
L291a	28	8.2	6.5	0.628	0.866	0.881	0.882	**0.275**	0.119	0.031	0.015	0.013
L41.1	57	8.5	5.0	0.604	0.819	0.858	0.861	**0.262**	0.102	0.015	0.051	0.043
L41.2	152	16.0	20.8	0.617	0.977	0.987	0.988	**0.369**	*	0.005	0.01	*
L291b	17	6.1	4.6	0.807	0.798	0.809	0.810	-0.011	0.002	†	0.017	0.017
Overall	52	10.0	8.2	0.731	0.871	0.900	0.903	**0.161**				

Weir (1996) using the ENA correction described in Chapuis & Estoup (2007).
Significant values are in bold, $P < 0.05$.
*, number of alleles exceeds software capability.
†, software does not correct these loci.

In the initial step of DAPC, 83 principal components of the PCA were retained as input to discriminant analysis, which accounted for more than 90% of the total variance. Based on the BIC, we chose models with 5–12 clusters to provide useful data summaries. Visual inspection of scatterplots identified the following K = 4 clusters (Figure 2.5): (1) Norway, English Channel, Bay of Biscay; (2) Portugal north and Malaga; (3) Portugal south, Canaries, Guinea, Ghana and USA; (4) Tangier and Senegal. Poland had an intermediate position, between group one and group two. The horizontal axis separated Tangier and Senegal from the rest, while the vertical axis set northern locations (Norway, Bay of Biscay, English Channel and Poland) apart from more southern sites (Portugal south, Canaries, Guinea and Ghana) and USA. The mean assignment rate for each individual to the correct genetic cluster was 78%. However, after the cross-validation, the assignment power dropped to 31%.

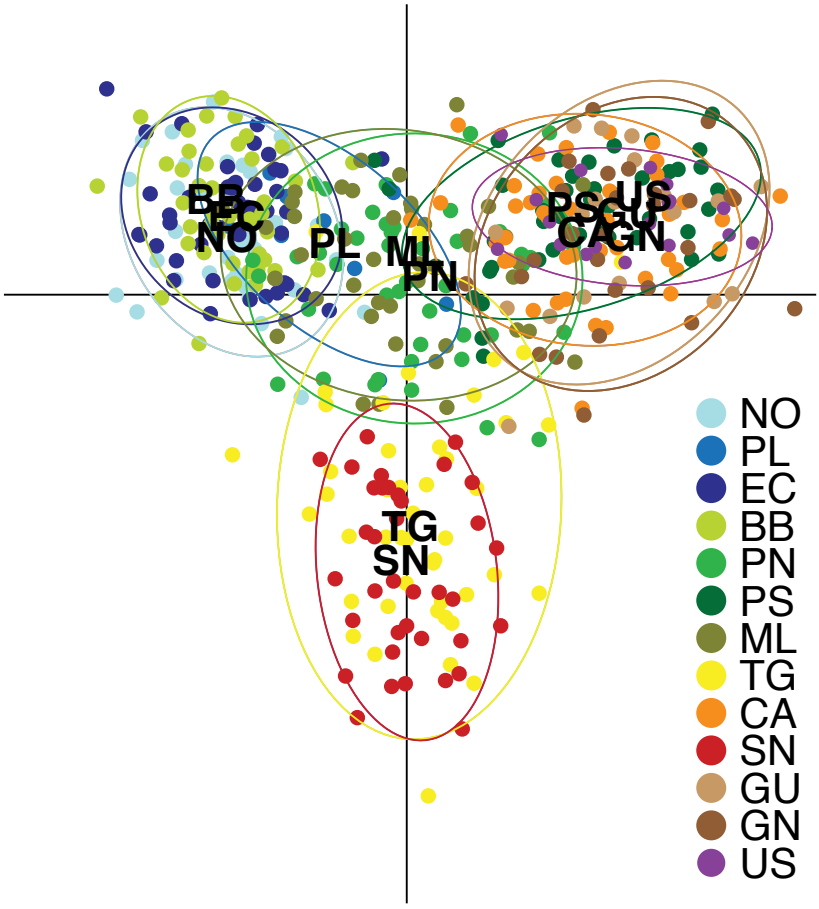


Figure 2.5 Discriminant analysis of principal components (DAPC) of multi-locus *Engraulis encrasicolus* and *E. eurystole* genotype data for all study locations. Individual genotypes appear as circles, and sample code (defined in Table 2.1) represents the centre of dispersion of each group. Horizontal and vertical axes are the first two principal components, respectively.

DISCUSSION

The distribution of genetic diversity and isolation patterns among populations of the European anchovy indicates post-glacial dispersals both to the western and north-eastern Atlantic. Based on our results we suggest that anchovies inhabiting the north-western Atlantic, nominally designated as *E. eurystole*, are conspecific with *E. encrasicolus* and do not merit species status. These anchovies represent an extension of the range of European anchovies across the Atlantic. The north-eastern and north-western Atlantic may have been rapidly recolonized since the LGM by large numbers of anchovies, such that leading-edge populations retained the genetic diversity of parent populations.

Genetic population structure

In the present work, population genetic structure of North Atlantic anchovies was not concordant between mtDNA and nuclear microsatellites. Mitochondrial DNA displays a deep divergence between two clades, with little apparent geographical structure, apart from the clinal frequencies of clades, while microsatellites provide evidence of the existence of four main clusters (Norway–Bay of Biscay–English Channel; Portugal north–Malaga; Portugal south–Canaries–Guinea-Bissau–Ghana–USA; Tangier–Senegal; Poland occupies an intermediate position between the first and the second group). These results were consistent with other studies, so far as the locations of samples overlapped with those in our work (e.g. Magoulas *et al.*, 2006; Bouchenak-Khelladi *et al.*, 2008; Sanz *et al.*, 2008; Zarraonaindia *et al.*, 2012). The distributions of mtDNA, allozyme, microsatellite and nuclear gene markers in the various studies consistently showed a complex population structure for European anchovies in the Atlantic Ocean and Mediterranean Sea.

The inclusion of fish in non-spawning condition from 8 out of the 13 sample locations requires the potential caveat of underestimating population differentiation, as transient migrants, not representative of the assumed local population origins, may have been sampled. Indeed, we detected some degree of connectivity between populations (e.g. Norway, English Channel and Bay of Biscay) that possibly reflects ephemeral migrants between those locations. Both marker types showed clear

differences in population structure, with mtDNA displaying a strong geographical cline for two clades. This is in contrast to microsatellite-based inferences of four genetic clusters corresponding loosely to large geographical regions. Combining mtDNA and microsatellites markers has the benefit of simultaneously assessing both historical and contemporary imprints on population structure (Limborg *et al.*, 2012). However, besides the effect of differences in mutation rates, we cannot eliminate other possible causes for mitonuclear discrepancies, such as sex-biased dispersal or selection.

Recent colonization of northern European seas

Many of the north-eastern Atlantic mtDNA haplotypes are shared with southern locations (Figures 2.1a) and microsatellites also show a close relationship between northern populations and those inhabiting the Bay of Biscay (Figure 2.5). Because north-eastern Atlantic areas would have been uninhabitable by anchovies during LGM conditions, our data are best explained by a southern origin of recently established northern populations. The geographical range expansion to the northern Atlantic Ocean must have occurred post-LGM, after the retreats of the British and the Fennoscandian ice sheets, because the minimum temperature presently tolerated by the European anchovy is 0–2 ºC (Alheit *et al.*, 2012) (Figure 2.1b). During the LGM this temperature range could only be found further south, in the English Channel (de Vernal & Hillaire-Marcel, 2006) (Figure 2.1c). Besides temperature, the abundance of anchovies is also dependent on local productivity, and during the LGM the nearshore areas of the North Atlantic experienced significantly reduced productivity (de Vernal & Hillaire-Marcel, 2006), which could have also had an impact on the persistence of the species in that region.

We detected a clear signal of demographic expansion for clade A (Figure 2.4a, c). Yet, for a population expansion of clade A to be more recent than the LGM, the mutation rate of cyt *b* in European anchovy would have to be at least 8×10^{-8}. Calibrated mutation rates for the cyt *b* locus in most marine taxa (Bowen *et al.*, 2001; Lessios, 2008) and other engraulids (Grant *et al.*, 2012) are closer to 2×10^{-8}. Therefore, population expansion in clade A is probably much older than the LGM. The signal of demographic expansion is less pronounced in clade B. However, like

clade A, much of the genetic diversity observed in clade B is most likely older than the LGM (Figure 2.4b, d) and there is no evidence to suggest a population bottleneck or pronounced demographic expansion during the last 20 kyr. Therefore, this species appears to have been able to track its thermal habitat preferences along coastlines and shift its geographical range latitudinally during recent climatic oscillations, without experiencing population bottlenecks or severe losses of genetic diversity, as has been the case in other species. For example, in Pacific herring (*Clupea pallasii*) Bayesian skyline plots display rapid population growth after the LGM and a flat population history during previous climatic fluctuations, suggesting the erosion of genetic signals from prior climate-related population disturbances (Grant *et al.*, 2012). In the European anchovy the signal of relatively ancient expansions is still visible and has not been eroded during more recent climatological oscillations.

The European anchovy does not display genetic signatures of recent, pronounced population contractions or expansions, particularly in the northern range of the distribution, where those effects would be most expected. This species displays high mtDNA and microsatellite diversity levels at all locations at its distributional margins. This observation probably reflects mass dispersal at the colonization front. In fact, there are records of swift appearances of anchovies in the North Sea during the 1990s, after a 30-year absence, and more recently the distribution range has extended to the Shetland Islands, southern Norway and into the Baltic Sea (Petitgas *et al.*, 2012). Fast colonizations of northern regions during favourable conditions are therefore plausible events and the dispersal capacities are pivotal for preserving the species genetic diversity under range contractions or shifts (Arenas *et al.*, 2011). We posit that European anchovies are capable of fast-tracking optimal habitat conditions, and their swift mass movements prevent the loss of genetic diversity that would otherwise result from climate-related habitat disturbances, such as occurred during the LGM.

Western North Atlantic anchovies

Our results show that populations of the putative species *E. eurystole* in the north-western Atlantic belong to the same clade A observed in European anchovies. The cyt *b* haplotype network places most clade A haplotypes of *E. eurystole* in the European subclades, with one haplotype shared between the eastern Atlantic (Canary Islands)

and the USA. Moreover, divergences between western and eastern Atlantic locations are smaller than among eastern Atlantic locations. Depending on effective population size and mutation rate, populations persisting in the western Atlantic over one or more glacial cycles might be expected to show considerable divergence from European source populations and to have deeply coalescing haplotype genealogies consistent with long persistence (Wares & Cunningham, 2001). Yet, as this is not the case, not only does *E. eurystole* appear to be conspecific with *E. encrasicolus* (virtually no differences in morphology; Whitehead *et al.*, 1988) but north-western Atlantic populations are likely to have been derived from the eastern Atlantic after the LGM.

We suggest that north-western Atlantic populations originated from genetically diverse sources (putatively western and central African populations), producing a non-monophyletic population in the USA, which displays similar divergences from all other locations. Furthermore, divergence is probably recent, given the close haplotype relationship with the eastern counterparts (Figure 2.2). Our results are most consistent with a massive wave of colonists, comparable to those moving into northern European waters in response to sea surface temperature increases. Given that we estimated a high genetic diversity in putative source populations, the diversity in the western Atlantic could be the result of a single colonization event and not necessarily the outcome of multiple colonizations.

Species with amphi-Atlantic distributions imply present or historical trans-Atlantic migration(s) within periods of favourable ecophysiological conditions. Western Atlantic anchovies may have reached the Americas along one of two possible current-mediated dispersal routes, the ‘northern’ route or the ‘equatorial’ route. Opportunities for dispersal across the North Atlantic are likely to be limited to interglacial periods, when warming allows the stepping-stone colonizations of mid-North Atlantic islands that are typical of east-to-west dispersals (Vermeij, 2005). North Atlantic warming provided the conditions required for anchovy spawning (14 ºC: Motos, 1996) and larval growth (16 ºC: Urtizberea *et al.*, 2008) in the waters around the Shetland Islands, which may have acted as stepping-stone areas. However, the north-eastern Atlantic samples consist mostly of clade B individuals, contrasting with the western Atlantic, which is composed entirely of clade A. Moreover, European anchovies have never been recorded in the Faroe Islands, Iceland or Greenland, and the differentiation

between the north-eastern and the western Atlantic and the lack of clade B individuals renders a 'northern' route unlikely.

Alternatively, an 'equatorial' route implies that anchovy populations off West and Central Africa have reached the Americas with the North Equatorial Current (NEC), running from West Africa to north-eastern South America. The NEC flows presently at a maximum speed of 15 cm s^{-1}, meaning that by purely passive drift, anchovies would take *c*. 200 days, probably much less if adult swimming is taken into account, to cross over the 2500 km from Guinea to Brazil continental platforms. A closer genetic relationship between western Atlantic and West Africa locations and the shared haplotype between the western Atlantic and Canary Islands further support the hypothesis of an 'equatorial' dispersal route. Dispersals from African populations in which clade A haplotypes predominate preclude the need to invoke clade sorting or founder effects to explain the absence of clade B haplotypes in north-western samples.

After crossing the Atlantic, anchovies could have followed northward currents (e.g. the Caribbean, Antilles and Florida and the Gulf Stream) through the Caribbean Sea to higher latitudes, up to their temperature tolerances. Apparently, high temperatures would have not limited dispersals of anchovies in these regions because temperatures over the last 100 ka were never higher than today (Emiliani, 1966). Therefore, anchovies could have crossed this hypothetical barrier at any time, not being restricted to specific periods. However, anchovy dispersals might be favoured during cooling periods by the enhancement of tropical currents (Fratantoni *et al.*, 2000). These fish may have found more suitable conditions to cross the Atlantic Ocean and the Caribbean Sea after the LGM, more likely during the Dryas stadials (18–15 ka; 14–13.7 ka; 12.8–11.5 ka) (Roberts, 1998) or during cooling periods in the late Holocene, from after the Hypsithermal period to the Little Ice Age (*c*. 0.5 ka; Denton & Karlén, 1973).

Examples of trans-Atlantic dispersal events include west-to-east migrations (e.g. the rock hind *Epinephelus adscensionis*; Carlin *et al.*, 2003) and east-to-west dispersals observed in four tropical fish, *Acanthurus monroviae* and *Parablennius pilicornis* (Luiz-Júnior *et al.*, 2004), *Epinephelus marginatus* (Joyeux *et al.*, 2001) and *Aulostomus strigosus* (Bowen *et al.*, 2001). However, most of the colonization

routes proposed are fairly uncontroversial as these species have latitudinally narrower geographical distributions. For instance, *Clupea harengus*, with a geographical distribution restricted to the northern Atlantic Ocean, exhibits no genetic differentiation between the north-western Atlantic (Nova Scotia) and north-eastern Atlantic (North Sea) (Bekkevold *et al.*, 2005). Although no attempt was made to propose a colonization route to or from either side of the Atlantic, a hypothetical route for *C. harengus* would certainly include a northern route. The difference with anchovies is that although we would expect the colonization route to be mostly along the continental platform, the evidence points instead to a southern equatorial route.

CONCLUSIONS

The results from this study provide valuable insights into the dynamics of anchovy population expansions and dispersals in response to ocean-climatic fluctuations in the Atlantic Ocean. Alternating cooling and warming periods in the North Atlantic over the course of the Quaternary might have displaced the ranges of these populations and set the stage for dispersal and recolonization. We found similar allele frequencies on either side of the English Channel and no significant decrease in genetic diversity in populations north of the English Channel. As anchovies are coastal species, climatic fluctuations were expected to enhance latitudinal range movements, but not longitudinal migrations across the Atlantic. However, climatic oscillations affected water mass circulation and trade winds, and specific cooling periods may have favoured these trans-Atlantic migrations. Whereas pre-LGM trans-Atlantic colonizations cannot be excluded, results are more compatible with post-LGM colonizations. Colonizations of the European anchovy in the northern part of the East Atlantic are unlikely to reflect first-time occupations, but instead represent the last of a series of range displacement cycles that reflect ongoing gene flow with southern core populations that have prevented divergence. The case of European anchovies contributes to our knowledge of how climate cycles conditioned species range contractions and expansions and show that the genetic characteristics of marginal populations, relative to central populations, can be determined by the dispersal dynamics of species.

ACKNOWLEDGEMENTS

This work was only possible due to the collaboration of many colleagues in collecting and shipping samples: E. Torstensen (IMR, Norway), M. Wyszynski (SFI, Poland), S. Vaz, F. Coppin, Y. Verin and D. Leroy (IFREMER, France), I. Zarraonaindia (UPV/EHU, Spain), K. Erzini (CCMAR, Portugal), A. Dias (Portugal), J. Tornero and J. Alarcón (IEO, Spain), M. Santamaria (Canary Islands), A. Hervé (Senegal), L. Fonseca (Guinea-Bissau), H. Dankwa (WRI, Ghana), J. Vasslides and R. Hagan (RUMFS, USA) and E. Wiley and A. Bentley (KNHM, USA). F. Lima produced Figure 2.1b and 2.1c. We thank W. S. Grant for discussion and comment on earlier versions of this manuscript, N. Coelho and M. Valente for technical help, A. Yu for sharing unpublished microsatellites, J. Neiva, R. L. Cunha, S. Teixeira, K. Nicastro and G. Zardi for discussions, Nina O. Therkildsen for sharing the DAPC code, and Christine Maggs, M. Limborg and several anonymous referees for valuable comments and suggestions. This study was funded in part by the ESF MarinERA-NET-project/189570 'Marine phylogeographic structuring during climate change: the signature of leading and rear edge of range shifting populations'. G.S.'s research was supported by a doctoral grant from FCT-Portugal (SFRH/BD/36600/2007).

REFERENCES

Agostini, V.N. & Bakun, A. (2002) "Ocean Triads" in the Mediterranean Sea: physical mechanisms potentially structuring reproductive habitat suitability (example application to European anchovy, *Engraulis encrasicolus*). *Fisheries Oceanography*, **11**, 129–142.

Alheit, J., Pohlmann, T., Casini, M., Greve, W., Hinrichs, R., Mathis, M., O'Driscoll, K., Vorberg, R. & Wagner, C. (2012) Climate variability drives anchovies and sardines into the North and Baltic Seas. *Progress in Oceanography*, **96**, 128–139.

Arenas, M., Ray, N., Currat, M. & Excoffier, L. (2011) Consequences of range contractions and range shifts on molecular diversity. *Molecular Biology and Evolution*, **29**, 207–218.

Bekkevold, D., Andre, C., Dahlgren, T.G., Clausen, L.A.W., Torstensen, E., Mosegaard, H., Carvalho, G.R., Christensen, T.B., Norlinder, E. & Ruzzante, D.E. (2005) Environmental correlates of population differentiation in Atlantic herring. *Evolution*, **59**, 2656–2668.

Borrell, Y.J., Piñera, J.A., Sánchez Prado, J.A. & Blanco, G. (2012) Mitochondrial DNA and microsatellite genetic differentiation in the European anchovy *Engraulis encrasicolus* L. *ICES Journal of Marine Science: Journal du Conseil*, **69**, 1357–1371.

Bouchenak-Khelladi, Y., Durand, J.D., Magoulas, A. & Borsa, P. (2008) Geographic structure of European anchovy: a nuclear-DNA study. *Journal of Sea Research*, **59**, 269–278.

Bowen, B.W., Bass, A.L., Rocha, L.A., Grant, W.S. & Robertson, D.R. (2001) Phylogeography of the trumpetfishes (*Aulostomus*): ring species complex on a global scale. *Evolution*, **55**, 1029–1039.

Briers (2003) *VBA add-in Mantel.xla 1.2.4*. Available at: http://www.lifesciences.napier.ac.uk/staff/rob/software/index.html (accessed August 2012).

Carlin, J.L., Robertson, D.R. & Bowen, B.W. (2003) Ancient divergences and recent connections in two tropical Atlantic reef fishes *Epinephelus adscensionis* and *Rypticus saponaceous* (Percoidei: Serranidae). *Marine Biology*, **143**, 1057–1069.

Chapuis, M.-P. & Estoup, A. (2007) Microsatellite null alleles and estimation of population differentiation. *Molecular Biology and Evolution*, **24**, 621–631.

CLIMAP (1981) *Seasonal reconstruction of the Earth's surface at the last glacial maximum*. Geological Society of America Map Chart Series. Geological Society of America, Boulder, CO.

Denton, G.H. & Karlén, W. (1973) Holocene climatic variations – their pattern and possible cause. *Quaternary Research*, **3**, 155–205.

Drummond, A., Ashton, B., Buxton, S., Cheung, M., Cooper, A., Duran, C., Field, M., Heled, J., Kearse, M., Markowitz, S., Moir, R., Stones-Havas, S., Sturrock, S., Thierer, T. & Wilson, A. (2011) *Geneious v. 5.1*. Available at: http://www.geneious.com/.

Drummond, A.J., Suchard, M.A., Xie, D. & Rambaut, A. (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, **29**, 1969–1973.

Emiliani, C. (1966) Isotopic paleotemperatures. *Science*, **154**, 851–857.

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

Fratantoni, D.M., Johns, W.E., Townsend, T.L. & Hurlburt, H.E. (2000) Low-latitude circulation and mass transport pathways in a model of the tropical Atlantic Ocean. *Journal of Physical Oceanography*, **30**, 1944–1966.

Fu, Y.X. (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, **147**, 915–925.

Goudet, J. (1995) FSTAT version 1.2: a computer program to calculate F-statistics. *The Journal of Heredity*, **86**, 485–486.

Grant, W.S., Lecomte, F. & Bowen, B.W. (2010) Biogeographical contingency and the evolution of tropical anchovies (genus *Cetengraulis*) from temperate anchovies (genus *Engraulis*). *Journal of Biogeography*, **37**, 1352–1362.

Grant, W.S., Liu, M., Gao, T. & Yanagimoto, T. (2012) Limits of Bayesian skyline plot analysis of mtDNA sequences to infer historical demographies in Pacific herring (and other species). *Molecular Phylogenetics and Evolution*, **65**, 203–212.

Harpending, H. (1994) Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. *Human Biology*, **66**, 591–600.

Hedrick, P.W. & Goodnight, C. (2005) A standardized genetic differentiation measure. *Evolution*, **59**, 1633–1638.

Jombart, T. (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics,* **24**, 1403–1405.

Jombart, T. (2013) *An introduction to adegenet 1.4-0*. Available at: ftp://ie.freshrpms.net/pub/CRAN/web/packages/adegenet/vignettes/adegenet-basics.pdf.

Jombart ,T., Devillard, S., Dufour, A.B. & Pontier, D. (2008) Revealing cryptic spatial patterns in genetic variability by a new multivariate method. *Heredity,* **101**, 92–103.

Jost, L. (2008) G_{ST} and its relatives do not measure differentiation. *Molecular Ecology*, **17**, 4015–4026.

Joyeux, J.C., Floeter, S.R., Ferreira, C.E.L. & Gasparini, J.L. (2001) Biogeography of tropical reef fishes: the South Atlantic puzzle. *Journal of Biogeography*, **28**, 831–841.

Keenan, K., McGinnity, P., Cross, T.F., Crozier, W.W. & Prodöhl, P.A. (2013) diveRsity: an R package for the estimation and exploration of population genetics parameters and their associated errors. *Methods in Ecology and Evolution*, **4**, 782–788.

Kelly, R.P., Oliver, T.A., Sivasundar, A. & Palumbi, S.R. (2010) A method for detecting population genetic structure in diverse, high gene-flow species. *Journal of Heredity*, **101**, 423–436.

Lambeck, K., Esat, T.M. & Potter, E.K. (2002) Links between climate and sea levels for the past three million years. *Nature*, **419**, 199–206.

Lessios, H.A. (2008) The great American schism: divergence of marine organisms after the rise of the Central American Isthmus. *Annual Review of Ecology, Evolution, and Systematics*, **39**, 63–91.

Librado, P. & Rozas, J. (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25**, 1451–1452.

Limborg, M., Hanel, R., Debes, P., Ring, A., André, C., Tsigenopoulos, C. & Bekkevold, D. (2012) Imprints from genetic drift and mutation imply relative divergence times across marine transition zones in a pan-European small pelagic fish (*Sprattus sprattus*). *Heredity*, **109**, 96–107.

Luiz-Júnior, O.J., Floeter, S.R., Gasparini, J.L., Ferreira, C.E.L. & Wirtz, P. (2004) The occurrence of *Acanthurus monroviae* (Perciformes: Acanthuridae) in the south-western Atlantic, with comments on other eastern Atlantic reef fishes occurring in Brazil. *Journal of Fish Biology*, **65**, 1173–1179.

Maggs, C.A., Castilho, R., Foltz, D., Henzler, C., Jolly, M.T., Kelly, J., Olsen, J., Perez, K.E., Stam, W., Vainola, R., Viard, F. & Wares, J. (2008) Evaluating signatures of glacial refugia for North Atlantic benthic marine taxa. *Ecology*, **89**, S108–S122.

Magoulas, A., Tsimenides, N. & Zouros, E. (1996) Mitochondrial DNA phylogeny and the reconstruction of the population history of a species: the case of the European anchovy (*Engraulis encrasicolus*). *Molecular Biology and Evolution*, **13**, 178–190.

Magoulas, A., Castilho, R., Caetano, S., Marcato, S. & Patarnello, T. (2006) Mitochondrial DNA reveals a mosaic pattern of phylogeographical structure in Atlantic and Mediterranean populations of anchovy (*Engraulis encrasicolus*). *Molecular Phylogenetics and Evolution*, **3**, 734–746.

McInerny, G.J., Turner, J.R.G., Wong, H.Y., Travis, J.M.J. & Benton, T.G. (2009) How range shifts induced by climate change affect neutral evolution. *Proceedings of the Royal Society B: Biological Sciences*, **276**, 1527–1534.

van der Meer, M.H., Jones, G.P., Hobbs, J.P.A. & van Herwerden, L. (2012) Historic hybridization and introgression between two iconic Australian anemonefish and contemporary patterns of population connectivity. *Ecology and Evolution*, **2**, 1592–1604.

Molfetti, É., Torres Vilaça, S., Georges, J.-Y., Plot, V., Delcroix, E., Le Scao, R., Lavergne, A., Barrioz, S., dos Santos, F.R. & de Thoisy, B. (2013) Recent demographic history and present fine-scale structure in the northwest Atlantic leatherback (*Dermochelys coriacea*) turtle population. *PLoS ONE*, **8**, e58061.

Motos, L. (1996) Reproductive biology and fecundity of the Bay of Biscay anchovy population (*Engraulis encrasicolus* L.). *Scientia Marina*, **60**, 195–207.

van Oosterhout, C., Hutchinson, W.F., Wills, D.P.M. & Shipley, P. (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, **4**, 535 – 538.

Parada, C., van der Lingen, C.D., Mullon, C. & Penven, P. (2003) Modelling the effect of buoyancy on the transport of anchovy (*Engraulis capensis*) eggs from spawning to nursery grounds in the southern Benguela: an IBM approach. *Fisheries Oceanography*, **12**, 170–184.

Pennings, P.S., Achenbach, A. & Foitzik, S. (2011) Similar evolutionary potentials in an obligate ant parasite and its two host species. *Journal of Evolutionary Biology*, **24**, 871–886.

Petitgas, P., Alheit, J., Peck, M., Raab, K., Irigoien, X., Huret, M., van der Kooij, J., Pohlmann, T., Wagner, C., Zarraonaindia, I. & Dickey-Collas, M. (2012) Anchovy population expansion in the North Sea. *Marine Ecology Progress Series*, **444**, 1–13.

Posada, D. (2004) *Collapse 1.2: describing haplotypes from sequence alignments*. Available at: http://darwin.uvigo.es/software/collapse.html (accessed August 2012).

Pritchard, J.K., Stephens, M. & Donnelly, P. (2000) Inference of population structure from multilocus genotype data. *Genetics*, **155**, 945–959.

Rambaut, A. & Drummond, A.J. (2007) *Tracer 1.5*. Available at: http://beast.bio.ed.ac.uk/Tracer.

Ramos-Onsins, S.E. & Rozas, J. (2002) Statistical properties of new neutrality tests against population growth. *Molecular Biology and Evolution*, **19**, 2092–2100.

Renssen, H. & Vandenberghe, J. (2003) Investigation of the relationship between permafrost distribution in NW Europe and extensive winter sea-ice cover in the North Atlantic Ocean during the cold phases of the Last Glaciation. *Quaternary Science Reviews*, **22**, 209–223.

Roberts, N. (1998) *The Holocene: an environmental history*. John Wiley & Sons, Malden, MA.

Rogers, A. & Harpending, H. (1992) Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution*, **9**, 552–569.

Salicru, M., Menendez, M.L., Morales, D. & Pardo, L. (1993) Asymptotic distribution of (h, φ)-entropies. *Communications in Statistics – Theory and Methods*, **22**, 2015–2031.

Sambrook, J. & Russell, D. (2001) *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory Press, New York.

Sanz, N., Garcia-Marín, J.L., Viñas, J., Roldán, M. & Pla, C. (2008) Spawning groups of European anchovy: population structure and management implications. *ICES Journal of Marine Science: Journal du Conseil*, **65**, 1635–1644.

Teacher, A.G.F. & Griffiths, D.J. (2011) HapStar: automated haplotype network layout and visualization. *Molecular Ecology Resources*, **11**, 151–153.

Toonen, R. & Hughes, S. (2001) Increased throughput for fragment analysis on ABI Prism 377 automated sequencer using a membrane comb and STRand software. *Biotechniques*, **31**, 1320–1324

Urtizberea, A., Fiksen, Ø., Folkvord, A. & Irigoien, X. (2008) Modelling growth of larval anchovies including diel feeding patterns, temperature and body size. *Journal of Plankton Research*, **30**, 1369–1383.

Vermeij, G.J. (2005) From Europe to America: Pliocene to recent trans-Atlantic expansion of cold-water North Atlantic molluscs. *Proceedings of the Royal Society B: Biological Sciences*, **272**, 2545–2550.

de Vernal, A. & Hillaire-Marcel, C. (2006) Provincialism in trends and high frequency changes in the northwest North Atlantic during the Holocene. *Global and Planetary Change*, **54**, 263–290.

Wares, J.P. & Cunningham, C.W. (2001) Phylogeography and historical ecology of the North Atlantic intertidal. *Evolution*, **55**, 2455–2469.

Weir, B.S. (1996) *Genetic data analysis II*. Sinauer Associates, Sinauer Associates, Sunderland, MA.

Whitehead, P.J., Nelson, W.S. & Wongratana, T. (1988) *Clupeoid fishes of the world (suborder Clupeoidei)*. Food and Agriculture Organization of the United Nations, Rome.

Zarraonaindia, I., Pardo, M.A., Iriondo, M., Manzano, C. & Estonba, A. (2009) Microsatellite variability in European anchovy (*Engraulis encrasicolus*) calls for further investigation of its genetic structure and biogeography. *ICES Journal of Marine Science: Journal du Conseil*, **66**, 2176–2182.

Zarraonaindia, I., Iriondo, M., Albaina, A., Pardo, M.A., Manzano, C., Grant, W.S., Irigoien, X. & Estonba, A. (2012) Multiple SNP markers reveal fine-scale population and deep phylogeographic structure in European anchovy (*Engraulis encrasicolus* L.). *PLoS ONE*, **7**, e42201.

SUPPORTING INFORMATION

Anchovies go north and west without losing diversity: post-glacial range expansions in a small pelagic fish

Gonçalo Silva, John B. Horne and Rita Castilho

Appendix S1 Primer sequences, polymerase chain reaction conditions, sources and summary statistics of molecular markers.

Loci information									Genetic diversity				
Amplified region	**Name (acc. no.)**	**Motif**	**Primers 5' → 3'**	**Reference**	**Cycles (*n*)**	***Ta* (°C)**	***At* (s)**	**Mg (mM)**	**Ar**	**n_a**	**H_O**	**H_E**	**G_{IS}**
mtDNA	**Cyt *b***	-	F: CCTACTCAAGATCGCTAACGA R: AGTTTAACTTTAGAATGCTAGCTTTGG	(Arnaud-Haond, unpublished) (modified from Inoue *et al.*, 2001)	40	48	60	2	1045	226	-	-	-
nuclear microsatellites	**Ee 10 (AY241273)**	[(GT)9CT]2[(GT)2CT]3	F#: **VIC**- GGTGGATGAAGTGGCAATCT R: CTGGGGTGGCATAACTGAAG	(Landi *et al.*, 2005)	35	57	30	2	197-319	72	0.842	0.877	0.014
	Ee2-135 (FJ534738)	(ATTAG)10	F#: **PET**-AGGGCAGTGACAGGAGAGTC R: TCGTTACCCTGCGTTTATACTG	(Pakaki *et al.*, 2009)	40	50	30	2	113-170	14	0.756*	0.845	0.072*
	Ee2-407 (FJ534751)	(CA)13	F#: **VIC**-AGGAATCTCCTTCCCGTCTC R: GTGGGTCTGTGGGTGTTTTG		35	57	30	2	137-293	55	0.828*	0.863	0.042*
	Ee2-452a (FJ534754)	(AC)13	F#: **FAM**-CCCAACCCTAGGGAGACATC R: TCGTTCAGCAAGCATACACC		35	57	30	2	248-344	48	0.893	0.926	0.018
	Ee2-508 (FJ534759)	(AGG)8	F#: **VIC**-CACATGCTCGCTAAACATTG R: ACCTGATGCTGCTTGGTAGC		35	52	30	2	154-198	26	0.601	0.871	-0.040
	Ee2-91a (FJ534732)	(AGG)12	F#: **NED**-AGAGCAGGTTCTTGCTGTGG R: TGTGGTGCGCTACTATCAGG		35	55	30	2	230-290	28	0.628	0.866	-0.045
	Ee2-91b (FJ534732)	(CCGCA)8	F#: **PET**-GGTCTTGAGCTTGGCATAGG R: CCGGAAGACACTCTGCACAC		35	55	30	2	100-170	17	0.807	0.798	-0.018
	EJ41.1 (AF344659)	(CACAA)8	F#: **PET**-TCTACCCCTGGAGGACACAC R: ACAGGGGGTTGAGAAAGAGG	(Chiu *et al.*, 2002)	35	54	30	1.5	139-273	57	0.604*	0.819	0.007*
	EJ41.2	(TCTA)30	F#: **FAM**-GCACTGACCTCTTTCTCAAC R: AAATATGGTGGTTCATCTCG	(Chiu et al., unpublished)	35	54	30	2	156-381	152	0.617*	0.977	0.071*

#, fluorescence dye; *Ta*, annealing temperature; *At*, annealing time; Mg, Magnesium concentration; *Ar*, allele range (bp); n_a, number of alleles; H_O, observed heterozygosity; H_E, expected heterozygosity; G_{IS}, inbreeding coefficient; *, significant ($P < 0.05$)

REFERENCES

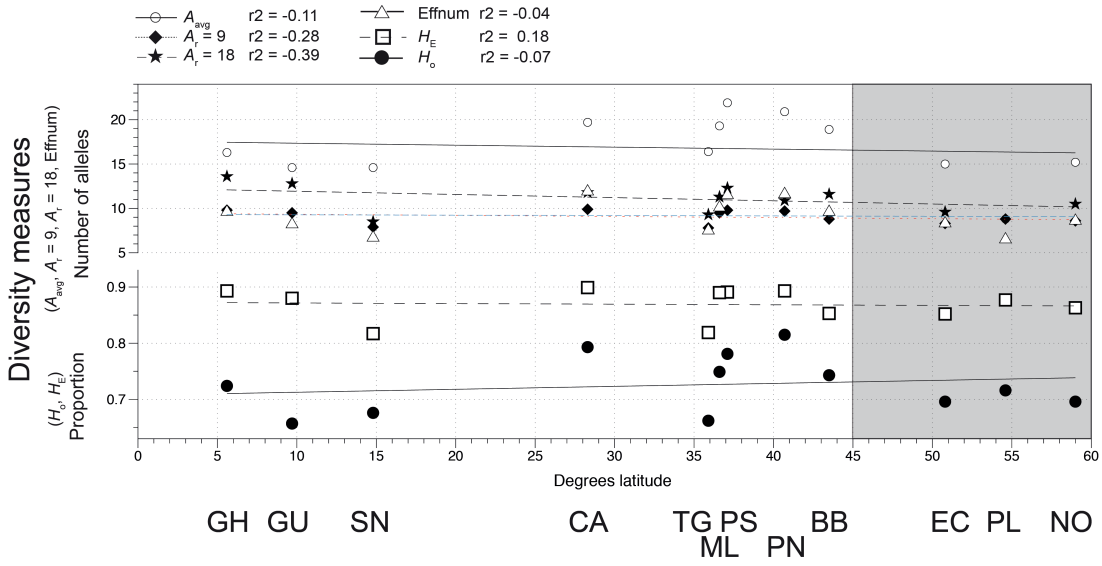
Chiu, T.S., Lee, Y.J., Huang, S.W. & Yu, H.T. (2002) Polymorphic microsatellite markers for stock identification in Japanese anchovy (*Engraulis japonicus*). *Molecular Ecology Notes*, **2**, 49–50.

Inoue, J.G., Miya, M., Tsukamoto, K. & Nishida, M. (2001) Complete mitochondrial DNA sequence of the Japanese anchovy *Engraulis japonicus*. *Fisheries Science*, **67**, 828–835.

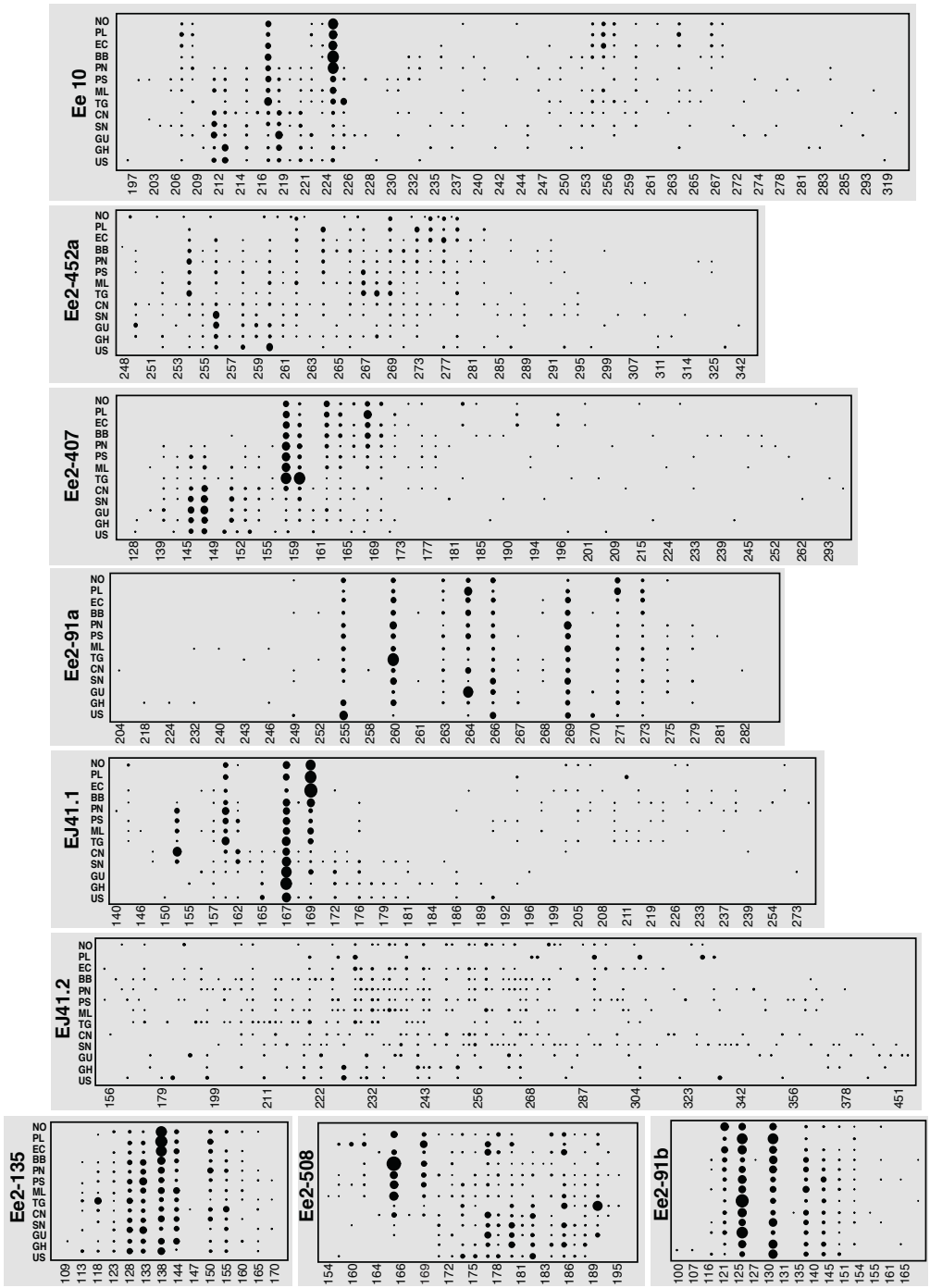
Landi, M., Garoia, F., Piccinetti, C. & Tinti, F. (2005) Isolation of polymorphic microsatellite loci from the European anchovy, *Engraulis encrasicolus*. *Molecular Ecology Notes*, **5**, 266–268.

Pakaki, V., Magoulas, A. & Kasapidis, P. (2009) New polymorphic microsatellite loci for population studies in the European anchovy, *Engraulis encrasicolus* (L.). *Molecular Ecology Resources*, **9**, 1406–1409.

Appendix S2 Plot of diversity measures of *Engraulis encrasicolus* along the latitudinal gradient. Average number of alleles (A_{avg}), allelic richness for $n = 9$ ($A_r = 9$), allelic richness for $n = 18$ ($A_r = 18$) (removing Poland), effective number of alleles (*Effnum*), observed and expected heterozygosity (H_O and H_E, respectively). Sample abbreviations are defined in Table 2.1.



Appendix S3 *Engraulis encrasicolus* and *E. eurystole* microsatellite allele frequencies for each locus; the diameters of the circles are proportional to allele frequencies. Sample abbreviations are defined in Table 2.1.



CHAPTER III ● THERMAL ADAPTATION AND CLINAL MTDNA VARIATION OF THE EUROPEAN ANCHOVY

Gonçalo Silva[1], Fernando P. Lima[2], Paulo Martel[3] and Rita Castilho[1]*

[1] Centro de Ciências do Mar (CCMAR, CIMAR - Associate Laboratory), Universidade do Algarve, Campus de Gambelas, 8005-139 Faro, Portugal; [2] CIBIO, Research Center in Biodiversity and Genetic Resources, University of Porto, Campus Agrário de Vairão, 4485-661 Vairão, Portugal; [3] Centro de Biomedicina Molecular e Estrutural Instituto de Biotecnologia e Bioengenharia (CBME - Associate Laboratory), Universidade do Algarve, Campus de Gambelas, 8005-139 Faro, Portugal.

*Corresponding author at: Centro de Ciências do Mar (CCMAR, CIMAR-Associate Laboratory), Universidade do Algarve, Campus de Gambelas, 8005-139 Faro, Portugal. Fax: +351 289 800 069.

E-mail address: rcastil@ualg.pt

Submitted to *Proceedings of the Royal Society B* - in revision

ABSTRACT

Natural populations of widely distributed organisms often exhibit genetic clinal variation over their geographical ranges. The European anchovy, *Engraulis encrasicolus*, illustrates this by displaying a two-clade mitochondrial structure clinally arranged along the eastern Atlantic. One clade has low frequencies at higher latitudes, while the other has an anti-tropical distribution, with frequencies decreasing towards the tropics. The distribution pattern of these clades has been explained as a consequence of secondary contact after an ancient geographic isolation. However, it is not unlikely that selection acts on mitochondria whose genes are involved in relevant oxidative phosphorylation processes. In this study, we performed selection tests on a fragment of 1044 bp of the mitochondrial cytochrome *b* gene using 455 individuals from 18 locations. We also tested correlations of six environmental features, temperature, salinity, apparent oxygen utilisation and nutrient concentrations of phosphate, nitrate and silicate, on a compilation of mitochondrial clade frequencies from 62 sampling sites comprising 2782 specimens from previously published studies. Positive selection in a single codon was detected only in the anti-tropical clade and temperature was the most relevant environmental predictor, contributing with 57% of the variance in the geographic distribution of clade frequencies. These findings strongly suggest that temperature is shaping the contemporary distribution of mtDNA clade frequencies in the European anchovy.

Keywords: Positive selection, *Engraulis encrasicolus*, adaptation, sea temperature, high metabolism, mitochondrial cytochrome *b* gene

INTRODUCTION

"I have called the principle, by which each slight variation, if useful, is preserved by the term of Natural Selection." (Darwin, 1859)

Mitochondrial DNA (mtDNA) has been widely used in evolutionary biology research over the past 20 years under the implicit assumption of neutrality (Avise, 2000). However, there is strong evidence that this molecule may be under positive selection, often related to thermal adaptation and aerobic capacity (Galtier *et al.*, 2009, and references therein). The assumption that mtDNA polymorphisms are neutral has been tested in the historical demographic context, but rarely these tests have been taken further. Genetic variation affected by selection and not chiefly by demography can compromise mtDNA markers usefulness to correctly estimate demographic changes, population structure or to date biogeographic events. In this event, molecular markers under selection may be useful to understand the processes that shape species distribution patterns and local adaptation.

Mitochondrial genes are involved in oxidative phosphorylation processes (OXPHOS complex) by which means the electron transport chain (ETC) creates a trans-membrane proton gradient that generates ATP (Adenosine Triphosphate - ATP) (Mitchell, 2011). The ETC is formed by protein complexes of subunits that are encoded in either nuclear or mitochondrial DNA. Non-synonymous single nucleotide polymorphisms in any of the genes encoding ETC subunits can potentially affect the quality of electron flow or influence other relevant binding sites, such as that of coenzyme Q or CoQ (Beckstead *et al.*, 2009). It is therefore plausible that non-synonymous changes in the mtDNA will impact the fitness of organisms given the pivotal role of mitochondrial bioenergetics on adaptation to environmental variability (Gershoni *et al.*, 2009).

An increasing number of studies have detected positive selective sweeps in the mitochondria, including adaptation to extreme O_2 requirements of flying capacity in bats (Shen *et al.*, 2010), low energy diet in large body mammals (da Fonseca *et al.*, 2008), high altitude resistance in alpacas and monkeys (Hochachka *et al.*, 1983; da Fonseca *et*

al., 2008; Yu *et al.*, 2011) and climate-mediated adaption in humans (Mishmar *et al.*, 2003; Ruiz-Pesini *et al.*, 2004; Balloux *et al.*, 2009). Although there are few studies of mtDNA selection in marine fish, selection in mitochondria has been invoked to explain patterns of genetic variation in the slippery-dick labrid (*Halichoeres bivittatus*) (Haney *et al.*, 2010), the association between the distribution of mitochondrial lineages and sea surface temperature in the walleye pollock (*Theragra chalcogramma*) (Grant *et al.*, 2006) and the cause of mito-nuclear co-evolution that increases aerobic capacity and swimming performances in billfishes (Xiphiidae and Istiophoridae families; (Dalziel *et al.*, 2006)). Selection is also suspected to have promoted amino acid changes in proton pumping that influenced fitness in Pacific salmon species (Garvin *et al.*, 2011) and the Atlantic herring (*Clupea harengus*) (Teacher *et al.*, 2012).

The European anchovy provides an ideal system to investigate adaptive selection. It is distributed throughout tropical, subtropical and cold-temperate coastal areas (ca. 60 ºN - 40 ºS), facing contrasting environmental features, which implies an impressive tolerance to a broad range of temperatures (2 − 30 ºC) and salinities (5 – 41‰). This species also shows both morphological and genetic variability across its distributional area, displaying a dual-clade mitochondrial structure, arranged into a clinal frequency in the eastern Atlantic (Silva *et al.*, *in press*). Clade A is present throughout the whole geographic distribution, but with lower frequencies at higher latitudes, while clade B has an anti-tropical distribution, with frequencies decreasing towards the tropics. This structure may reflect post-glacial secondary contact after an ancient isolation (Magoulas *et al.*, 1996). Nevertheless, one cannot exclude the relevance of other processes such as sex-biased dispersal, nuclear allelic convergence, incomplete mtDNA lineage sorting, adaptive introgression, demographic disparities, gamete incompatibility or, as considered in the present work, adaptive selection (Grant, 2005; Kristoffersen & Magoulas, 2008) in promoting the observed genetic divergence.

Recently, a mito-genomic survey of a widely distributed marine mammal, the killer whale, showed high levels of amino-acid conservation and only two positively selected codons, both in the cytochrome *b* (cyt *b*) gene, correlated with temperature adaptation (Foote *et al.*, 2010). Here, we focus on this gene to explore a putative instance of positive

selection shaping the distribution of the two European anchovy *Engraulis encrasicolus* genetic lineages. We posit that the cyt *b* of the European anchovy may be potentially affected by positive selective regimes that influence metabolism and constrain the distribution of mtDNA clades. We expect to detect non-synonymous substitutions that would provide selective advantage to one of the clades, possibly altering the function of the protein, promoting a better adaptation to local environment. We correlate the present-day distribution of mtDNA lineages of anchovies with various environmental factors. Our study is part of an emerging effort to better understand the role of natural selection in shaping the geographical distribution of genetic variation of organisms and their adaptation to changing environments.

MATERIAL AND METHODS

SAMPLES COLLECTION, DNA EXTRACTION AND AMPLIFICATION AND SEQUENCE ALIGNMENT

We collected 455 specimens of the European anchovy *E. encrasicolus*, from 18 locations, from Norway to South Africa and the Mediterranean, covering most of the geographic distribution of the species, with the exception of the western Atlantic, the Baltic and Black seas (figure 3.1 and table S1, electronic supplementary material). We chose not to include samples from the Baltic and the Black Seas and from the western Atlantic because the genetic pool of the individuals from these regions is influenced either by the extremely enclosed geomorphology of the sea basins or transatlantic currents (Grant, 2005; Zarraonaindia *et al.*, 2012).

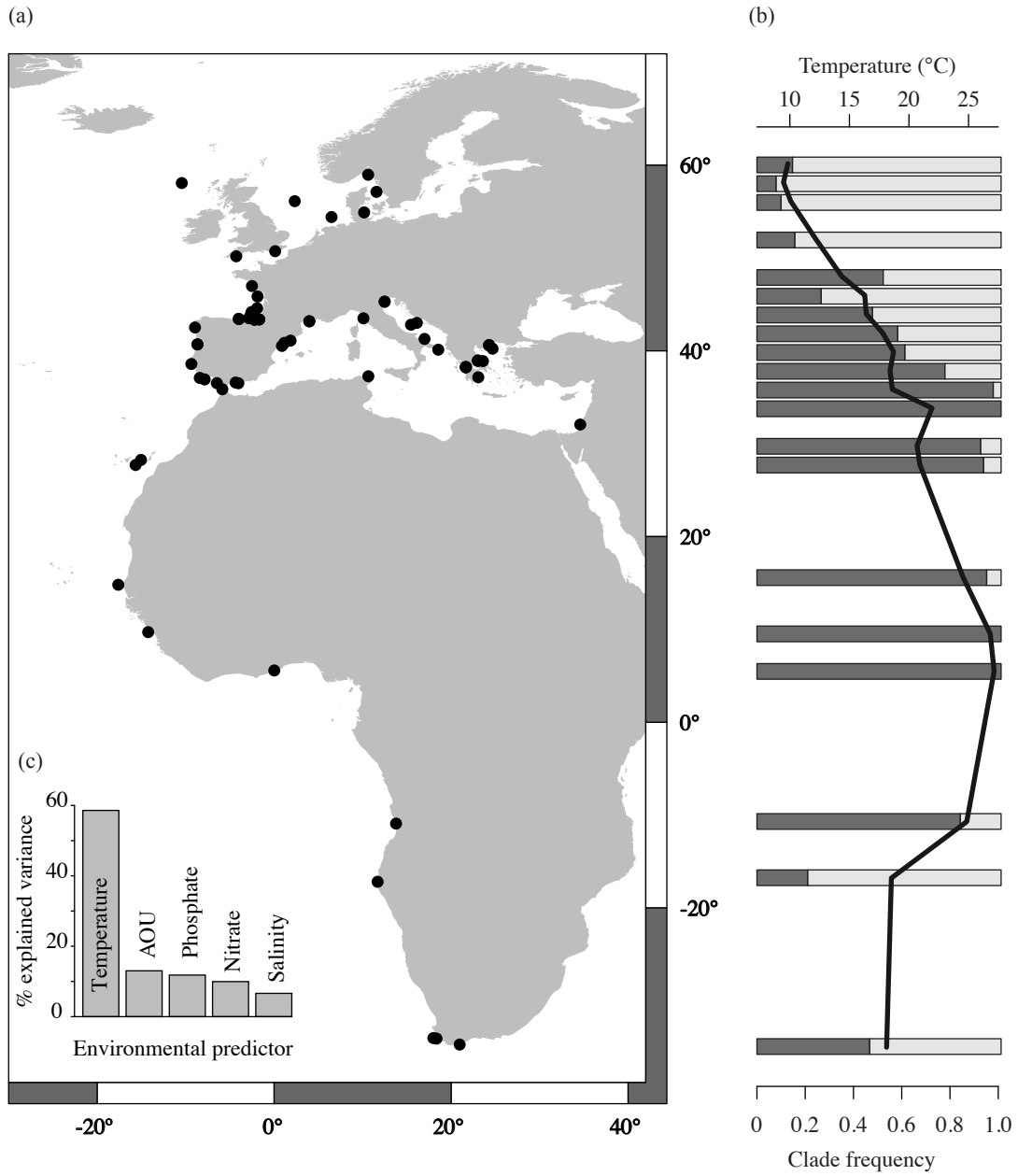


Figure 3.1. (a) Locations used for environmental correlates of mitochondrial clades frequencies; (b) Climatological sea temperature from World Ocean Atlas 2009 between 0 and 10m (black line) and clade frequency (clade A: dark grey, clade B: light grey); (c) Environmental variables importance after hierarchical partitioning analysis; AOU – Apparent Oxygen Utilisation.

Table 3.1. Sample information. Sampling points = number of independent sampling points in each region (for complete information see Supporting Information, table S1); when more than a sampling point is present, an approximate location is indicated.

basin	country	region	code in map	longitude	latitude	N	clade A	clade B	sampling points
Atlantic	Norway	North Sea	1	10.63	58.98	24	2	22	1
Atlantic	Denmark	North Sea	2	11.57	57.14	15	1	14	1
Atlantic	Scotland	North Sea	3	-4.05	57.12	35	6	29	2
Atlantic	Germany	North Sea	4	8.33	54.68	39	4	35	2
Atlantic	France	English Channel	5	0.12	50.75	27	3	24	1
Atlantic	UK	English Channel	6	-4.28	50.20	25	5	20	1
Atlantic	France	Bay of Biscay	7	-1.95	44.60	126	55	71	3
Atlantic	Spain	Bay of Biscay	8	-2.75	44.06	210	115	95	8
Atlantic	France	Gulf of Lion	9	3.98	43.20	50	21	29	1
Atlantic	France	Northeastern Atlantic	10	3.98	43.20	22	5	17	1
Atlantic	Spain	Galicia	11	-8.94	42.53	29	26	3	1
Atlantic	Portugal	Northeastern Atlantic	12	-8.58	38.81	217	188	29	5
Atlantic	Morocco	Saharan upwelling	13	-5.85	35.88	62	60	2	1
Atlantic	Spain	Canaries	14	-15.35	28.00	76	70	6	2
Atlantic	Senegal	West Africa	15	-17.61	14.82	34	32	2	1
Atlantic	Guinea-Bissau	West Africa	16	-14.23	9.72	20	20	0	1
Atlantic	Ghana	Gulf of Guinea	21	0.02	5.59	25	25	0	1
Atlantic	Angola	Benguela	17	13.78	-10.90	24	20	4	1
Atlantic	Namibia	Benguela	18	11.70	-17.17	24	5	19	1
Atlantic	South Africa	Agulhas	19	19.11	-34.25	59	24	35	3
Mediterranean	Croatia	Adriatic Sea	20	16.13	43.03	20	5	15	1
Mediterranean	Greece	Aegean Sea	22	23.63	39.31	606	469	137	8
Mediterranean	Greece	Ionian Sea	23	21.67	38.26	361	129	232	4
Mediterranean	Israel	Levantine Sea	24	34.59	32.08	26	26	0	1
Mediterranean	Italy	Adriatic Sea	25	15.20	42.98	282	38	244	5
Mediterranean	Italy	Ligurian Sea	26	10.09	43.52	55	27	28	1
Mediterranean	Spain	Balearic Sea	27	1.31	40.85	86	38	48	3
Mediterranean	Spain	Alboran Sea	28	-4.20	36.57	115	106	9	2
Mediterranean	Spain	Gulf of Cadiz	29	-6.47	36.54	60	55	5	1
Mediterranean	Tunisia	Gulf of Tunis	30	10.65	37.28	28	16	12	1
	Grand Total					2782	1596	1186	65

Fish were purchased at small coastal fish markets, as artisanal fisherman do not venture far, to ensure the correct origin of fish, or were collected on scientific cruises (see acknowledgements). A small portion of white muscle or fin tissue were preserved in 96% ethanol and stored at -20ºC. DNA extraction, polymerase chain reaction (PCR), purification and sequencing were performed for a 1044 bp fragment of the mitochondrial cyt *b* as described in Silva *et al.* (Silva *et al.*, *in press*). Sequences were deposited in GenBank (see Data Accessibility section). The sequences were aligned using Clustal W (Thompson *et al.*, 1994) and visually inspected in Geneious 5.4 (Drummond *et al.*, 2011).

TESTS OF RECOMBINATION AND SELECTION

We tested the alignment for evidence of mitochondrial cyt *b* recombinants using GARD (Genetic Algorithms for Recombination Detection) analysis (Kosakovsky Pond *et al.*, 2006a) implemented in the online interface www.datamonkey.org (Delport *et al.*, 2010). To assess if selection was acting on cyt *b*, the Z-test (Nei & Gojobori, 1986) was performed in Mega 5 (Tamura *et al.*, 2011). We further implemented likelihood and Bayesian-based methods to identify site-specific cyt *b* positive selection where the rate of non-synonymous substitution (d*N*) is greater than the rate of synonymous substitution (d*S*). We applied SLAC (single likelihood ancestor counting), FEL (fixed effects likelihood) (Kosakovsky Pond & Frost, 2005), IFEL (internal fixed effects likelihood) (Kosakovsky Pond *et al.*, 2006b), FUBAR (fast unconstrained Bayesian approximation) (Murrell *et al.*, 2013) and MEME (mixed effects model of evolution) (Murrell *et al.*, 2012) to our data. Simulation suggests that MEME and FUBAR are substantially more powerful and equally accurate than the other methods (Murrell *et al.*, 2012; Murrell *et al.*, 2013). These methods are generally biased against detecting positive selection in conservative gene sequences, even when single amino acid changes can turn out to be adaptive. We applied all these methods to prevent against our results being an artefact of a particular methodology or a set of assumptions.

BIOCHEMICAL SOURCES OF INTRINSIC VARIATION

The cyt *b* dataset was aligned to a reference sequence available on GenBank (ACC number: NC_009581). Additionally we used TreeSAAP (Woolley *et al.*, 2003) to categorize 539 biochemical/structural physico-chemical changes owing to amino acid replacements into eight magnitude categories and determine whether the observed magnitude of amino acid changes deviates significantly from neutral expectations. We run analyses accordingly to Woolley *et al.* (2003). We considered only amino acid replacements with significant magnitude categories 6–8 ($p < 0.001$). The crystallographic structure of the cytochrome *bc1* complex interacting with cytochrome *c* was taken from the Protein Databank, PDB 3CX5 (Berman *et al.*, 2000). The homology model for the *E. encrasicolus* structure, based on a sequence variant with a Methionine residue at position 368 (the yeast sequence has a Valine residue at the homologous position), was obtained from the ModBase database of homology models (Pieper *et al.*, 2002).

ENVIRONMENTAL CORRELATES OF MITOCHONDRIAL CLADE FREQUENCIES

We compiled *E. encrasicolus* mitochondrial clade frequencies of 62 sampling sites comprising 2782 specimens from previous studies (Magoulas *et al.*, 1996; Grant *et al.*, 2005; Magoulas *et al.*, 2006; Bouchenak-Khelladi *et al.*, 2008; Borrell *et al.*, 2012; Zarraonaindia *et al.*, 2012; Silva *et al.*, *in press*) (table S1, electronic supplementary material) and used general linear models (GLM) of the binomial family (logit) to evaluate the correlation between clade frequencies and a variety of environmental variables. Data on temperature, salinity, apparent oxygen utilisation and nutrient concentrations (phosphate, nitrate and silicate) for depths < 10m were obtained from the World Ocean Atlas 2009 one-degree objectively analysed climatology data sets (Boyer *et al.*, 2009) in NetCDF format and imported as geo-referenced layers into R 2.15.3 (R Development Core Team, 2013) using the *ncdf* (Pierce, 2011) and *raster* (Hijmans, 2013) packages. The package *hier.part* (Walsh & Nally, 2013) was used to quantify the independent correlation of each predictor variable with the clade frequency, a method called hierarchical partitioning (Chevan & Sutherland, 1991; Nally, 1996).

RESULTS

RECOMBINATION AND SELECTION

From the 455 individuals analysed for the cyt *b* fragment, 246 polymorphic sites yielded 316 haplotypes, 8 amino acid variable sites and 11 amino acid type sequences (figure 3.2).

79	96	125	192	303	329	368	NA	NB	N
I	A	M	V	V	V	V	279	91	370
M	A	M	V	V	V	V	1	0	1
I	A	M	V	V	I	V	1	0	1
I	A	M	A	V	V	V	1	0	1
I	T	M	V	V	V	V	1	0	1
I	A	M	V	L	V	V	0	2	2
I	A	V	V	V	V	V	0	1	1
I	A	M	V	V	V	A	1	1	2
I	A	M	V	V	V	M	1	63	64
I	A	V	V	V	V	M	0	1	1
I	A	M	V	I	V	M	0	1	1

Figure 3.2. Amino acid substitutions in the mitochondrial fragment of cytochrome *b* of *Engraulis encrasicolus*. NA: number of clade A individuals; NB: number of clade B individuals; N: total number of individuals. Colours are meant to evidence differences and similarities between amino acid positions. Box indicates amino acid change under selection.

No evidence for recombination was found with GARD. Positive selection over all sites was detected on cyt *b* (Clade A: −5.81; $p = 1$; Clade B: −8.75; $p = 1$). The amino acid changes under selection were in a total of four, three under purifying selection and one under positive selection (codon 368) (table 3.2). However, the significant positive selected site located in the 9th trans-membrane helix of the cyt *b*, was detected only in clade B by with FUBAR, IFEL and FEL methods. Codon 368 is present in 64

individuals, 63 of which belonging to clade B. This codon presents in clade B a mutation of a Valine or an Alanine into a Methionine (figure 3.2).

Table 3.2. Positively and negatively selected sites in cytochrome *b* gene estimated by FUBAR, SLAC, IFEL, FEL and MEME models (*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$).

alignment position		278	302	812	1007
Codon		125	133	303	368
selection type		purifying	purifying	purifying	positive
all	FUBAR	*	***		*
	SLAC	**	*		
	IFEL	*			*
	FEL	*	**		**
	MEME				
clade A	FUBAR	***	***	**	
	SLAC	***	*		
	IFEL				
	FEL	***	**	*	
	MEME				
clade B	FUBAR				****
	SLAC				**
	IFEL				***
	FEL				****
	MEME				***

ADAPTATION AT THE MOLECULAR LEVEL

The cyt *b* sequences of yeast *S. cerevisae* and *E. encrasicolus* were aligned, showing 50% identity. Alignment of the model with the cyt *b* monomer in the 3CX5 produced an RMSD of 0.7 Å and allowed identification of the yeast structural homologue of Met 368 in the *Engraulis* sequence (Valine 369 in yeast) (figure 3.3). TreeSAAP identified 15 significant physico-chemical properties potentially influenced by positive selection in codon 368 (table S2, electronic supplementary material).

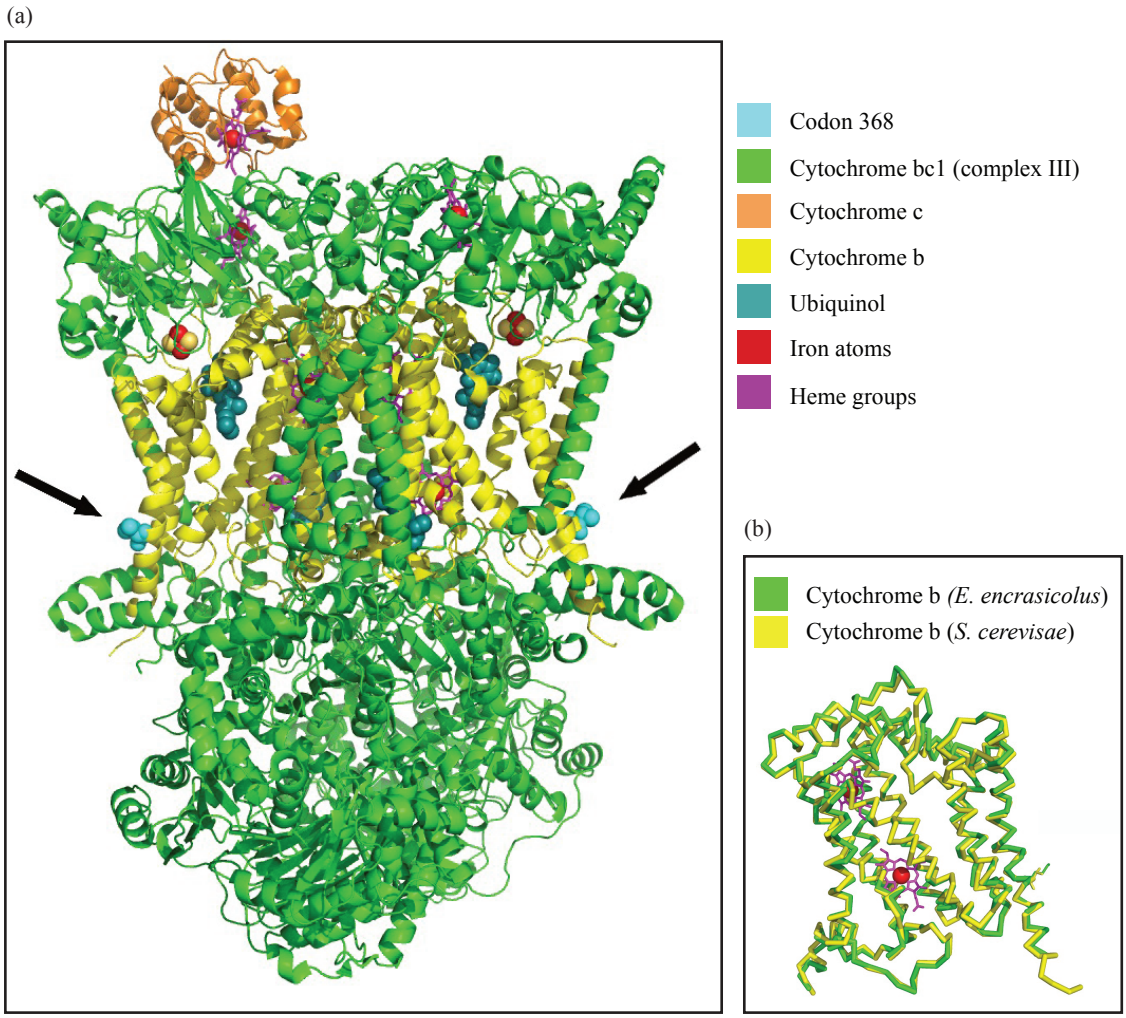


Figure 3.3. (a) Crystal structure of yeast (*Saccharomyces cerevisae*) cytochrome *bc1* (complex III) complexed with cytochrome *c*. The black arrows point to the two V369 residues (368 in the *Engraulis encrasicolus* sequence); (b) Backbone alignment of the yeast (*S. cerevisae*) and *E. encrasicolus* cytochrome *b* structures.

ENVIRONMENTAL CORRELATES OF MITOCHONDRIAL CLADES FREQUENCY

The clade frequencies shifted smoothly along latitudinal gradients in the Atlantic Ocean and between sea basins within the Mediterranean (figure 3.1). Clade A was found in all the sampling locations, whereas clade B was absent from locations off Senegal, Guinea, Ghana and Israel. Clade A was present in higher frequencies mostly at lower latitudes (off Portugal, Morocco, Canary Islands, Senegal, Guinea, Ghana, Angola, northern and central Aegean and Israel), while clade B was present in higher frequencies at higher

latitudes in the Atlantic Ocean (from the Norwegian coast to the Bay of Biscay and from the Namibia coast to South African waters) and in most of northern Mediterranean locations (off Gulf of Lion, Adriatic, Ionian and southern Aegean). Locations in the Bay of Biscay, Ligurian Sea and off Tunisia presented ratios between 0.4 and 0.6. From the six tested environmental variables, silicate was not considered in the final GLM because its inclusion did not significantly improve the model (χ^2 = 1.39, d.f. = 1, p = 0.24). The coefficients for the remaining variables are shown in table 3.3. Strikingly, sea temperature was the best clade frequency predictor (with a relative importance of 58.6%), followed by apparent oxygen utilization, phosphate, nitrate and salinity, with relative importance of 13.0%, 11.8%, 10% and 6.6%, respectively (figure 3.1).

Table 3.3. Results of the general linear models relating clade frequency with predictor variables from World Ocean Atlas 2009. Silicate is not listed because it did not contribute significantly to the model (see results).

Parameters	**estimate**	**std. error**	**z value**	**Pr(>|z|)**	**Sign.**
Intercept	-3.886	1.072	-3.625	2.89E-04	< 0.001
Temperature	-0.405	0.032	-12.547	<2.00E-16	< 0.001
Nitrate	-0.270	0.053	-5.055	4.29E-07	< 0.001
Salinity	0.293	0.036	8.101	5.44E-16	< 0.001
Phosphate	4.269	0.609	7.013	2.33E-12	< 0.001
apparent oxygen utilization	3.519	0.690	5.097	3.45E-07	< 0.001

DISCUSSION

Our results contribute to a better understanding of the role of natural selection in shaping the distribution of marine organisms, in particular the influence of sea temperature on the distribution of mitochondrial lineages in the European anchovy. Here we identified one putatively adaptive change in the mitochondrial cyt *b* gene, associated with clade B, more abundant in low temperature environments, suggesting that selection is acting on *E. encrasicolus* mito-genome.

GENETIC CLINES AND ENVIRONMENTAL CORRELATES OF MITOCHONDRIAL CLADE FREQUENCY

The European anchovy is widely distributed implying an adaptation to distinct environmental features, such as the steep thermal cline in the eastern Atlantic or salinity gradients between the Baltic Sea and the Atlantic Ocean. The two mtDNA clades found in the European anchovy are sympatric over most of the distribution range and exhibit a remarkable latitudinal cline in the eastern Atlantic (Silva *et al.*, *in press*). Previous studies (Grant, 2005; Magoulas *et al.*, 2006; Kristoffersen & Magoulas, 2008; Zarraonaindia *et al.*, 2012) assumed the observed two-clade pattern as a consequence of ancient isolations followed by secondary contact. However, genetic clines may represent a balance between selection, genetic drift and dispersal, along time and space (Barton & Hewitt, 1985; Barton & Gale, 1993). These clines are present in different small pelagic fish species and have been related to both historical factors and hydrographic barriers to dispersal in sardines (Chlaida *et al.*, 2009) or maintained by selective pressures in the Atlantic herring (Teacher *et al.*, 2012). One possible explanation for the origin and persistence of the dual-clade structure in the European anchovy may be an adaptation to the physical properties of the environment, in particular sea temperature, as suggested by the GLM (figure 3.1; table 3.3). Temperature along the distribution range of the European anchovy varies clinally (figure 3.1) and contributes 58.6% to the model, accounting for five to six times more variance in the geographic distribution of clade frequencies than any other environmental predictor. The second best predictor for the distribution of the mtDNA clade frequencies was apparent oxygen utilisation (13%). Oxygen availability is of extreme importance in ectothermic small pelagic fish, especially at higher latitudes where water temperature is low and consequently metabolism decreases. Although anchovies have high capacity for migration, environmental affinity precludes dispersal, contributing to population structure (Magoulas *et al.*, 2006; Zarraonaindia *et al.*, 2012; Viñas *et al.*, 2013; Silva *et al.*, *in press*). Temperature as been identified as one of the major selective forces acting on mtDNA (Ballard & Whitlock, 2004). Temperature-mediated selection was found in humans (Mishmar *et al.*, 2003; Ruiz-Pesini *et al.*, 2004; Balloux *et al.*, 2009), where genetic differentiation between pairs of populations is correlated to difference in temperature (Balloux *et al.*, 2009). In the marine realm, the influence of

temperature in shaping mitochondrial diversity was described in the walleye pollock (*Theragra chalcogramma*) where sea surface temperature and mitochondrial lineages were significantly correlated, showing a latitudinal clinal distribution and higher genetic diversity than under a mutation-drift equilibrium model (Grant *et al.*, 2006).

ADAPTATION AT THE MOLECULAR LEVEL

Small pelagic fish are ectothermic and metabolic rates increase with water temperature increments (Elliott, 1976). Anchovies have high metabolic requirements and more than 95% of the myofibrils are adjacent to mitochondria, suggesting a high dependence on aerobic metabolism (Johnston, 1982). When water temperature decreases, body temperature and metabolic rates decrease, probably affecting swimming performance (Blier & Guderley, 1993), muscle associated energetic needs (Johnston, 1982; Johnston & Dunn, 1987), egg size and fecundity (Ballard & Rand, 2005).

The substitution of a Valine for a Methionine in codon 368 could play an important role in the ETC, enhancing the electron transfer process from cyt *b* to cyt *c*. However, analysis of the crystal structure of the cyt *b* (cytochrome *bc1*, complex III) shows that the Valine replacement by Methionine at position 368 is not likely to affect efficiency of the electron transfer mechanism (figure 3.3), for the following reasons. First, a direct effect is unlikely, due to the marginal positioning of residue 368, well clear off the cyt *b* electron pathways. Second, conformational effects are also unlikely, because a single replacement of Valine by Methionine should not cause major structural changes, particularly at the protein surface, as is the case here. The latter consideration is supported by the similarity between the crystal structure and the ModBase model at this position (figure 3.3). Another possibility to explain positive selection in the codon 368 is that this substitution affects protein trafficking or membrane integration, or stability of the putative cytochrome *c* reductase-oxidase super-complex as indicated by the distribution of amino acid residues in the 18 non-redundant families of thermophilic proteins (KUMS000101; table S2). A change in heat capacity (HUTJ700101; table S2) could be related with a variation in the entropic component the free energy of folding of cyt *b* (hydrophobic

effect). But its significance depends on the magnitude of the heat capacity change, which should be very small for a single residue replacement. Also this effect tends to be much less important for a protein working in a non-polar environment like a lipid membrane. The biochemical intricacy of the phosphorilative oxidation processes preclude us from predicting the exact functional implications of the substitution, and we are limited to the suggestion that it will have an impact on the overall ATP production by the respiratory chain, and consequently on the overall metabolic performance of clade B.

CONCLUSIONS

Over the past 20 years, most of genetic studies performed in marine species were based on mtDNA markers and motivated to solve stock boundaries for fishery management (Grant *et al.*, 2006). However, these studies assumed that mtDNA was neutral and did not account for the influence of selective pressures in the estimation of population structure. Our results contribute to the growing body of evidence that mtDNA of natural populations is affected by selective pressures, that need to be accounted for in historical interpretations of biogeographic scenarios. Moreover, our findings strongly suggest that the contemporary distribution of mtDNA clade frequencies in the European anchovy is being shaped by temperature. Molecular adaptations to different metabolic requirements may be the key to understand how species will adapt to future climate change.

REFERENCES

Avise, J.C. (2000) *Phylogeography: the history and formation of species*. Harvard University Press, Cambridge, Massachussets, USA.

Ballard, J.W.O. & Whitlock, M.C. (2004) The incomplete natural history of mitochondria. *Molecular Ecology*, **13**, 729–744.

Ballard, J.W.O. & Rand, D.M. (2005) The population biology of mitochondrial DNA and its phylogenetic implications. *Annual Review of Ecology, Evolution, and Systematics*, **36**, 621–642.

Balloux, F., Handley, L.-J.L., Jombart, T., Liu, H. & Manica, A. (2009) Climate shaped the worldwide distribution of human mitochondrial DNA sequence variation. *Proceedings of the Royal Society B: Biological Sciences*, **276**, 3447–3455.

Barton, N.H. & Hewitt, G.M. (1985) Analysis of hybrid zones. *Annual Review of Ecology and Systematics*, **16**, 113–148.

Barton, N.H. & Gale, K.S. (1993) Genetic analysis of hybrid zones. *Hybrid Zones and the Evolutionary Process* (ed. by R.G. Harrison), pp. 13–45. Oxford University Press, New York.

Beckstead, W.A., Ebbert, M.T.W., Rowe, M.J. & McClellan, D.A. (2009) Evolutionary pressure on mitochondrial Cytochrome *b* is consistent with a role of CytbI7T affecting longevity during caloric restriction. *PLoS One*, **4**, e5836.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. & Bourne, P.E. (2000) The Protein Data Bank *Nucleic Acids Research*, **28**, 235-242.

Blier, P.U. & Guderley, H.E. (1993) Mitochondrial activity in rainbow trout red muscle: the effect of temperature on the ADP-dependence of ATP synthesis. *Journal of Experimental Biology*, **176**, 145–158.

Borrell, Y.J., Piñera, J.A., Sánchez Prado, J.A. & Blanco, G. (2012) Mitochondrial DNA and microsatellite genetic differentiation in the European anchovy *Engraulis encrasicolus* L. *ICES Journal of Marine Science*, **69**, 1357–1371.

Bouchenak-Khelladi, Y., Durand, J.D., Magoulas, A. & Borsa, P. (2008) Geographic structure of European anchovy: a nuclear-DNA study. *Journal of Sea Research*, **59**, 269–278.

Boyer, T.P., Antonov, J.I., Baranova, O.K., Garcia, H.E., Johnson, D.R., Locarnini, R.A., Mishonov, A.V., O'Brien, T.D., Seidov, D., Smolyar, I.V. & Zweng, M.M. (2009) *World Ocean Database*. NOAA Atlas NESDIS 66, U.S. Gov. Printing Office,, Wash., D.C.

Chevan, A. & Sutherland, M. (1991) Hierarchical partitioning. *The American Statistician*, **45**, 90–96.

Chlaida, M., Laurent, V., Kifani, S., Benazzou, T., Jaziri, H. & Planes, S. (2009) Evidence of a genetic cline for *Sardina pilchardus* along the northwest African coast. *ICES Journal of Marine Science*, **66**, 264–271.

da Fonseca, R., Johnson, W., O'Brien, S., Ramos, M. & Antunes, A. (2008) The adaptive evolution of the mammalian mitochondrial genome. *BMC Genomics*, **9**, 119.

Dalziel, A.C., Moyes, C.D., Fredriksson, E. & Lougheed, S.C. (2006) Molecular evolution of Cytochrome c Oxidase in high-performance fish (Teleostei: Scombroidei). *Journal of Molecular Evolution*, **62**, 319–331.

Darwin, C. (1859) *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life*, 1st edn. John Murray, London.

Delport, W., Poon, A.F.Y., Frost, S.D.W. & Kosakovsky Pond, S.L. (2010) Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics*, **26**, 2455–2457.

Drummond, A., Ashton, B., Buxton, S., Cheung, M., Cooper, A., Duran, C., Field, M., Heled, J., Kearse, M., Markowitz, S., Moir, R., Stones-Havas, S., Sturrock, S., Thierer, T. & Wilson, A. (2011) *Geneious v. 5.1*. Biomatters Ltd, available at: http://www.geneious.com.

Elliott, J.M. (1976) The energetics of feeding, metabolism and growth of brown trout (*Salmo trutta* L.) in relation to body weight, water temperature and ration size. *Journal of Animal Ecology*, **45**, 923–948.

Foote, A.D., Morin, P.A., Durban, J.W., Pitman, R.L., Wade, P., Willerslev, E., Gilbert, M.T.P. & da Fonseca, R.R. (2010) Positive selection on the killer whale mitogenome. *Biology Letters*, **7**, 116–118.

Galtier, N., Nabholz, B., Giémin, S. & Hurst, G.D.D. (2009) Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Molecular Ecology*, **18**, 4541–4550.

Garvin, M.R., Bielawski, J.P. & Gharrett, A.J. (2011) Positive darwinian selection in the piston that powers proton pumps in complex I of the mitochondria of Pacific salmon. *PLoS One*, **6**, e24127.

Gershoni, M., Templeton, A.R. & Mishmar, D. (2009) Mitochondrial bioenergetics as a major motive force of speciation. *BioEssays*, **31**, 642–650.

Grant, W.S. (2005) A second look at mitochondrial DNA variability in European anchovy (*Engraulis encrasicolus*): assessing models of population structure and the Black Sea isolation hypothesis. *Genetica*, **125**, 293–309.

Grant, W.S., Leslie, R.W. & Bowen, B.W. (2005) Molecular genetic assessment of bipolarity in the anchovy genus *Engraulis*. *Journal of Fish Biology*, **67**, 1242–1265.

Grant, W.S., Spies, I.B. & Canino, M.F. (2006) Biogeographic evidence for selection on mitochondrial DNA in north Pacific walleye pollock *Theragra chalcogramma*. *Journal of Heredity*, **97**, 571–580.

Haney, R.A., Silliman, B.R. & Rand, D.M. (2010) Effects of selection and mutation on mitochondrial variation and inferences of historical population expansion in a Caribbean reef fish. *Molecular Phylogenetics and Evolution*, **57**, 821–828.

Hijmans, R.J. (2013) Raster: Geographic data analysis and modeling. *R package version 2.1-48*,

Hochachka, P.W., Stanley, C., Merkt, J. & Sumar-Kalinowski, J. (1983) Metabolic meaning of elevated levels of oxidative enzymes in high altitude adapted animals: An interpretive hypothesis. *Respiration Physiology*, **52**, 303–313.

Johnston, I.A. (1982) Quantitative analyses of ultrastructure and vascularization of the slow muscle fibres of the anchovy. *Tissue and Cell*, **14**, 319–328.

Johnston, I.A. & Dunn, J. (1987) Temperature acclimation and metabolism in ectotherms with particular reference to teleost fish. In: *Symposia of the Society for the Experimental Biology* eds. K. Bowler and B.J. Fuller), pp. 67–93. Cambridge University Press, Cambridge, UK.

Kosakovsky Pond, S.L. & Frost, S.D.W. (2005) Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Molecular Biology and Evolution*, **22**, 1208–1222.

Kosakovsky Pond, S.L., Posada, D., Gravenor, M.B., Woelk, C.H. & Frost, S.D.W. (2006a) GARD: a genetic algorithm for recombination detection. *Bioinformatics*, **22**, 3096–3098.

Kosakovsky Pond, S.L., Frost, S.D.W., Grossman, Z., Gravenor, M.B., Richman, D.D. & Brown, A.J.L. (2006b) Adaptation to different human populations by HIV-1 revealed by codon-based analyses. *PLoS Comput Biol*, **2**, e62.

Kristoffersen, J.B. & Magoulas, A. (2008) Population structure of anchovy *Engraulis encrasicolus* L. in the Mediterranean Sea inferred from multiple methods. *Fisheries Research*, **91**, 187–195.

Magoulas, A., Tsimenides, N. & Zouros, E. (1996) Mitochondrial DNA phylogeny and the reconstruction of the population history of a species: the case of the European anchovy (*Engraulis encrasicolus*). *Molecular Biology and Evolution*, **13**, 178–190.

Magoulas, A., Castilho, R., Caetano, S., Marcato, S. & Patarnello, T. (2006) Mitochondrial DNA reveals a mosaic pattern of phylogeographical structure in Atlantic and Mediterranean populations of anchovy (*Engraulis encrasicolus*). *Molecular Phylogenetics and Evolution*, **3**, 734–746.

Mishmar, D., Ruiz-Pesini, E., Golik, P., Macaulay, V., Clark, A.G., Hosseini, S., Brandon, M., Easley, K., Chen, E., Brown, M.D., Sukernik, R.I., Olckers, A. & Wallace, D.C. (2003) Natural selection shaped regional mtDNA variation in humans. *Proceedings of the National Academy of Sciences*, **100**, 171–176.

Mitchell, P. (2011) Chemiosmotic coupling in oxidative and photosynthetic phosphorylation. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, **1807**, 150–1538.

Murrell, B., Wertheim, J.O., Moola, S., Weighill, T., Scheffler, K. & Kosakovsky Pond, S.L. (2012) Detecting individual sites subject to episodic diversifying selection. *PLoS Genetics*, **8**, e1002764.

Murrell, B., Moola, S., Mabona, A., Weighill, T., Sheward, D., Kosakovsky Pond, S.L. & Scheffler, K. (2013) FUBAR: a Fast, Unconstrained Bayesian AppRoximation for inferring selection. *Molecular Biology and Evolution*, **30**, 1196–1205.

Nally, R.M. (1996) Hierarchical partitioning as an interpretative tool in multivariate inference. *Australian Journal of Ecology*, **21**, 224–228.

Nei, M. & Gojobori, T. (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular Biology and Evolution*, **3**, 418–426.

Pieper, U., Eswar, N., Stuart, A.C., Ilyin, V.A. & Sali, A. (2002) ModBase, a database of annotated comparative protein structure models. *Nucl. Acids Res.*, **30**, 255-259.

Pierce, D. (2011) ncdf: Interface to Unidata netCDF data files. *R package version 1.6.6.*,

R Development Core Team (2013) R: A language and environment for statistical computing.

Ruiz-Pesini, E., Mishmar, D., Brandon, M., Procaccio, V. & Wallace, D.C. (2004) Effects of purifying and adaptive selection on regional variation in human mtDNA. *Science*, **303**, 223–226.

Shen, Y.-Y., Liang, L., Zhu, Z.-H., Zhou, W.-P., Irwin, D.M. & Zhang, Y.-P. (2010) Adaptive evolution of energy metabolism genes and the origin of flight in bats. *Proceedings of the National Academy of Sciences*, **107**, 8666–8671.

Silva, G., Horne, J. & Castilho, R. (*in press*) Anchovies go north and west without losing diversity: post-glacial range expansions in a small pelagic fish. *Journal of Biogeography*,

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. & Kumar, S. (2011) MEGA5: Molecular Evolutionary Genetics Analysis using maximum likelihood,

evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, **28**, 2731–2739.

Teacher, A., Andre, C., Merila, J. & Wheat, C. (2012) Whole mitochondrial genome scan for population structure and selection in the Atlantic herring. *BMC Evolutionary Biology*, **12**, 248.

Thompson, J.D., Higgins, D.G. & Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, **22**, 4673–4680.

Viñas, J., Sanz, N., Peñarrubia, L., Araguas, R.-M., García-Marín, J.-L., Roldán, M.-I. & Pla, C. (2013) Genetic population structure of European anchovy in the Mediterranean Sea and the northeast Atlantic Ocean using sequence analysis of the mitochondrial DNA control region. *ICES Journal of Marine Science*,

Walsh, C. & Nally, R.M. (2013) hier.part: Hierarchical Partitioning. *R package version 1.0-4*,

Woolley, S., Johnson, J., Smith, M.J., Crandall, K.A. & McClellan, D.A. (2003) TreeSAAP: Selection on amino acid properties using phylogenetic trees. *Bioinformatics*, **19**, 671–672.

Yu, L., Wang, X., Ting, N. & Zhang, Y. (2011) Mitogenomic analysis of Chinese snub-nosed monkeys: evidence of positive selection in NADH dehydrogenase genes in high-altitude adaptation. *Mitochondrion*, **11**, 497–503.

Zarraonaindia, I., Iriondo, M., Albaina, A., Pardo, M.A., Manzano, C., Grant, W.S., Irigoien, X. & Estonba, A. (2012) Multiple SNP markers reveal fine-scale population and deep phylogeographic structure in European anchovy (*Engraulis encrasicolus* L.). *PLoS ONE*, **7**, e42201.

SUPPORTING INFORMATION

Thermal adaptation and clinal mtDNA variation of the European anchovy

Gonçalo Silva, Fernando P. Lima, Paulo Martel and Rita Castilho

Table S1 - Data information of samples used for environmental correlates analyses and number of individuals sequenced used for selection tests.

Sea Basin	Country	Location	Map Code	Reference	Year	Month	Longitude	Latitude	N	A	B	Nseq
Atlantic	Norway	Oslo	1	Silva et al., submitted	2007	October	10.63	58.98	24	2	22	24
Atlantic	Denmark	Denmark	2	Zarraonindia et al., 2012	2007	July	11.57	57.14	15	1	14	
Atlantic	Scotland	Scotland	3	Zarraonindia et al., 2012	2009	February	2.33	56.14	11	1	10	
Atlantic	Scotland	Scotland	3	Zarraonindia et al., 2012	2009	May	-10.43	58.09	24	5	19	
Atlantic	Germany	Kiel	4	Zarraonindia et al., 2012	2006	November	10.17	54.92	28	3	25	
Atlantic	Germany	Germany	4	Zarraonindia et al., 2012	2007	April	6.49	54.43	11	1	10	
Atlantic	France	English Channel	5	Silva et al., submitted	2007	October	0.12	50.75	27	3	24	27
Atlantic	France	St Jean de Luz	6	Magoulas et al., 2006	2001	April	-1.68	43.40	50	21	29	
Atlantic	France	St Jean de Luz	6	Magoulas et al., 1996	1993	March	-1.68	43.40	47	19	28	
Atlantic	France	Nantes	6	Borrel et al., 2012	2009	May	-2.50	47.00	29	15	14	
Atlantic	Spain	Bay of Biscay-5029	7	Zarraonindia et al., 2012	2008	May	-1.87	45.87	29	13	16	
Atlantic	Spain	Bay of Biscay-5020	7	Zarraonindia et al., 2012	2009	May	-1.92	44.60	27	11	16	
Atlantic	Spain	Bay of Biscay-5001	7	Zarraonindia et al., 2012	2009	May	-2.21	43.35	45	31	14	
Atlantic	Spain	Bay of Biscay	7	Silva et al., submitted	2007	-	-2.85	43.54	23	9	14	23
Atlantic	Spain	East Cantabrian Sea	7	Borrel et al., 2012	2009	April	-2.60	44.00	15	9	6	
Atlantic	Spain	West Cantabrian Sea	7	Borrel et al., 2012	2009	April	-4.01	43.47	39	29	10	
Atlantic	Spain	Getaria Coast	7	Borrel et al., 2012	2009	May	-2.54	44.20	12	3	9	
Atlantic	Spain	Valdearenas	7	Borrel et al., 2012	2009	April	-3.97	43.44	20	10	10	
Atlantic	Spain	Galicia	8	Zarraonindia et al., 2012	2010	March	-8.94	42.53	29	26	3	
Atlantic	Portugal	Portugal South	9	Zarraonindia et al., 2012	2008	February	-9.35	38.60	28	22	6	
Atlantic	Portugal	Aveiro	9	Magoulas et al., 2006	1998	March	-8.66	40.71	70	63	7	25
Atlantic	Portugal	Olhão	9	Magoulas et al., 2006	1997	July	-7.85	36.94	57	51	6	
Atlantic	Portugal	Armação de Pêra	9	Silva et al., submitted	2007	July	-8.40	37.10	29	24	5	29
Atlantic	Spain	Cadiz	10	Zarraonindia et al., 2012	2009	April	-6.47	36.54	60	55	5	
Atlantic	Morocco	Tangier	11	Magoulas et al., 2006	1997	December	-5.85	35.88	62	60	2	38
Atlantic	Spain	Canary islands	12	Zarraonindia et al., 2012	2007	May	-15.65	27.72	28	26	2	
Atlantic	Spain	Canary islands	12	Magoulas et al., 2006	1999	March	-15.04	28.27	48	44	4	24
Atlantic	Senegal	Dakar	13	Magoulas et al., 2006	1999	March	-17.61	14.82	34	32	2	25

Sea Basin	Country	Location	Map Code	Reference	Year	Month	Longitude	Latitude	N	A	B	Nseq
Atlantic	Guinea-Bissau	Guinea-Bissau	14	Silva et al., submitted	2006	April	-14.23	9.72	20	20	0	20
Atlantic	Ghana	Accra	15	Silva et al., submitted	2008	April	0.02	5.59	25	25	0	25
Atlantic	Angola	Angola	16	This work	2007	August	13.78	-10.90	24	20	4	24
Atlantic	Namibia	Namibia	17	This work	2007	August	11.70	-17.17	24	5	19	24
Atlantic	South Africa	South Africa	18	Zarraonindia et al., 2012	2009	September	18.00	-34.00	30	4	26	
Atlantic	South Africa	Hout Bay	18	Grant et al., 2005	1987	August	18.35	-34.06	18	16	2	
Atlantic	South Africa	South Africa	18	This work	2007	August	20.97	-34.70	11	4	7	11
Mediterranean	Italy	Livorno	19	Magoulas et al., 2006	1998	May	10.09	43.52	55	27	28	
Mediterranean	France	Gulf of Lion	20	This work	2008	June	3.98	43.20	22	5	17	22
Mediterranean	France	Sete	20	Magoulas et al., 1996	1992	December	3.98	43.20	50	21	29	
Mediterranean	Croatia	Split	21	Borrel et al., 2012	2009	June	16.13	43.03	20	5	15	
Mediterranean	Italy	Adriatic Sea	22	Zarraonindia et al., 2012	2007	October	15.47	42.83	27	3	24	
Mediterranean	Italy	Chioggia	22	Magoulas et al., 1996	1993	November	12.49	45.31	57	7	50	
Mediterranean	Italy	Bari	22	Magoulas et al., 2006	1997	July	17.00	41.30	70	8	62	
Mediterranean	Italy	Chioggia	22	Magoulas et al., 2006	1997	July	12.49	45.31	57	5	52	28
Mediterranean	Italy	Otranto	22	Magoulas et al., 2006	1998	August	18.55	40.14	71	15	56	
Mediterranean	Spain	Tarragona	23	Zarraonindia et al., 2012	2009	March	1.17	40.88	33	11	22	
Mediterranean	Spain	Barcelona	23	Grant et al., 2005	1988	September	1.86	41.12	16	7	9	
Mediterranean	Greece	Aegean Sea	24	Zarraonindia et al., 2012	2008	July	24.68	40.25	32	31	1	
Mediterranean	Greece	Gulf of Kavala	24	Magoulas et al., 1996	1989	May	24.29	40.65	141	121	20	
Mediterranean	Greece	Gulf of Kavala	24	Magoulas et al., 1996	1989	October	24.30	40.60	57	46	11	
Mediterranean	Greece	Pagasitikos Gulf	24	Magoulas et al., 1996	1992	September	23.60	38.91	20	17	3	
Mediterranean	Greece	Oreoi Strait	24	Kristoffersen & Magoulas 2008	2000	December	23.03	38.96	126	105	21	
Mediterranean	Greece	Oreoi Strait	24	Kristoffersen & Magoulas 2008	2000	December	23.03	38.96	56	44	12	
Mediterranean	Greece	Oreoi Strait	24	Kristoffersen & Magoulas 2008	2000	December	23.03	38.96	115	85	30	
Mediterranean	Greece	Saronikos Gulf	24	Magoulas et al., 1996	1993	June	23.07	37.17	59	20	39	
Mediterranean	Greece	Patraikos Gulf	25	Magoulas et al., 1996	1989	August	21.67	38.26	118	36	82	
Mediterranean	Greece	Patras	25	Kristoffersen & Magoulas 2008	2001	January	21.67	38.26	103	42	61	
Mediterranean	Greece	Patraikos Gulf	25	Magoulas et al., 1996	1989	October	21.67	38.26	54	21	33	
Mediterranean	Greece	Patras	25	Kristoffersen & Magoulas 2008	2001	February	21.67	38.26	86	30	56	
Mediterranean	Tunisia	Tunis	26	This work	2009	February	10.65	37.28	28	16	12	28
Mediterranean	Spain	Alboran Sea	27	Zarraonindia et al., 2012	2009	October	-4.04	36.53	68	63	5	
Mediterranean	Spain	Malaga	27	Magoulas et al., 2006	1998	November	-4.35	36.60	47	43	4	25
Mediterranean	Israel	Telaviv	28	This work	2008	July	34.59	32.08	26	26	0	26

* - sample used only for selection tests due to the absence of sampling month; N - number of individuals; A - clade A absolute frequencies; B - clade B absolute frequencies; Nseq - number of individuals sequenced.

Table S2 - Significant physico-chemical properties of cytochrome *b* codon 368 amino acid changes ($p < 0.001$): TreeSAAP analyses of magnitude categories ranging from 6 to 8.

Property	Property reference	Category	Z-score value
Normalized frequency of C-terminal beta-sheet	CHOP780209	6	3.257
Normalized frequency of alpha-helix, with weights	LEVM780101	6	4.761
Distribution of amino acid residues in the 18 non-redundant families of thermophilic proteins	KUMS000101	6	4.787
Normalized positional residue frequency at helix termini C4	AURR980112	6	5.140
Amino acid composition	DAYM780101	6	5.310
Alpha-helix indices for alpha/beta-proteins	GEIM800104	6	6.918
Amino acid distribution	JUKT750101	7	3.402
AA composition of EXT of single-spanning proteins	NAKH920103	7	4.368
Normalized frequency of beta-sheet in alpha/beta class	PALJ810112	7	4.897
Normalized frequency of alpha-helix in alpha/beta class	PALJ810109	7	5.045
Heat capacity	HUTJ700101	7	5.750
Normalized composition from animal	NAKH900106	8	3.510
Normalized composition of mt-proteins	NAKH900104	8	3.510
Relative preference value at C5	RICJ880110	8	4.250
A parameter defined from the residuals obtained from the best correlation of the Chou-Fasman parameter of beta-sheet	CHAM830102	8	5.045

CHAPTER IV ● PLEISTOCENE CLIMATIC OSCILLATIONS AND THE EVOLUTION OF THE OLD WORLD ANCHOVIES *ENGRAULIS SPP.*

Gonçalo Silva[1], Regina Cunha[1] and Rita Castilho[1]*

[1] Centro de Ciências do Mar (CCMAR, CIMAR - Associate Laboratory), Universidade do Algarve, Campus de Gambelas, 8005-139 Faro, Portugal;

*Corresponding author at: Centro de Ciências do Mar (CCMAR, CIMAR-Associate Laboratory), Universidade do Algarve, Campus de Gambelas, 8005-139 Faro, Portugal. Fax: +351 289 800 069.

E-mail address: rcastil@ualg.pt

In preparation

ABSTRACT

Aim As part of an emerging effort to understand how marine pelagic fishes evolved, we examined the patterns of genetic variation among Old World Anchovies (OWA) complex *Engraulis spp* to estimate the origin and analyze the connectivity between the Atlantic and Pacific anchovies and propose a biogeographic scenario for the group.

Location Adults were collected from the northwestern Atlantic, eastern Atlantic (from Norway to South Africa) and from the western Pacific (Japan, Sydney and New Zealand).

Methods A 1045 bp fragment of the mtDNA cytochrome *b* gene was sequenced (n = 373) and nuclear 8 microsatellite loci were genotyped (n = 531) for anchovies from 16 locations. Populations were surveyed for diversity, differentiation and connectivity with a range of summary statistics. Multivariate discriminant analysis of principal components was employed to detect reproductive isolation among nominal species. A phylogeny of the species complex was performed to establish the phylogenetic relationship between taxa. Additionaly, we also presented a mtDNA cytochrome *b* gene phylogeny of the family Engraulidae to date lineage-splitting events.

Results Although Atlantic and Pacific anchovies show independent histories, both ancient and contemporary gene-flow between the two oceanic basins was found. In the Atlantic Ocean, no genetic differentiation was depicted between *E. encrasicolus*, *E. capensis* and *E. eurystole*, with taxa grouping in two previously described clades with a latitudinal frequency cline. Within the Pacific, two different clades likely corresponding to *E. japonicus* and *E. australis* revealed both ancient and contemporary admixture. The origin of the OWA occurred at about 3 Ma, while divergence between the Atlantic and Pacific anchovies dates from 0.62 Ma.

Main conclusions Anchovies high ability for dispersal and tolerance to wide range of temperatures allowed them to disperse from the Pacific to the Atlantic Ocean, to maintain the connectivity between the two oceans through time and thus prevent speciation. Anchovies probably colonized the Atlantic Ocean migrating along south Asia, Middle East and eastern Africa continental platforms reaching the Cape of the Good Hope.

Keywords: Old World Anchovies, *Engraulis*, historical biogeography, phylogeography, inter-oceanic migrations, trans-Equatorial dispersals, speciation, Pleistocene.

INTRODUCTION

Periodic climatic events affect the evolution of the species shaping their biogeographic and macroecological patterns (Jansson & Dynesius, 2002). Speciation among marine taxa is mostly related to geological and climatic events, both implying different time scales (Briggs, 1987a, b). Although cyclical periodicity of range shifts may enhance secondary contacts and prevent speciation (Jansson & Dynesius, 2002), the isolation experienced by some peripheral populations promotes differentiation. Species that were formed during the Pleistocene often exhibit shallow divergences due to their recent isolation or incipient speciation (Rocha & Bowen, 2008).

Theoretically, physical barriers that could trigger vicariant allopatric or peripatric speciation are absent in the marine environment (Palumbi, 1992; Palumbi *et al.*, 1997). Nonetheless, organisms often show limited distributional ranges imposed by intrinsic physiological constraints. Throughout the Pleistocene climatic oscillations, organisms were able to cross temporarily interrupted barriers to dispersal during specific periods of time (Mirams *et al.*, 2011). Transitions over soft barriers include e.g., long-distance migrations through warm Equatorial waters (Burridge, 2002) or interoceanic migrations (Bowen *et al.*, 2001; Rocha *et al.*, 2005; Bowen *et al.*, 2006).

The Old World anchovies (OWA) complex (family Engraulidae) are coastal fish species distributed along offshore areas above the continental platforms of the Atlantic and Pacific oceans, restricted sea basins in the Mediterranean Sea, the Baltic and the Black seas as well as inshore environments such as estuaries, inlets and bays (Whitehead *et al.*, 1988). This species complex comprises five nominal species: the Japanese anchovy *E. japonicus* Temminck & Schlegel, 1846 and the Australian anchovy *E. australis* Shaw, 1790 in the western Pacific; the Cape anchovy *E. capensis* Gilchrist, 1913 in the southeastern Atlantic Ocean; the silver anchovy *E. eurystole* Swain & Meek, 1884 in the western Atlantic Ocean; the European anchovy *E. encrasicolus* (Linnaeus, 1758) in the eastern Atlantic, Baltic Sea, Mediterranean Sea and the Black Sea. Recently, the white anchovy *E. albidus* Borsa, Collet & Durand, 2004 was added to the OWA group, based on ecological, morphological and genetic differences between two estuarine and two pelagic populations in the Mediterranean Sea (Figure 1; Borsa, 2002; Borsa *et al.*, 2004). The original description of OWA

nominal species was mainly taxonomically and geographically based, with taxa longitudinally (e.g. *E. encrasicolus* and *E. eurystole*) and latitudinaly (e.g. *E. japonicus* and *E. australis*) disjunct (Whitehead *et al.*, 1988). Despite a large body of literature focusing on this group, the phylogenetic relationship between the OWA species remain poorly understood. This group is thought to have diverged recently from the remaining Engraulidae at about one million years ago (Grant *et al.*, 2010).

The genetic divergence between OWA species is shallow and there are several conflicting aspects in the taxonomy and molecular classification of the group. Based on morphological characters, Whitehead et al. (1988) proposed that the whole group should be considered a single species and molecular studies revealed that some of the described species have no genetic basis. No significant genetic differences were found between *E. encrasicolus* and *E. eurystole* (Silva *et al.*, *in press*) and the existence of shared mtDNA haplotypes between *E. encrasicolus* and *E. capensis* (Grant *et al.*, 2005) or between *E. australis* and *E. japonicus* (Grant & Bowen, 2006), seriously compromised the current taxonomy of this group. The description of typical nuclear alleles of *E. albidus* anciently introgressed in *E. capensis* (Bouchenak-Khelladi *et al.*, 2008) or the paraphyly of *E. encrasicolus* mtDNA lineages (Grant *et al.*, 2005; Grant *et al.*, 2010) generated further controversy.

Pleistocene climate swings influenced anchovies range shifts, leading to transequatorial dispersals during episodes of global cooling or through deep cold water (Grant *et al.*, 2005; Grant & Bowen, 2006) and promoted inter-oceanic migrations (Grant & Bowen, 2006). Anchovies are thought to have colonized the Atlantic Ocean through the southern Indian Ocean (Grant & Bowen, 2006). Anchovies in the Atlantic Ocean likely experienced several extinction-colonization cycles at the extremes of the distribution driven by Pleistocene climate shifts (Grant & Bowen, 2006; Alheit *et al.*, 2012; Silva *et al.*, *in press*). Moreover, the western Atlantic was colonized after the last glacial maximum (LGM) from anchovies dispersing from western African populations of the European anchovy (Silva *et al.*, *in press*). In the Pacific Ocean, the Japanese and the Australian anchovies diverged 105 ka (thousand years ago) to 420 ka (Liu *et al.*, 2006). The genetic signature of Pacific anchovies revealed persistence on separated hemispheres over several glacial cycles, although recent dispersers were identified (Grant & Bowen, 2006).

Thus far, studies involving OWA were based on allozymes and on a fragment of the mitochondrial cytochrome *b* (cyt *b*) gene (521 bp; Grant *et al.*, 2005; Grant & Bowen, 2006). In this study, we focused on all five OWA nominal species using a larger portion (1,045 bp) of cyt *b* and eight nuclear microsatellites to provide a novel perspective of the OWA evolution and to analyze the connectivity between the Atlantic and Pacific anchovies. We also dated main lineage splitting events within Engraulidae to determine the age of the OWA, and analyzed routes of dispersal to propose a biogeographic scenario for this group.

MATERIAL AND METHODS

SAMPLE COLLECTION, DNA EXTRACTION AND PCR AMPLIFICATION

Samples of OWA likely representing five putative species were collected from 16 locations from both Atlantic and Pacific oceans (Figure 4.1; Table 4.1). Fish were purchased at small coastal fish markets, as artisanal fisherman do not venture far, or were collected on scientific cruises (see acknowledgements). A small portion of white muscle or fin was preserved in 96% ethanol and stored at -20°C. Total genomic DNA was extracted by a saline method (Sambrook & Russell, 2001). DNA extraction, polymerase chain reaction (PCR), purification of the PCR product, sequencing for a fragment of the mitochondrial cyt *b* (1045 bp) and microsatellite genotyping of eight loci were performed as described in Silva *et al.* (*in press*). Sequences were deposited in GenBank (accession numbers: JQ716609–JQ716731, JQ716748–JQ716756, JX683020–JX683113, KF601435–KF601478 and KJ007642- KJ007734).

Table 4.1 Sample locations, sample abbreviations, collection dates, sample sizes and summary statistics for a 1045 bp sequence fragment of the mtDNA cytochrome *b* and eight nuclear microsatellites of European anchovy (*Engraulis encrasicolus*).

Location	Code	Nominal species	Long	Lat	Year	Mitochondrial Cytochrome *b*				Microsatellites						
						n	n_h	***h***	π	***N***	***Aavg***	A_r = 9	A_r =18	***Effnum***	H_O	H_E
Norway	NO	*E. encrasicolus*	10.6	59.0	2007	24	17	0.953	0.009	40	13.1	7.4	10.3	6.97	0.714	0.850
Poland	PL	*E. encrasicolus*	16.5	54.6	2008	9	7	0.917	0.014	9	8.4	8.4	–	6.08	0.736	0.864
English Channel	EC	*E. encrasicolus*	0.1	50.8	2007	27	18	0.963	0.010	45	13.0	7.3	9.8	6.60	0.729	0.837
Bay of Biscay	BB	*E. encrasicolus*	-2.9	43.5	2007	23	19	0.980	0.015	45	16.0	7.9	10.4	6.93	0.762	0.837
Portugal - North	PN	*E. encrasicolus*	-8.8	40.7	1998	25	24	0.997	0.013	45	17.3	8.3	11.5	8.38	0.825	0.882
Portugal - South	PS	*E. encrasicolus*	-8.4	37.1	2007	29	27	0.995	0.011	43	17.9	7.6	12.6	8.59	0.777	0.880
Canary Islands	CA	*E. encrasicolus*	-15.0	28.3	1999	24	23	0.996	0.007	42	17.8	8.4	12.1	9.93	0.792	0.888
Senegal	SN	*E. encrasicolus*	-17.6	14.8	1999	25	25	1.000	0.006	37	12.5	6.9	9.4	4.97	0.693	0.797
Guinea-Bissau	GU	*E. encrasicolus*	-14.2	9.7	2006	20	20	1.000	0.006	19	13.1	8.4	12.9	6.75	0.653	0.868
Ghana	GH	*E. encrasicolus*	0.0	5.6	2008	25	25	1.000	0.006	27	15.5	8.9	12.9	8.66	0.766	0.883
Namibia	NM	*E. capensis*	11.7	-17.2	2007	24	17	0.920	0.010	32	14.4	8.9	11.6	8.38	0.769	0.875
South Africa	SA	*E. capensis*	21.0	-34.7	2007	13	10	0.923	0.019	21	12.6	9.1	11.5	8.15	0.768	0.887
USA	US	*E. eurystole*	-66.1	41.5	2006	12	9	0.909	0.004	18	13.5	9.3	13.5	7.98	0.793	0.894
Japan	JP	*E. japonicus*	139.9	35.6	2006	24	24	1.000	0.011	30	17.1	7.8	13.0	9.01	0.699	0.869
Australia	AU	*E. australis*	151.0	-35.0	2008	34	33	0.998	0.007	44	21.4	9.4	13.6	10.53	0.742	0.899
New Zealand	NZ	*E. australis*	175.0	-36.7	2005	35	15	0.709	0.001	34	14.4	7.4	11.4	6.77	0.717	0.849
Total						373	269	0.993	0.023	531	47.6	32.0	39.6	7.00	0.746	0.866

Long, longitude; Lat, latitude; *N*, number of individuals; *n*, number of haplotypes; *h*, haplotype diverstity; π, nucleotide diversity; *Aavg*, average number of alleles, *Ar*, allelic richness; *Effnum*, Effective number of alleles; H_0, observed mean heterozygozity; H_E, expected mean heterozygozity.

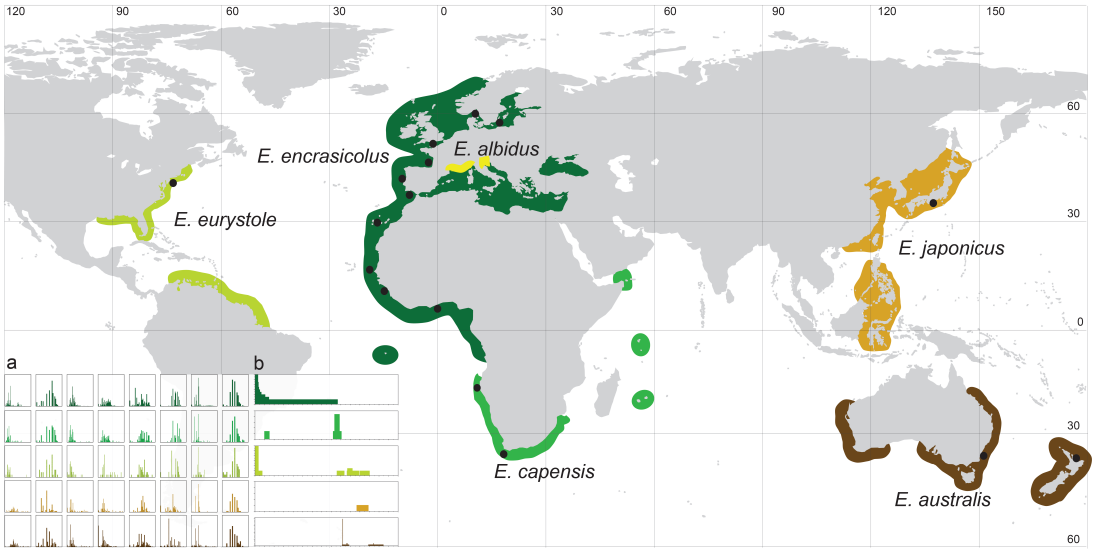


Figure 4.1 Present-day distribution of putative Old World anchovies species (Whitehead *et al.*, 1988); black dots represent sample locations. a) distribution of nuclear microsatellite alleles per locus (columns) and per putative species (rows); b) distribution of mitochondrial haplotypes per putative species (rows).

GENETIC ANALYSIS

Cyt *b* sequences were aligned using CLUSTALX 2.0.3 with default settings, implemented in GENEIOUS 5.4 (Drummond A. *et al.*, 2011) and checked manually. Sequences were reduced to haplotypes using COLLAPSE 1.2 (Posada, 2004). For cyt *b*, we calculated number of individuals (n), number of haplotypes (n_h) and haplotype (h) and nucleotide diversities (π) in ARLEQUIN 3.5.1.2 (Excoffier & Lischer, 2010). Summary statistics, number of individuals (n), mean number of alleles (n_a), observed heterozygosity (H_O) and expected heterozygosity (H_E) were calculated for each location and for each locus with GENODIVE (Meirmans & Van Tienderen, 2004).

POPULATION STRUCTURE, DIFFERENTIATION AND CONNECTIVITY

To examine the relationship between mitochondrial haplotypes, a minimum spanning network was constructed with ARLEQUIN 3.5.1.2 (Excoffier & Lischer, 2010) and visualized with HAPSTAR (Teacher & Griffiths, 2011). Pairwise genetic differentiation was estimated with G_{st_est} (Hedrick & Goodnight, 2005) and Jost's D_{est} value (Jost,

2008), both within and between putative species, following Pennings *et al.* (2011) for mtDNA and using the R package DIVERSITY (Keenan *et al.*, 2013) for microsatellites.

We used discriminant analysis of principal components (DAPC) (Jombart *et al.*, 2010) to analyze the structure of Old World anchovies, a non-model (i.e., principal component analysis; PCA)-based approach. Unlike Bayesian clustering (Pritchard *et al.*, 2000; Guillot *et al.*, 2005; Corander *et al.*, 2008), DAPC has the advantage of being free of assumptions regarding Hardy–Weinberg or linkage equilibrium (Jombart *et al.*, 2008; Jombart *et al.*, 2010). The method relies on allele data transformation using Principal Component Analysis (PCA) as a prior step to Discriminant Analysis (DA). DA defines a model in which genetic variation is partitioned into a between-group and a within-group component. Groups can be defined *a priori* (i.e., populations, species, locations collection sites, cohorts etc.) or can be inferred using first sequential K-means (Legendre & Legendre, 1998). This analysis was performed using the ADEGENET package (Jombart, 2008) for the R 3.0.2 software (R Development Core Team, 2013). We ran two DAPC analyses. First we assigned each individual to its nominal species (*a priori* species assignment), and secondly we used the "find.clusters" function to determine the number of genetic clusters and assignment of individuals to each of those clusters. We therefore obtained an unbiased interpretation of the genetic structure of the OWA species complex by removing the effect of *a priori* species assignment. Retaining too many PCs can lead to overfitting the discriminant functions, which could model any structure and virtually discriminate any set of clusters. ADEGENET proposes an optimization procedure to evaluate the optimal numbers of PCs to retain. The procedure is based on the calculation of the α-score, which measures the difference between the proportion of successful reassignment of the analysis (observed discrimination) and values obtained using random groups (random discrimination). The number of retained PCs can be chosen so as to optimize the a-score. The optimal number of clusters was chosen on the basis of the lowest associated Bayesian Information Criterion (BIC). Then we used DAPC to assign individuals into clusters, retaining the number of principal components using 85% of the cumulative deviance.

ESTIMATION OF MIGRATION RATES

We used the coalescent-based program Migrate-N (Beerli & Felsenstein, 1999; Beerli & Felsenstein, 2001) to compare different biogeographic hypotheses for the past and present migration of OWA. This approach assumes the Wright-Fisher model, where locations have a constant effective size through time, the rate of mutation is constant, and locations exchange migrants with constant rates per generation, but those rates can vary among locations. We conducted the analyses with two sets of data, mitochondrial DNA and microsatellites, structured into two groups according to geographical regions: southern Atlantic Ocean (pooled southern locations Senegal, Guinea-Bissau, Ghana, Namibia, Angola and South Africa) and Pacific Ocean (Japan, Australia and New Zealand). We tested four variations of the two-population (Atlantic-Pacific) migration model: bidirectional migration (full model, 4 parameters), strict Atlantic to Pacific migration (3 parameters), strict Pacific to Atlantic migration (3 parameters) and panmictic model that assumes the Atlantic and Pacific are part of a panmictic population (1 parameter). Testing the directionality of gene flow is justified because the dominant ocean current between the ocean basins, the Agulhas flow, runs westerly from the Indian to the Atlantic Ocean and is thought to play a limiting role in marine dispersal in the opposite direction (Lutjeharms, 2006; Ivanova, 2009). Initial values were calculated using F_{ST}. Mutation rates were set to be constant among loci. The Migrate-N run parameters were calibrated on the full model for convergence of the Markov chain Monte Carlo sampling method. The prior distributions were uniform for mutation-scaled population size parameters theta (θ), that are four times the product of the effective population size and the mutation rate, and mutation-scaled migration rates M, that is, immigration rate scaled by the mutation rate, over the range of θ = 0.05–0.5 and prior migration rate M = 0–100 for mtDNA, and θ = 10–100 and M = 0–50 for microsatellites. These settings resulted in converged posterior distributions with a clear maximum for each estimate. The Bayesian run for mtDNA consisted of one long chain with a total of 15 million states visited and 50,000 states recorded for the generation of posterior distribution histograms for each locus after discarding the first 10,000 genealogies as burn-in; for all loci, a total of 48 million states were visited and 160,000 samples were recorded. For all the analyses we used an adaptive heating scheme with 4 simultaneous chains using different acceptance ratios (temperature settings were 1.0; 1.5; 3.0; $1x10^6$); the analyses were run on a

cluster computer using 4 compute nodes per run. The Bayesian run for microsatellites consisted of one long chain with a total of 6 million states were visited and 20,000 states were recorded for the generation of posterior distribution histograms for each locus after discarding the first 10,000 genealogies as burn-in; for all loci, a total of 48 million states were visited and 160,000 samples were recorded. For all the analyses we used an adaptive heating scheme with 4 simultaneous chains using different acceptance ratios (temperature settings were 1.0; 1.5; 3.0; 1,000,000.0); the analyses were run on a cluster computer using 4 compute nodes per run. Overall loci information was combined into a single estimate by Bézier approximation of the thermodynamic scores as described by Beerli & Palczewski (2010) and we averaged the Bézier score over three different runs and used as input to evaluate multiple models using Bayes factors (Bloomquist *et al.*, 2010).

PHYLOGENETIC ANALYSES

To set the biogeographical scenario for the OWA complex we performed a phylogenetic analysis based on a fragment of the mitochondrial cytochrome *b* gene with 110 representative individuals from the five putative OWA *Engraulis* species (110 taxa, 1045bp). We used *E. mordax* and *Anchovia clupeoides* as outgroups (for accession numbers see Support Information Table S2). The Akaike Information Criterion (AIC) (Akaike, 1974) implemented in MODELTEST 3.7 (Posada & Crandall, 1998), selected the GTR+I+Γ as the evolutionary model that best fitted the data set. The inferred parameters were used in maximum likelihood (ML) and Bayesian Inference (BI) analyses. BI analyses based on the mitochondrial data set were conducted with MRBAYES 3.2.1 (Ronquist *et al.*, 2012). Metropolis-coupled Markov chain Monte Carlo (MCMC) analyses were run for 6,000,000 generations, and sampled every 100 generations. Convergence between the two runs was assessed by examining the potential scale reduction factors (PSRF) and effective sample size (ESS). The data set was analyzed under the GTR+I+Γ evolutionary model. Length of burnin was determined by examination of traces in TRACER 1.4 (Rambaut & Drummond, 2007). The first 60,000 generations were discarded, and robustness of the inferred trees was evaluated using Bayesian posterior probabilities (BPPs). PHYML 3.0 (Guindon & Gascuel, 2003) was used to estimate the ML tree, and to test by non-

parametric bootstrapping the robustness of the inferred trees using 1,000 pseudoreplicates. The GTR+I+Γ was selected as the best-fit evolutionary model and inferred model parameters were used in the ML analysis.

To estimate OWA origin and date lineage-splitting events within Engraulidae we used a Bayesian relaxed molecular-clock approach as implemented in BEAST 1.7.4 (Drummond *et al.*, 2012) based on partial mitochondrial cyt *b* sequences (1,040 bp). We selected at least one representative species per genus (except *Papuengraulis*; 57 taxa; accession numbers in supporting information Table S2) and six outgroup species *Chirocentrus dorab*, *Clupea harengus*, *Denticeps clupeoides*, *Ilisha africana*, *Sardina pilchardus*, *Sundasalanx mekongensis* (for accession numbers see Support Information Table S2), selected from Lavoue *et al.* (2010). We used a Yule tree prior that assumes a constant speciation rate among lineages. Estimates of divergence are generally based upon known historical events, geologic or fossil, which can be used as calibration points for a taxon-specific mutation rate estimate (McCormack *et al.*, 2011). We used three calibration points based on the fossil record. One refers to the earliest record of Engraulidae [6–12 Ma] from the Miocene - lower Pliocene of Cyprus (Grande & Nelson, 1985). The second calibration corresponds to *E. japonicus* [2–0 Ma] from Kokubu group, Japan (Yabumoto, 1988), and the third to the divergence between *Cetengraulis mysticetus* and *C. edentulus* due to the closure of the Isthmus of Panama (Lessios *et al.*, 1999; Grant *et al.*, 2010). Calibrations using the three fossils were modeled with a lognormal distribution, where 95% of the prior weight fell within the geological interval in which each fossil was discovered. For the Engraulidae [6–12 Ma], the parameters of the lognormal calibration prior were: hard minimum bound 6.0, mean 1.099 and standard deviation 0.422. For *E. japonicus* [2–0 Ma], the parameters of the lognormal calibration prior were: hard minimum bound 0, mean 0 and standard deviation 0.425. For the divergence between *Cetengraulis mysticetus* and *C. edentulous* we used a calibration according to Coppard *et al.* (2013), where the closure occurred between 3.1–2.8 Ma. Lognormal calibration was set to: hard minimum bound 2.8, mean -1.9, and standard deviation 0.424. MCMC analyses were run for 40,000,000 generations with a sample frequency of 10,000, following a discarded burn-in of 4,000,000 steps. The convergence to the stationary distributions was confirmed by inspection of the MCMC samples using TRACER 1.4 (Rambaut & Drummond, 2007). BI and dating analyses were performed on the

CCMAR Computational Cluster Facility (http://gyra.ualg.pt) at the University of Algarve, and the ML analysis was performed on the web server Mobyle (http://mobyle.pasteur.fr/).

RESULTS

MTDNA

A total of 373 individuals from five OWA putative species were analysed, yielding 269 haplotypes. Haplotype diversity (*h*) was generally high, ranging from 0.917 to 1.000 in the eastern Atlantic Ocean (from Norway to South Africa), 0.909 in the western Atlantic, and from 0.709 to 1.000 in the Pacific Ocean (Table 4.1). Nucleotide diversity (π) was low, ranging from 0.4% (USA) to 1.9% (South Africa) in the Atlantic Ocean and from 0.1% to 1.1% in the Pacific Ocean (Table 4.1).

No genetic differentiation was found between *E. encrasicolus*, *E. capensis* and *E. eurystole*. Alternatively, haplotypes from the Atlantic taxa grouped into two previously described clades (Magoulas *et al.*, 1996) with a latitudinal frequency cline. Clade A haplotypes were more frequent at lower latitudes (From Portugal North to Ghana), while clade B predominated in samples north of the English Channel and from Namibia to South Africa. Individuals from the western Atlantic all grouped within clade A. Both clades were present in all locations, with the exception of West Africa (Senegal, Guinea-Bissau and Ghana) and USA, where clade B is absent. In the Pacific Ocean, three lineages were found: one corresponding to anchovies from Japan, but including two shared haplotypes from two individuals from South Africa, one endemic to southern Pacific and a third clade of two japanese anchovies grouping on the southern Pacific.

The haplotype network revealed four main clades, two in Atlantic Ocean and two in the Pacific Ocean (Figure 4.2). Clades from the two oceanic basins were separated by a minimum of 30–32 mutations. Within the Atlantic Ocean, putative species *E. encrasicolus*, *E. capensis* and *E. eurystole* were not monophyletic, but alternatively fit on European anchovy clades A and B. Two individuals from South Africa clustered in the northern Pacific clade. The "southern" Pacific clade mostly includes haplotypes from *E. australis*, but also two haplotypes from *E. japonicus* that are separated nine mutations from the most frequent haplotype of this clade, likely representing

intermediate haplotypes between the two clades. The four clades exhibited different haplotype patterns: *E. encrasicolus* clade A and *E. australis* were characterized by multiple star-like radiations with relatively shallow genetic divergences; *E. encrasicolus* clade B and *E. japonicus* lacked distinct star patterns and exhibited many unsampled or extinct haplotypes.

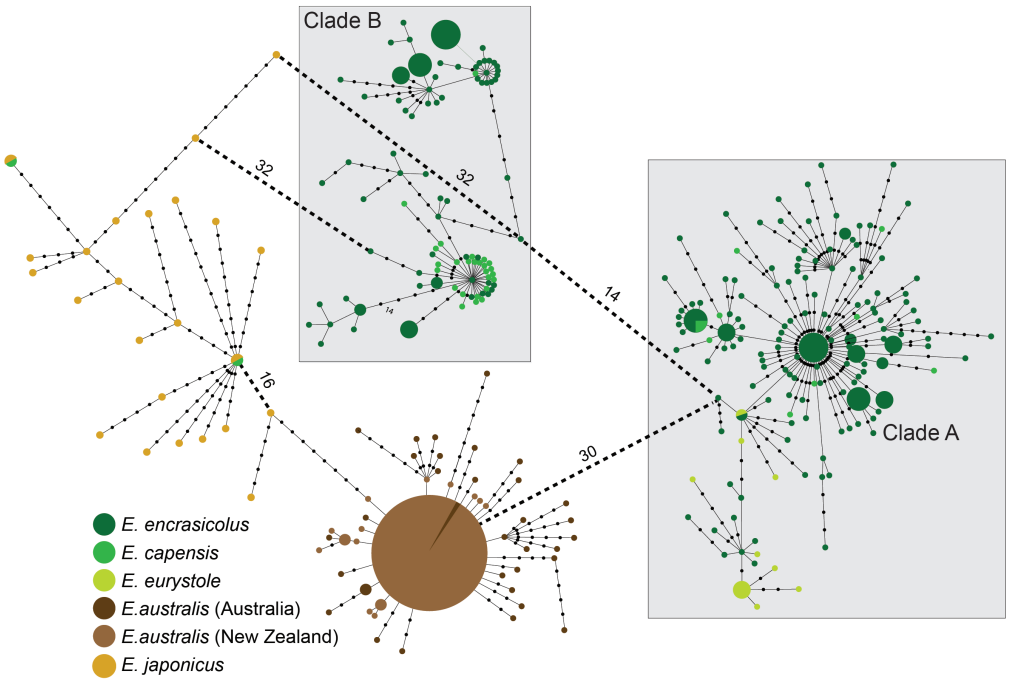


Figure 4.2 Minimum spanning tree for the Old World anchovies, based on the mitochondrial cytochrome *b* (1045 bp, 373 individuals). The colour and the size of the circles represent the geographic origin (according to Figure 1, with the exception of New Zealand haplotypes which are represented here in light brown) and frequency of each haplotype, respectively. The smallest colored circles represent a singleton haplotype. Black circles represent either extant unsampled sequences or extinct ancestral sequences. The length of the lines connecting haplotypes is proportional to the number of mutational differences separating the haplotypes, except when otherwise indicated.

Estimates with mtDNA data of the mutation scaled population size parameter were the same between ocean basins (Θ = 0.1). Our comparison of candidate models of gene flow between populations, clearly reject panmixia, and showed that the full geneflow model (model 1) fitted our mtDNA data best (Table 4.2), with highly asymmetrical immigration between ocean basins with the Pacific population providing five times as many immigrants into the Atlantic Ocean than vice-versa (M = 14.5 vs 2.9 respectively).

Table 4.2 Bayes factors model comparison of migration models for Old World Anchovies between the Atlantic (ATL) and the Pacific (PAC) oceans.

Marker	Models		Model parameters	Bézier	dBézier	Probability
mtDNA	Model 1	ATL <-> PAC	****	-5339,10	0,00	1
	Model 2	ATL <- PAC	*0**	-5396,31	-57,20	1,4342E-25
	Model 3	ATL -> PAC	**0*	-5375,88	-36,77	1,071E-16
	Model 4	(ATL+PAC)	*0*0	-5524,26	-185,15	3,87457E-81
Microsatellites	Model 1	ATL <-> PAC	****	-329230,64	-232312,46	0
	Model 2	ATL <- PAC	*0**	-96918,17	0,00	1
	Model 3	ATL -> PAC	**0*	-205619,62	-108701,45	0
	Model 4	(ATL+PAC)	*0*0	-384554,60	-287636,43	0

Model parameters code as follows: asterisk (*) indicates that a particular migration rate was estimated by the model and 0 indicates that no migration was allowed. The first sign indicates theta for Atlantic, the second sign migration to Atlantic, the third sign theta for Pacific and the forth sign migration to Pacific.

MICROSATELLITE DNA

Multilocus genotypes from the 16 locations from the Atlantic Ocean and Pacific Ocean were obtained for 462 anchovies. The number of alleles per locus varied from 17 (locus Ee2-91b) to 82 (locus Ee10) over all locations (Table S1 in Supplementary information). Mean allelic richness, standardized for comparison across a minimum common sample size of 9 individuals, ranged from 6.9 (Senegal) to 9.3 (USA) in the Atlantic Ocean and from 7.4 to 9.4 in the Pacific Ocean. Expected heterozygosity (H_E) varied between 0.797 (Senegal) and 0.894 (USA) in the Atlantic Ocean and between 0.849 and 0.899 in the Pacific Ocean, while the observed heterozygosity (H_O) varied between 0.653 (Guinea-Bissau) and 0.825 (Portugal north) in the Atlantic Ocean and between 0.699 and 0.742 in the Pacific Ocean (Table 4.1).

The DAPC results show that *E. encrasicolus*, *E. eurystole* and *E. capensis* are closer to each other than to the *E. australis* and *E. japonicus* clusters (Figure 4.3a). The probability of assignment of individuals to their nominal species (Figure 4.3b) shows that *E. eurystole* and *E. capensis* individuals were mostly assigned to the *E. encrasicolus* clusters with probabilities close to 0.8, and *E. australis* and *E. japonicus*

were mostly assigned to their own clusters also with high probabilities. The *a posteriori* DAPC analysis (Figure 4.3c) shows two clusters, one comprising 51% (167 out of 324) *E. encrasicolus* individuals and 17 *E. eurystole* and 45 *E. capensis,* 76 *E. australis* and 29 *E. japonicus* and another with 49% *E. encrasicolus.*

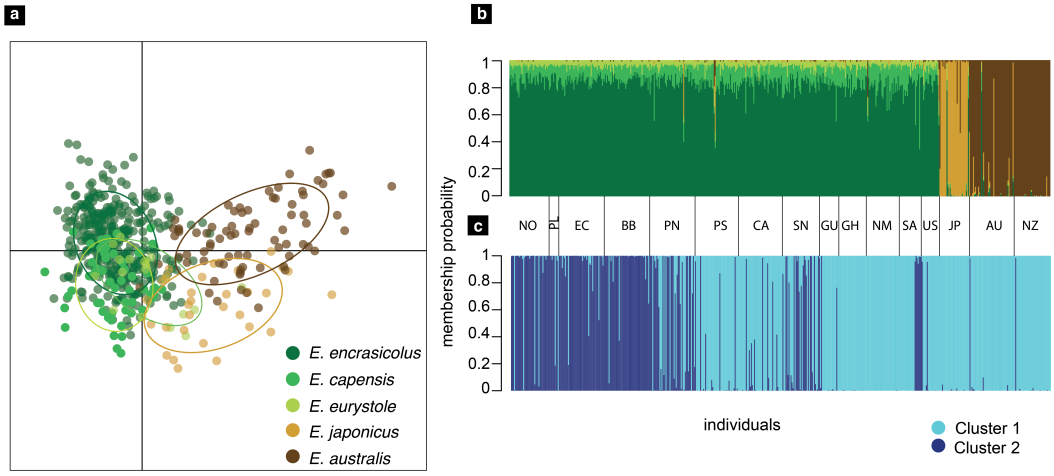


Figure 4.3 a) Discriminant analysis of principal components (DAPC) of multi-locus Old World Anchovies genotypes; individual genotypes appear as circles ellipses represents the centre of dispersion of each putative species (see Figure 4.1; Table 4.1). Horizontal and vertical axes are the first two principal components, respectively; b) scatter plot of *a priori*-defined nominal species; c) individuals assigned to their genetic cluster without forcing them into pre-determined groups.

Estimates with microsatellite data of the mutation scaled population size parameter were higher in the Pacific Ocean ($\Theta_{Atlantic}$ = 11.6 and $\Theta_{Pacific}$ = 98.4). Our comparison of candidate models of gene flow between populations, clearly reject panmixia, and showed that the Pacific to Atlantic model (model 2) fitted our microsatellite data best (Table 4.2), with an extremely low gene flow from the Pacific into the Atlantic (M = 1.5).

PHYLOGENETIC ANALYSES

Maximum likelihood and Bayesian phylogenetic approaches based on the mtDNA cyt *b* data set yielded the topology depicted in Figure 4. BI analyses revealed the existence of three well-supported clades. Two of them in the Atlantic (clades A and B) and the third in the Pacific corresponding to *E. australis*. In the BI analysis *E.*

japonicus was retrieved as paraphyletic, as well as for the Atlantic clades (A and B) in the ML tree.

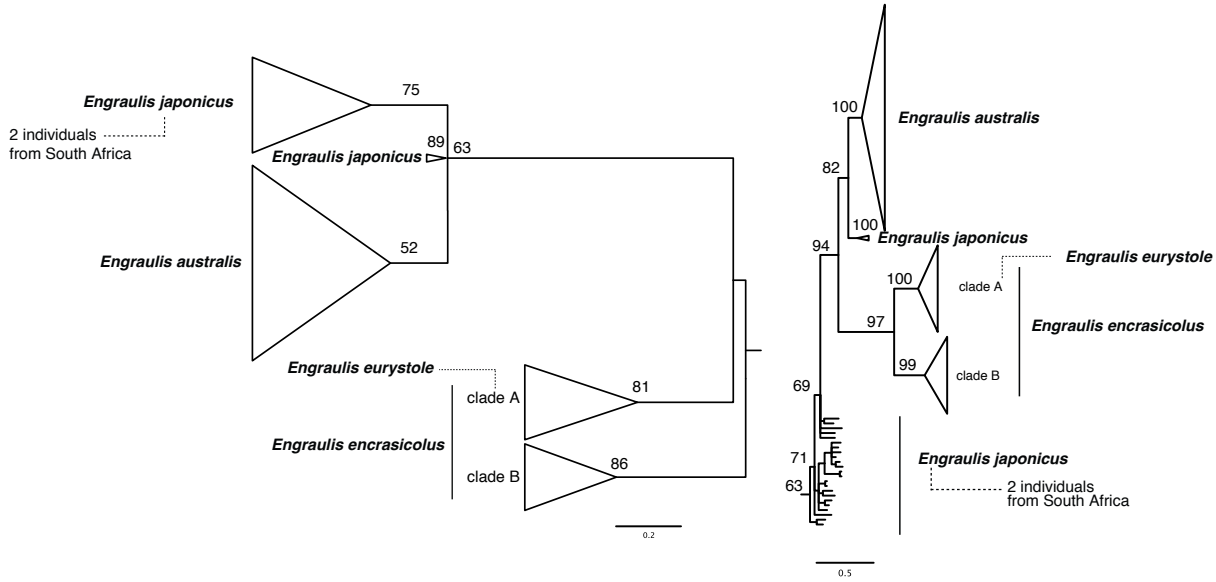


Figure 4.4 Maximum-likelihood (left) and Bayesian (right) trees inferred from a fragment of 1045bp of cyt *b* for Old World anchovies *Engraulis spp.*. Maximum-likelihood bootstrap values larger than 50% and Bayesian posterior probabilities (BPP) greater than 75 for major supported clades are shown above branches; the trees were rooted using *Anchovia clupeoides* and *Engraulis mordax* as outgroups.

Beast dating analysis based on the mtDNA cyt *b* data set of the family Engraulidae estimated the origin of the OWA at about 3.16 Ma [2.08–4.29] and the divergence between the Atlantic and Pacific anchovies at 0.62 Ma ago [0.39–0.90] (Figure 4.5). The age of the most recent common ancestor of the Atlantic anchovies was estimated at 0.37 [0.19–0.62] Ma, while the trans-Atlantic colonization occurred at 0.09 [0.03–0.20]. The split between Pacific anchovies was estimated at about 0.25 [0.11–0.45] Ma.

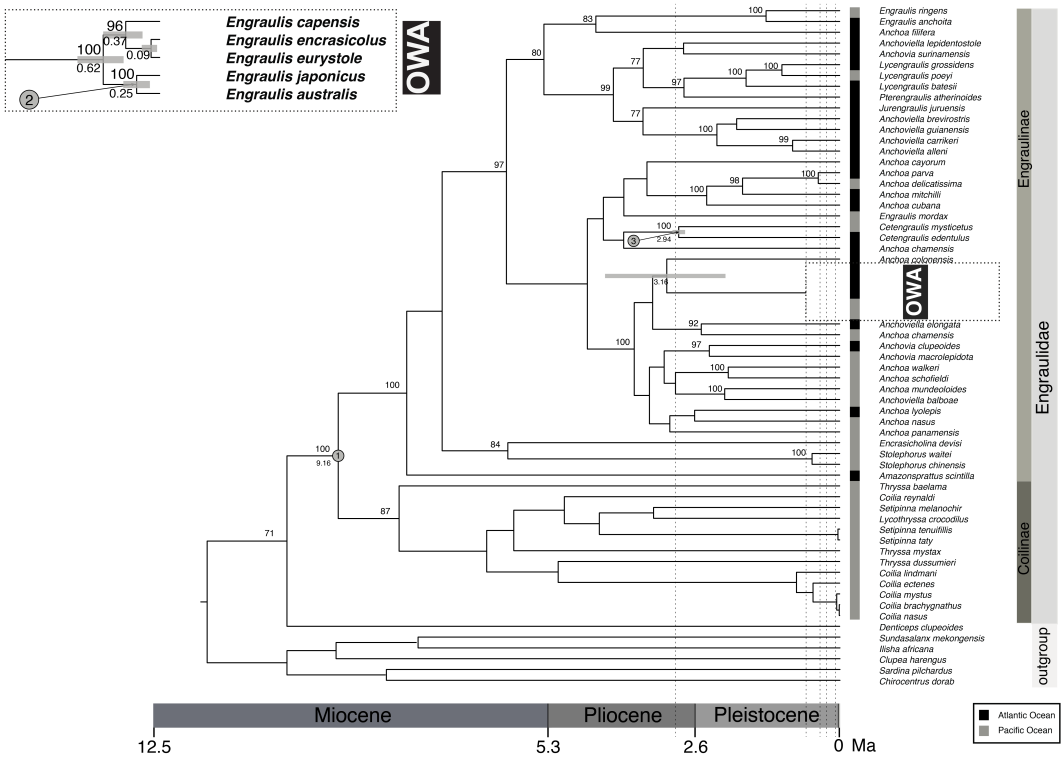


Figure 4.5 Bayesian dating analysis of the Engraulidae family inferred from a fragment of 1040 bp of the mitochondrial cytochrome *b* gene; numbers above and below branches represent node posterior probabilities and ages of the nodes, respectively. Vertical dashed lines indicate relevant lineage splitting-events. The zoomed dashed box represents Old World Anchovies (OWA) taxa.

DISCUSSION

The genetic patterns among the OWA complex indicate no genetic diferentiation between putative species from the Atlantic Ocean, and shallow divergences with ongoing geneflow between Atlantic and Pacific anchovies. The two species from the Pacific (*E. japonicus* and *E. australis*) only received statistical support in the Bayesian analysis and *E. japonicus* was retrieved as paraphyletic. Microsatellite data showed contemporary admixture. Based on our results, we suggest that previous described species of OWA are conspecific and should be considered a single species with regional variants. Nevertheless, we were not able to obtain specimens from *E. albidus* and therefore, no definitive conclusions regarding its taxonomy can be taken. Regardless the addition of more sequence data the origin of the OWA remains uncertain. The colonization of the Atlantic Ocean likely occurred through South Africa, with anchovies dispersing across the northern Indian Ocean along the

continental platform of south Asia, Middle East and eastern Africa. The Atlantic and Pacific OWA revealed independent historical demographies with contemporary gene flow. Before addressing the main interpretations and conclusions of these results, two main caveats must be addressed. First, data from the Pacific Ocean rely on few sampling sites. Second, despite rare records in the eastern African coast, no intermediate locations were sampled between the Atlantic Ocean and the Pacific Ocean. Therefore, we may have a restricted representation of the genetic diversity from the Indo-Pacific region. However, the results presented in this study provide solid arguments to set a biogeographical scenario for the OWA evolution.

AGE ESTIMATES SHOWED SHALLOW LINEAGE DIVERGENCE WITHIN OWA

Our Bayesian dating analysis revealed the existence of shallow mitochondrial divergences within OWA lineages (Figure 4.5). Most the global mtDNA phylogenies for marine taxa show evidence of long-term isolation between Atlantic and Indian-Pacific Oceans (Grant & Bowen, 1998). Our results showed that the divergence between the Pacific and the Atlantic OWA occurred during the Pleistocene at 620 ka (Figure 4.5). This estimate is more recent than previously assessed (Grant & Bowen, 2006; Grant *et al.*, 2010), which can be explained by the use of different methodologies. Here, we used a relaxed molecular-clock dating analysis and fossil calibrations whereas the former studies estimated this divergence using a fixed mutation rate.

FROM THE PACIFIC TO THE ATLANTIC: OPEN-OCEAN VS. COASTAL DISPERSAL

Anchovies were thought to have reached the Atlantic Ocean through an open-ocean route that involved dispersal from the Pacific towards the Atlantic via South Africa (Grant *et al.*, 2005; Grant & Bowen, 2006). This dispersal route would imply more than 8000 km of open-ocean migration between the western Australia to South Africa, without stepping-stones. Grant and his colleagues (2005; 2006) considered that this route would be the only possible pathway because the known distribution of anchovies seemed to reflect temperate-water habitat requirements like occurs with other pelagic species (e.g. sardines; Grant & Bowen, 1998). Nevertheless, a temperate

habitat may not be a limiting factor for the species involved, as they can be found in tropical waters both in the Atlantic and Pacific oceans (Whitehead *et al.*, 1988). Further, we obtained samples from tropical regions (Guinea-Bissau and Ghana) and thus, the geographical distribution of OWA seems to be wider than previously thought.

We propose an alternative route in which the dispersal of the anchovies from the Pacific to the Atlantic also occurs via the tip of South Africa but through coastal dispersal following continental platforms across the northern Indian Ocean and the eastern Africa coast (Figure 4.6). Anchovies are coastal pelagic fishes that live in average up to three years and it would be extremely unlikely to survive to such a trans-oceanic migration in a single generation with no stopovers (Lessios & Robertson, 2013). Moreover, open-ocean areas usually exhibit low productivity and scarce food resources (Sigman & Hain, 2012), which would create additional difficulties to large-scale migrations.

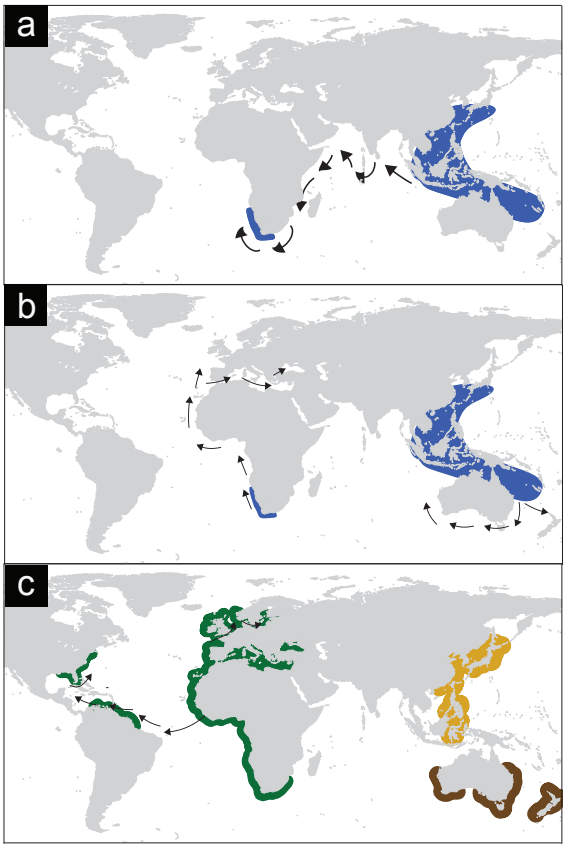


Figure 4.6. Hypothetical biogeographic scenario of Old World anchovies (OWA); a) colonisation route of Atlantic OWA at 620 ka (Lessios *et al.*, 1999); blue shadings represent putative distribution of ancestral populations; b) trans-Equatorial dispersals in each oceanic basin; c) present-day distribution of OWA, with post-LGM trans-Atlantic dispersal and northeastern Atlantic colonisation after ice sheets retreat around 8 ka.

The existence of shared haplotypes (Figure 4.2) and microsatellite alleles between Atlantic and Indo-Pacific OWA revealed by our study points to the existence of recurrent migration events and contemporary gene flow between ocean basins. The OWA high ability for migration and tolerance to wide temperature ranges provide the opportunity for trans-oceanic dispersal. Atlantic colonizers were likely seeded from an Indo-Pacific pool that could have included both north and south Pacific anchovy ancestors, departing from Philippines and Indonesia, and using Somalia, Mauritius and Seychelles as stepping-stones (Whitehead *et al.*, 1988). The studies performed thus far on the evolution of OWA (Grant & Bowen, 1998; Grant & Leslie, 2005; Grant & Bowen, 2006), already suggested dispersal from the Pacific to the Atlantic, probably via southern Africa, but had no evidence to support this hypothesis.

Pleistocene colonizations from the Pacific to the Atlantic through the Cape of the Good Hope were inferred for other coastal fish species, but mostly tropical (Bowen *et al.*, 2001; Rocha *et al.*, 2005; Bowen *et al.*, 2006; Lessios & Robertson, 2013). Although the Benguela upwelling in western South Africa constitutes a barrier for several species, its intensity has oscillated along the Pleistocene and allowed punctuated episodes of dispersal (Marlow *et al.*, 2000). Despite the few observed exceptions such as those in cichlids, rockfishes or in some butterflyfishes that are compatible with a Pleistocene time frame [0.01–2.6 Ma], most of the contemporary reef fishes show an earlier origin (Pliocene [2.6-5.3 Ma] or Miocene [5.3-23 Ma]) (reviewed in Rocha & Bowen, 2008). Our results indicate that mtDNA genealogy of OWA also coalesced during the Pleistocene at about 620 ka (Figure 4.5).

TRANS-EQUATORIAL DISPERSALS

OWA putative species have been commonly referred as having an anti-tropical distribution because inhabit, preferentially, temperate waters of the Atlantic and western Pacific (Grant *et al.*, 2005; Grant & Bowen, 2006). However, these fish can also be found in tropical waters (Guinea-Bissau and Ghana, this study) suggesting higher tolerance to warmer temperatures than previously thought (Grant *et al.*, 2005; Grant & Bowen, 2006). According to our results, anchovies from south Portugal to South Africa largely constitute a single population, which implies trans-Equatorial gene flow (Figure 4.3). Also, historical gene flow is also inferred from the reticulated haplotype network as South African anchovies do not show independent demographic

histories and share haplotypes with European fish.

We also identified both historical and contemporary trans-Equatorial dispersals in the Pacific Ocean. Two Japanese haplotypes were recovered as sister group of the southwestern Pacific lineage (*E. australis*) suggesting an ancient divergence, while microsatellites show evidence of contemporary gene flow (Figure 4.3). Anchovies seem to use the Central Indo-Pacific islands as stepping-stones for trans-Equatorial dispersals as corroborated by occasional records in the Phillipines and Indonesia (Whitehead *et al.*, 1988). Alternatively, Pacific OWA may inhabit permanently these areas, although in small numbers. Therefore, contemporary warm oceanic temperatures do not seem to constitute a barrier to the dispersal of OWA.

CONCLUSIONS

Our survey of the OWA complex *Engraulis spp.* indicate the existence of shallow mitochondrial divergences within OWA lineages. Results suggest that divergence between Atlantic and Pacific anchovies occurred recently during the Pleistocene at 620 ka. The occurrence of shared haplotypes and microsatellite alleles between Atlantic and Indo-Pacific anchovies revealed by our study suggests the existence of recurrent migration events and contemporary gene flow, even if at very low levels, between ocean basins. The life history traits of OWA related to migration abilities and tolerance to wide temperature ranges shaped the genetic architecture of the species.

ACKNOWLEDGMENTS

This work was only possible due to the collaboration of many colleagues in collecting and shipping samples: E. Torstensen (IMR, Norway), M. Wyszynski (SFI, Poland), S. Vaz, F. Coppin, Y. Verin and D. Leroy (IFREMER, France), I. Zarraonaindia (UPV/EHU, Spain), K. Erzini (CCMAR, Portugal), A. Dias (Portugal), J. Tornero and J. Alarcón (IEO, Spain), M. Santamaria (Canary Islands), A. Hervé (Senegal), L. Fonseca (Guinea-Bissau), H. Dankwa (WRI, Ghana), J. Vasslides and R. Hagan (RUMFS, USA) and E. Wiley and A. Bentley (KNHM, USA), Ana Tereza Gonçalves, Miguel Ruano, and Aníbal Guerra (Portugal), Rob Leslie (South Africa) and Stewart Grant (USA), Kevin Rowling (Australia) and Rob Tesker (New Zealand). We thank N. Coelho, C. Patrão and M. Valente for technical help. G.S.'s research was supported by a doctoral grant from FCT-Portugal (SFRH/BD/36600/2007).

REFERENCES

Akaike, H. (1974) A new look at the statistical model identifications. *IEEE Transactions on automatic control*, **19**, 716—723.

Alheit, J., Pohlmann, T., Casini, M., Greve, W., Hinrichs, R., Mathis, M., O'Driscoll, K., Vorberg, R. & Wagner, C. (2012) Climate variability drives anchovies and sardines into the North and Baltic Seas. *Progress In Oceanography*, **96**, 128—139.

Beerli, P. & Felsenstein, J. (1999) Maximum likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. *Genetics*, **152**, 763—773.

Beerli, P. & Felsenstein, J. (2001) Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. *Proceedings of the National Academy of Sciences*, **98**, 4563—4568.

Beerli, P. & Palczewski, M. (2010) Unified framework to evaluate panmixia and migration direction among multiple sampling locations. *Genetics*, **185**, 313—326.

Bloomquist, E.W., Lemey, P. & Suchard, M.A. (2010) Three roads diverged? Routes to phylogeographic inference. *Trends in Ecology & Evolution*, **25**, 626—632.

Borsa, P. (2002) Allozyme, mitochondrial-DNA, and morphometric variability indicate cryptic species of anchovy (*Engraulis encrasicolus*). *Biological Journal of Linnean Society*, **75**, 261–269.

Borsa, P., Collet, A. & Durand, J.D. (2004) Nuclear-DNA markers confirm the presence of two anchovy species in the Mediterranean. *Comptes Rendus Biologies*, **327**, 1113–1123.

Bouchenak-Khelladi, Y., Durand, J.D., Magoulas, A. & Borsa, P. (2008) Geographic structure of European anchovy: a nuclear-DNA study. *Journal of Sea Research*, **59**, 269–278.

Bowen, B.W., Muss, A., Rocha, L.A. & Grant, W.S. (2006) Shallow mtDNA coalescence in Atlantic pygmy angelfishes (genus *Centropyge*) indicates a recent invasion from the Indian Ocean. *Journal of Heredity*, **97**, 1—12.

Bowen, B.W., Bass, A.L., Rocha, L.A., Grant, W.S. & Robertson, D.R. (2001) Phylogeography of the trumpetfishes (*Aulostomus*): ring species complex on a global scale. *Evolution*, **55**, 1029–1039.

Briggs, J.C. (1987a) Antitropical distribution and evolution in the Indo-West Pacific Ocean. *Systematic Biology*, **36**, 237—247.

Briggs, J.C. (1987b) Antitropicality and vicariance. *Systematic Biology*, **36**, 206—207.

Burridge, C.P. (2002) Antitropicality of Pacific fishes: molecular insights. *Environmental Biology of Fishes*, **65**, 151—164.

Coppard, S.E., Zigler, K.S. & Lessios, H.A. (2013) Phylogeography of the sand dollar genus *Mellita*: Cryptic speciation along the coasts of the Americas. *Molecular Phylogenetics and Evolution*, **69**, 1033—1042.

Corander, J., Marttinen, P., Sirén, J. & Tang, J. (2008) Enhanced Bayesian modelling in BAPS software for learning genetic structures of populations. *Bmc Bioinformatics*, **9**, 539.

Drummond A., Ashton B, Buxton S, Cheung M, Cooper A, Duran C, Field M, Heled J, Kearse M, Markowitz S, Moir R, Stones-Havas S, Sturrock S, Thierer T & A, W. (2011) *Geneious*. Biomatters, Ltd.

Drummond, A.J., Suchard, M.A., Xie, D. & Rambaut, A. (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, **29**, 1969–1973.

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

Grande, L. & Nelson, G. (1985) Interrelationships of fossil and recent anchovies (Teleostei: Engrauloidea) and description of a new species from the Miocene of Cyprus. *American Museum Novitates*, 1—16.

Grant, W.S. & Bowen, B.W. (1998) Shallow population histories in deep evolutionary lineages of marine fishes: insights from sardines and anchovies and lessons for conservation. *The Journal of Heredity*, **89**, 415–426.

Grant, W.S. & Leslie, R.W. (2005) Bayesian analysis of allozyme markers indicates a single genetic population of kingklip *Genypterus capensis* off South Africa. *African Journal of Marine Science*, **27**, 479-485.

Grant, W.S. & Bowen, B.W. (2006) Living in a tilted world: climate change and geography limit speciation in Old World anchovies (*Engraulis*; Engraulidae). *Biological Journal of the Linnean Society*, **88**, 673–689.

Grant, W.S., Leslie, R.W. & Bowen, B.W. (2005) Molecular genetic assessment of bipolarity in the anchovy genus *Engraulis*. *Journal of Fish Biology*, **67**, 1242–1265.

Grant, W.S., Lecomte, F. & Bowen, B.W. (2010) Biogeographical contingency and the evolution of tropical anchovies (genus *Cetengraulis*) from temperate anchovies (genus *Engraulis*). *Journal of Biogeography*, **37**, 1352–1362.

Guillot, G., Mortier, F. & Estoup, A. (2005) Geneland: a computer package for landscape genetics. *Molecular Ecology Notes*, **5**, 712–715.

Guindon, S. & Gascuel, O. (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology*, **52**, 696–704.

Hedrick, P.W. & Goodnight, C. (2005) A standardized genetic differentiation measure. *Evolution*, **59**, 1633–1638.

Ivanova, E.V. (2009) *The global thermohaline paleocirculation*. Springer, New York.

Jansson, R. & Dynesius, M. (2002) The fate of clades in a world of recurrent climatic change: Milankovitch oscillations and evolution. *Annual Review of Ecology and Systematics*, **33**, 741—777.

Jombart, T. (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, **24**, 1403–1405.

Jombart, T., Devillard, S. & Balloux, F. (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*, **11**, 94.

Jombart, T., Devillard, S., Dufour, A.B. & Pontier, D. (2008) Revealing cryptic spatial patterns in genetic variability by a new multivariate method. *Heredity*, **101**, 92–103.

Jost, L. (2008) G(ST) and its relatives do not measure differentiation. *Molecular Ecology*, **17**, 4015–4026.

Keenan, K., McGinnity, P., Cross, T.F., Crozier, W.W. & Prodöhl, P.A. (2013) diveRsity: An R package for the estimation and exploration of population genetics parameters and their associated errors. *Methods in Ecology and Evolution*, **4**, 782–788.

Lavoué, S., Miya, M. & Nishida, M. (2010) Mitochondrial phylogenomics of anchovies (family Engraulidae) and recurrent origins of pronounced

miniaturization in the order Clupeiformes. *Molecular Phylogenetics and Evolution*, **56**, 480—485.
Legendre, P. & Legendre, L.F.J. (1998) *Numerical Ecology*. Elsevier Science, Amsterdam.
Lessios, H.A. & Robertson, D.R. (2013) Speciation on a round planet: phylogeography of the goatfish genus Mulloidichthys. *Journal of Biogeography*, **40**, 2373—2384.
Lessios, H.A., Kessing, B.D., Robertson, D.R. & Paulay, G. (1999) Phylogeography of the pantropical sea urchin *Eucidaris* in relation to land barriers and ocean currents. *Evolution*, **53**, 806—817.
Liu, J.X., Gao, T.X., Zhuang, Z.M., Jin, X.S., Yokogawa, K. & Zhang, Y.P. (2006) Late Pleistocene divergence and subsequent population expansion of two closely related fish species, Japanese anchovy (*Engraulis japonicus*) and Australian anchovy (*Engraulis australis*). *Molecular Phylogenetics and Evolution*, **40**, 712—723.
Lutjeharms, J.R.E. (2006) The coastal oceans of south-eastern Africa. *The sea: the global coastal ocean* (ed. by A.R. Robinson and K. Brink), p. 783—834. Harvard University Press.
Marlow, J.R., Lange, C.B., Wefer, G. & Rosell-Melé, A. (2000) Upwelling intensification as part of the Pliocene-Pleistocene climate transition. *Science*, **290**, 2288—2291.
McCormack, J.E., Heled, J., Delaney, K.S., Peterson, A.T. & Knowles, L.L. (2011) Calibrating divergence time on species trees versus gene trees: implications for speciation history of *Aphelocoma* jays. *Evolution*, **65**, 184—202.
Meirmans, P.G. & Van Tienderen, P.H. (2004) GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes*, **4**, 792–794.
Mirams, A.G.K., Treml, E.A., Shields, J.L., Liggins, L. & Riginos, C. (2011) Vicariance and dispersal across an intermittent barrier: population genetic structure of marine animals across the Torres Strait land bridge. *Coral Reefs*, **30**, 937—949.
Palumbi, S.R. (1992) Marine speciation on a small planet. *Trends in Ecology and Evolution*, **7**, 114—118.
Palumbi, S.R., Grabowsky, G., Duda, T., Geyer, L. & Tachino, N. (1997) Speciation and population genetic structure in tropical Pacific Sea urchins. *Evolution*, **51**, 1506—1517.
Pennings, P.S., Achenbach, A. & Foitzik, S. (2011) Similar evolutionary potentials in an obligate ant parasite and its two host species. *Journal of Evolutionary Biology*, **24**, 871–886.
Posada, D. & Crandall, K.A. (1998) Modeltest: testing the model of DNA substituition. *Bioinformatics*, **14**, 817—818.
Pritchard, J.K., Stephens, M. & Donnelly, P. (2000) Inference of population structure from multilocus genotype data. *Genetics*, **155**, 945–959.
R Development Core Team (2013) R: A language and environment for statistical computing.
Rambaut, A. & Drummond, A.J. (2007) *Tracer*.
Rocha, L.A. & Bowen, B.W. (2008) Speciation in coral-reef fishes. *Journal of Fish Biology*, **72**, 1101—1121.

Rocha, L.A., Robertson, D.R., Rocha, C.R., Van Tassell, J.L., Craig, M.T. & Bowen, B.W. (2005) Recent invasion of the tropical Atlantic by an Indo-Pacific coral reef fish. *Molecular Ecology*, **14**, 3921—3928.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A. & Huelsenbeck, J.P. (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*,

Sambrook, J. & Russell, D. (2001) *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory Press, New York.

Sigman, D.M. & Hain, M.P. (2012) The biological productivity of the ocean. *Nature Education Knowledge*, **3**, 21.

Silva, G., Horne, J. & Castilho, R. (*in press*) Anchovies go north and west without losing diversity: post-glacial range expansions in a small pelagic fish. *Journal of Biogeography*,

Teacher, A.G.F. & Griffiths, D.J. (2011) HapStar: automated haplotype network layout and visualization. *Molecular Ecology Resources*, **11**, 151–153.

Whitehead, P.J., Nelson, W.S. & Wongratana, T. (1988) *Clupeoid fishes of the World (Suborder Clupeoidei)*. FAO, Rome.

Yabumoto, Y. (1988) Pleistocene clupeid and engraulidid fishes from Kokubu group in Kagoshima Prefecture, Japan. *Bulletin of Kitakyushu Museum Natural History*, 55-74.

SUPPORTING INFORMATION

Pleistocene climatic oscillations and the evolution of the Old World Anchovies *Engraulis spp.*

Gonçalo Silva, Regina Cunha and Rita Castilho

Table S1 Summary statistics across eight microsatellite loci of the Old World Anchovies. A_{rang}, allele range (bp); $\boldsymbol{A_r}$, average number of alleles across locations; *Effnum*, effective number of alleles; H_O, observed heterozygosity; H_E, expected heterozygosity; *H*t, total heterozygosity; *H*'t, corrected total heterozygosity; G_{IS}, inbreeding coefficient.

Nuclear microsatellites	**Genetic diversity**								
Name (acc. no.)	$\boldsymbol{A_{rang}}$	$\boldsymbol{A_r}$	$\boldsymbol{A_r}$**=9**	***Effnum***	$\boldsymbol{H_O}$	$\boldsymbol{H_E}$	***H*t**	***H*'t**	$\boldsymbol{G_{IS}}$
Ee 10 (AY241273)	197-381	85	62	8.179	0.854	0.895	0.947	0.950	0.045
Ee2-135 (FJ534738)	102-170	24	17	5.906	0.758	0.848	0.870	0.871	0.107
Ee2-407 (FJ534751)	128-296	62	38	7.292	0.816	0.880	0.922	0.925	0.073
Ee2-452a (FJ534754)	245-357	55	37	11.139	0.886	0.929	0.951	0.952	0.047
Ee2-508 (FJ534759)	154-198	29	26	7.213	0.589	0.884	0.927	0.930	0.334
Ee2-91a (FJ534732)	197-298	39	19	6.451	0.613	0.866	0.903	0.906	0.292
Ee2-91b (FJ534732)	100-165	19	13	4.851	0.809	0.809	0.818	0.818	-0.000
EJ41.1 (AF344659)	140-283	68	44	4.941	0.644	0.816	0.866	0.869	0.211

* Primers information and PCR conditions in Silva *et al.* (*in press*).

Table S2 - Accession numbers of species used on phylogenetic analyses

Species	Accession Number	Family	Sub-family	Native
Coilia brachygnathus	EU694410.1	Engraulidae	Coilinae	Pacific
Coilia ectenes	NC_019625.1	Engraulidae	Coilinae	Pacific
Coilia lindmani	AP011558.1	Engraulidae	Coilinae	Indo-Pacific
Coilia mystus	EU694407.1	Engraulidae	Coilinae	Pacific
Coilia nasus	EU694403.1	Engraulidae	Coilinae	Pacific
Coilia reynaldi	AP011559.1	Engraulidae	Coilinae	Indo-Pacific
Lycothrissa crocodilus	JQ012420.1	Engraulidae	Coilinae	Indo-Pacific
Setipinna melanochir	AP011565.1	Engraulidae	Coilinae	Indo-Pacific
Setipinna taty	JQ012365.1	Engraulidae	Coilinae	Indian
Setipinna tenuifilis	JQ012398.1	Engraulidae	Coilinae	Indo-Pacific
Thryssa baelama	AP009616.1	Engraulidae	Coilinae	Indo-Pacific
Thryssa dussumieri	JQ012363.1	Engraulidae	Coilinae	Indo-Pacific
Thryssa mystax	JQ012366.1	Engraulidae	Coilinae	Indo-Pacific
Amazonsprattus scintilla	JQ012351.1	Engraulidae	Engraulinae	Atlantic
Anchoa cayorum	JQ012346.1	Engraulidae	Engraulinae	Atlantic
Anchoa chamensis	JQ012375.1	Engraulidae	Engraulinae	Pacific
Anchoa colonensis	JQ012383.1	Engraulidae	Engraulinae	Atlantic
Anchoa cubana	JQ012342.1	Engraulidae	Engraulinae	Atlantic
Anchoa delicatissima	JQ012348.1	Engraulidae	Engraulinae	Pacific
Anchoa filifera	JQ012387.1	Engraulidae	Engraulinae	Atlantic
Anchoa lamprotaenia	JQ012379.1	Engraulidae	Engraulinae	Atlantic
Anchoa lyolepis	JQ012344.1	Engraulidae	Engraulinae	Atlantic
Anchoa mitchilli	JQ012357.1	Engraulidae	Engraulinae	Atlantic
Anchoa mundeoloides	JQ012419.1	Engraulidae	Engraulinae	Pacific
Anchoa nasus	JQ012373.1	Engraulidae	Engraulinae	Pacific
Anchoa panamensis	JQ012392.1	Engraulidae	Engraulinae	Pacific
Anchoa parva	JQ012377.1	Engraulidae	Engraulinae	Atlantic
Anchoa schofieldi	JQ012349.1	Engraulidae	Engraulinae	Pacific
Anchoa walkeri	JQ012369.1	Engraulidae	Engraulinae	Pacific
Anchovia clupeoides	EU552570.1	Engraulidae	Engraulinae	Atlantic
Anchovia macrolepidota	JQ012394.1	Engraulidae	Engraulinae	Pacific
Anchovia surinamensis	JQ012402.1	Engraulidae	Engraulinae	Atlantic
Anchoviella alleni	JQ012333.1	Engraulidae	Engraulinae	Atlantic
Anchoviella balboae	JQ012371.1	Engraulidae	Engraulinae	Pacific
Anchoviella brevirostris	JQ012412.1	Engraulidae	Engraulinae	Atlantic
Anchoviella carrikeri	JQ012330.1	Engraulidae	Engraulinae	Atlantic
Anchoviella elongata	JQ012381.1	Engraulidae	Engraulinae	Atlantic
Anchoviella guianensis	JQ012327.1	Engraulidae	Engraulinae	Atlantic
Anchoviella lepidentostole	JQ012414.1	Engraulidae	Engraulinae	Atlantic
Cetengraulis edentulus	JQ012385.1	Engraulidae	Engraulinae	Atlantic
Cetengraulis mysticetus	JQ012390.1	Engraulidae	Engraulinae	Pacific
Encrasicholina devisi	JQ012364.1	Engraulidae	Engraulinae	Indian
Engraulis anchoita	JQ012416.1	Engraulidae	Engraulinae	Atlantic
Engraulis australis	KJ007734	Engraulidae	Engraulinae	Pacific
Engraulis capensis	KF601464	Engraulidae	Engraulinae	Atlantic
Engraulis encrasicolus	JQ716644	Engraulidae	Engraulinae	Atlantic
Engraulis eurystole	JQ716748	Engraulidae	Engraulinae	Atlantic
Engraulis japonicus	KJ007642	Engraulidae	Engraulinae	Pacific
Engraulis mordax	JQ012350.1	Engraulidae	Engraulinae	Pacific
Engraulis ringens	JQ012426.1	Engraulidae	Engraulinae	Pacific
Jurengraulis juruensis	JQ012329.1	Engraulidae	Engraulinae	Atlantic
Lycengraulis batesii	JQ012326.1	Engraulidae	Engraulinae	Atlantic
Lycengraulis grossidens	JQ012396.1	Engraulidae	Engraulinae	Atlantic
Lycengraulis poeyi	JQ012370.1	Engraulidae	Engraulinae	Pacific
Pterengraulis atherinoides	JQ012323.1	Engraulidae	Engraulinae	Atlantic
Stolephorus chinensis	AP011566.1	Engraulidae	Engraulinae	Indo-Pacific
Stolephorus waitei	AP011567.1	Engraulidae	Engraulinae	Indo-Pacific
Chirocentrus dorab	AP006229.1	Chirocentridae		
Clupea harengus	NC_009577.1	Clupeidae		
Sardina pilchardus	NC_009592.1	Clupeidae		
Denticeps clupeoides	NC_007889.1	Denticipitidae		

Ilisha africana	NC_009584.1	Pristigasteridae
Sundasalanx mekongensis	AP006232.1	Sundasalangidae

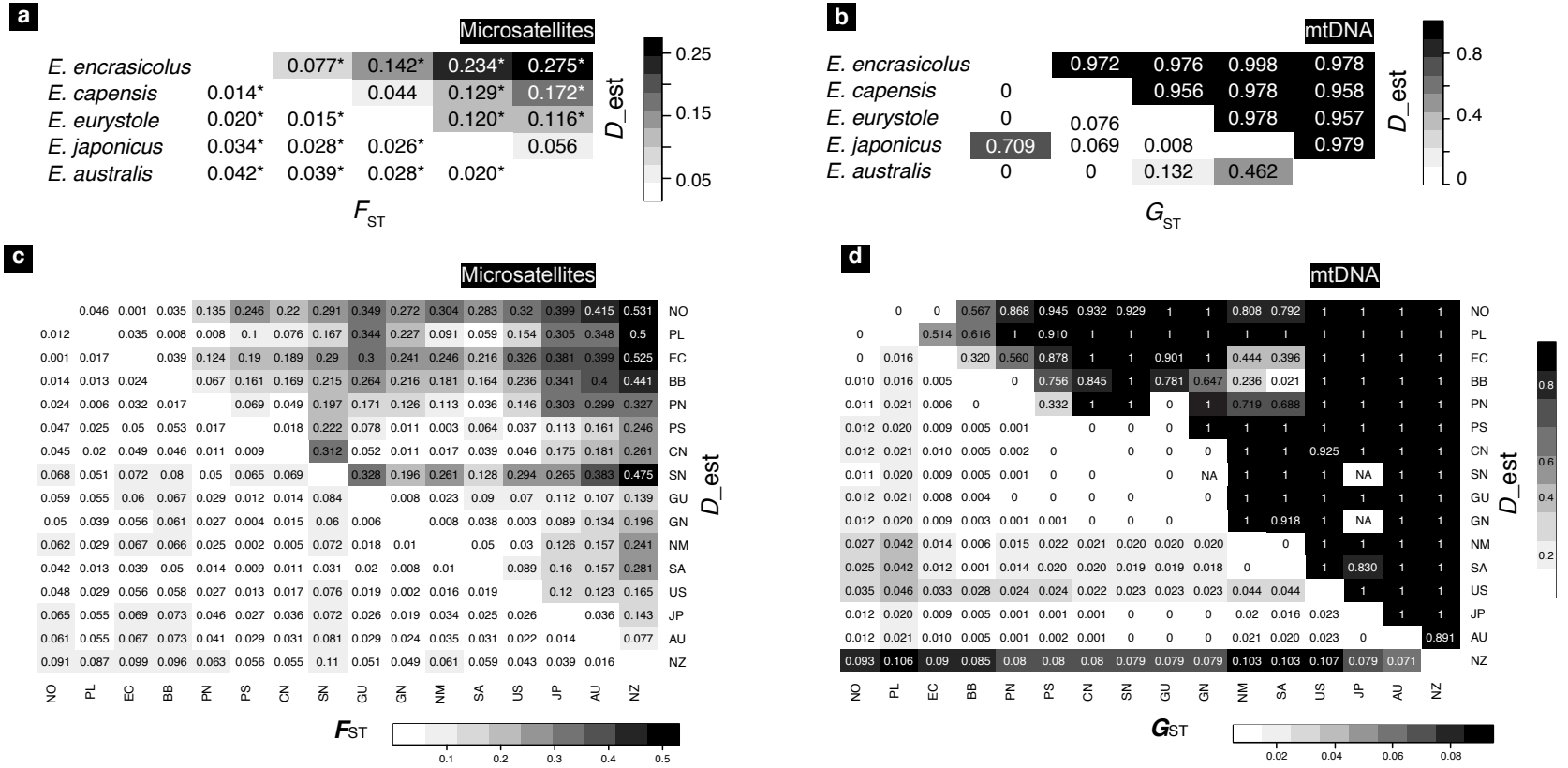


Figure S1 a) Pairwise-taxa F_{st} (below diagonal) and D_{est} values (above diagonal) for microsatellites; b) Pairwise-taxa G_{st} (below diagonal) and D_{est} values (above diagonal) for mtDNA; a) Pairwise-locations F_{st} (below diagonal) and D_{est} values (above diagonal) for microsatellites; b) Pairwise-locations G_{st} (below diagonal) and D_{est} values (above diagonal) for mtDNA; site abbreviations defined in Table 1.

CHAPTER V ● GENERAL CONCLUSIONS

This thesis aimed at unravelling the evolutionary history of the European anchovy (*Engraulis encrasicolus*) to understand the impact of past and present environmental fluctuations in shaping anchovies genetic diversity, population structure and historical demography. The European anchovy constitutes an interesting model organism for biogeographical studies as a marine small pelagic fish with broad distribution range, short generation time, large population sizes, high dispersal ability and high connectivity between populations. These characteristics allow studying the response of these type of organisms to climatic oscillations and to oceanographic processes. Given such biological traits, the European anchovy shows unexpected levels of population structure, clinal distribution of the mitochondrial clades, but low genetic differentiation from some of its congeners, despite large geographical distances. The species is part of a group designated as Old World anchovies, that includes, besides *E. encrasicolus*, other 4 nominal species, *E. eurystole*, *E. capensis* and *E. japonicus* and *E. australis*. We have included these entities, in order to have a more inclusive view of the species complex, sampling a total of 740 individuals from 23 locations worldwide. There is also a sixth more recent species (*E. albidus*) with an extremely restricted Mediterranean distribution, but no samples were possible to obtain. The numerous studies previously published were each individually rather restricted in geographic coverage of the eastern Atlantic distribution of the anchovies (Magoulas *et al.*, 1996; Grant, 2005; Magoulas *et al.*, 2006; Sanz *et al.*, 2008; Zarraonaindia *et al.*, 2009; Borrell *et al.*, 2012; Zarraonaindia *et al.*, 2012; Viñas *et al.*, 2013). The present work has provided results based on samples from Norway to South Africa, including samples from Guinea-Bissau, Ghana, Angola and Namibia. Although the Mediterranean was largely sampled before, we added two strategic samples in the North African Mediterranean (Tunisia) and eastern Levantine basin (Israel). Last but not least, we provide the first genetic results from the western Atlantic *E. eurystole*.

The most relevant contributions of this thesis are summarized below in separate sections, each corresponding to a specific topic.

LATITUDINAL DISPLACEMENTS, TRANS-OCEANIC COLONISATIONS AND INTER-OCEANIC TRANSITIONS

The palaeoclimatic oscillations challenged the physiologic conditions of anchovies and have driven important biogeographical shifts by imposing latitudinal displacements that avoided wholesale population reductions and genetic bottlenecks. Although anchovies show marked population structure, with a marked dual-clade mitochondrial structure and with strong geographic differentiation in some instances, their high capacity for dispersal allowed them to track suitable conditions and to rapidly colonise newly available habitats. At the northeastern Atlantic extreme range of distribution, the European anchovy was cyclically extirpated and forced to migrate southwards during cooling periods, while expanded northwards up to their physiological limits during periods of warming.

Climate shifts also provided the opportunity to anchovies to disperse to the western Atlantic. The western Atlantic putative *E. eurystole* was found to be conspecific with the eastern Atlantic counterpart *E. encrasicolus*. *Engraulis eurystole* has kept high genetic diversity, shares an haplotype with the Canary islands, and is most probably derived from the west and central coast African populations, mediated by the North Equatorial Current. To the best of our knowledge, it is the first time that an east-to-west transatlantic colonisation through the North Equatorial Current is proposed and supported by genetic evidence. The northeastern and northwestern Atlantic anchovy populations were likely recolonized rapidly since the LGM by large numbers of individuals, such that leading-edge populations retained most of the genetic diversity of parent populations.

The Atlantic OWA has most probably originated in Pacific waters. The most recent common ancestor of the European anchovy migrated to the Atlantic Ocean through the Indian Ocean, possibly in a stepping-stone colonisation along the Asian, Middle East and eastern African continental platforms. Depending on the calibration used, anchovies from the two oceanic basins diverged 0.62 Ma (Lessios *et al.*, 1999), coinciding with an interglacial period, with temperatures one to two degrees cooler than the present interglacial (Casper, 2010).

MTDNA AND NATURAL SELECTION

We found evidence that natural selection shaped mtDNA of the European anchovy, proposing that the cytochrome *b* gene is under positive selection in the anti-tropically distributed clade B. Temperature is the most relevant environmental predictor for clade geographical distribution contributing with 57% of the variance. However, no major structural changes in the conformation of the protein or evidence that the detected amino-acid substitution affected efficiency in the electron transfer process. The effect of selection compromises previous interpretations regarding the evolutionary history of clade B (Grant, 2005; Magoulas *et al.*, 2006; Borrell *et al.*, 2012), as selection may increase lineage coalescence times when compared to neutral expectations (Irwin, 2012), but it enriches the natural history of this species. The understanding of molecular adaptations to different environmental regimes may provide useful tools to understand how species will respond to future climate change.

PHYLOGEOGRAPHIC INFERENCES: GENETIC DIVERSITY, POPULATION STRUCTURE AND HISTORICAL DEMOGRAPHY

The genetic variation in the European anchovy is mostly determined by species life-history traits. Nevertheless, the effect of selection affecting vital genes, such as the mitochondrial cytochrome *b*, also contributes to shape the mitochondrial genetic architecture of this species. The contemporary distribution of mitochondrial clade frequencies revealed a latitudinal gradient of frequencies as a consequence of historical range shifts and by contemporary factors (e.g. selective sweeps that impose physiological limits to temperature tolerance). Despite anchovies have high ability for dispersal, the European anchovy clade B specimens revealed preference for local dispersal and reproduction, leaving genetics imprints of population structure even at neutral loci. Besides clade frequency shifting between geographic locations, no population structure was retrieved from mtDNA data. On the other hand, four major contemporary populations were identified in the Atlantic Ocean using neutral loci: 1) Tangier and Senegal; 2) Norway and the Baltic Sea southwards to Bay of Biscay; 3) western Atlantic and Canaries southwards to South Africa; 4) anchovies that inhabit areas between Portugal and the Mediterranean, including the Alboran Sea, constituting an intermediate group between the northern and the southern populations.

Nevertheless, the genetic break in the Bay of Biscay was detected in both mtDNA and nuclear microsatellites markers and is also common to many other marine invertebrate and vertebrate organisms, constituting a barrier to dispersal (Maggs *et al.*, 2008). The Bay of Biscay break is also consistent with the transition between the Northern European Seas and Lusitanian provinces (Spalding *et al.*, 2007).

Anchovies mtDNA revealed a slight decrease in haplotype diversities and allelic richness from the extremes of the distribution towards the tropics, but no decrease in heterozygosity was found for nuclear markers. Nevertheless, diversity values were generally high among molecular markers, concordant with species life-history traits. Imprints of paleoclimatic oscillations were observed at the northeastern extreme of the distribution range where populations show higher proportions of shared alleles/ haplotypes, contrasting to core populations (e.g. locations south to Bay of Biscay) that exhibited higher proportions of private alleles/ haplotypes.

The two mitochondrial clades of the European anchovy exhibit different demographic histories. Evidence supports demographic expansions older than the LGM in both clades, but the signal of expansion in clade B is less pronounced. Although strong climatic events (e.g. LGM) may erase past genetic imprints (Grant *et al.*, 2012), populations of the European anchovy were able to massively track suitable conditions, avoiding bottlenecks and wholesale reductions on genetic diversity. Clade B should exhibit a more pronounced signal of population expansion after the LGM, because it is more frequent in the recently deglaciated northeastern Atlantic. However, that was not observed, probably due to the effect of selection in clade B that could have increased mutation rate, masking a putative population expansion in the northern seas.

PHYLOGENETIC RELATIONSHIPS AND TAXONOMIC IMPLICATIONS AMONG THE OWA

Although the taxonomic review of the OWA complex was not a primary goal of this thesis, the phylogenetic approach based on molecular evidence contributed to shed light on that subject. The lack of reproductive isolation (shared mtDNA haplotypes and nuclear alleles) and morphological divergences among putative species of OWA support the collapse of this group complex into a single species (with two subspecies)

or two incipient species, one in the Atlantic Ocean and another in the Pacific Ocean.

Recent dispersers from the western Pacific into the Atlantic Ocean were depicted from shared genetic background between South African and Japanese anchovies (both mitochondrial haplotypes and nuclear alleles). Although the genetic compositions of anchovies from the two oceanic basins are largely distinct, the ongoing gene flow indicates that both pools still interbreed. This may be attributed to small coalescence time of divergence between anchovies from the two oceans and/ or to low migration rates that impedes species to evolve in complete isolation.

Within the Atlantic Ocean, the northwestern Atlantic *E. eurystole* is not monophyletic in either mtDNA or microsatellites, is most likely conspecific with *E. encrasicolus* and do not merit species status (Chapter II). In the eastern Atlantic, the nuclear data indicates that the individuals from *E. encrasicolus* mtDNA clades A and B interbreed, and also between these and *E. capensis*, supporting the view that these do not merit species status (Chapter IV). We propose that the distribution range of the European anchovy to be continuous from Norway and the Baltic Sea to South Africa, Mediterranean and Black Sea, and the western Atlantic. Because *E. encrasicolus* mtDNA clade B is under selection, its genetic constitution does not reflect the evolutionary history of the clade, and therefore there is no sense in estimating divergence time between the two mitochondrial lineages.

Engraulis albidus, the most recent nominal species will most probably not constitute a different taxonomic unit from *E. encrasicolus* for three main reasons. First, both taxa are pelagic with high dispersal ability and occur in sympatry, as far as the distribution ranges overlap. Although sympatric speciation is a possible process, anchovies life-history traits are not conducive to promote speciation in sympatry, except if selective pressures impose divergence (Irwin, 2012). The alternative hypothesis would be that the species has formerly appeared in isolation and present-day sympatry is a consequence of secondary contact. This explanation was previously suggested for the existence of *E. encrasicolus* mtDNA clades A and B (Magoulas *et al.*, 2006). Second, individuals far apart such as those from the eastern and the western Atlantic or from Norway and South Africa do not seem to be reproductively isolated (Chapter IV), the likelihood of *E. albidus* and *E. encrasicolus* being separate taxonomic units is quite low. Third, *Engraulis albidus* was designated as "inshore" due to the preference for brackish waters, estuaries and coastal lagoons, but *E. encrasicolus* also tolerates very low salinities (e.g. Baltic Sea). There are also no clear

distinctive morphological characters found, except some slightly differences in "raw multivariate scores from canonical analysis" or in the "mean ratio of head depth to head length" (Borsa *et al.*, 2004). Differences in allele frequencies of two nuclear introns were found between inshore and offshore populations, but *E. albidus* typical alleles were also found in South Africa (Bouchenak-Khelladi *et al.*, 2008), indicating past introgression and long distance dispersal. Moreover, within *E. encrasicolus* distribution, some locations were found to have different alleles composition (e.g. Tangier and Senegal), constituting distinct genetic assemblages, but not different species (Chapters II). If *E. albidus* constitutes a different species from *E. encrasicolus*, it would represent a rare instance of ecological divergence in a pelagic species with large distribution range (Grant *et al.*, 2005).

Within the Pacific Ocean, two haplotypes from Japan are more genetically similar to the Australian lineage than to the Japanese lineage. These individuals may represent an ancient colonization of anchovies from Australia into the northwestern Pacific that persisted in the Japanese waters through several climatic cycles (Chapter IV). At the nuclear level the genetic pools revealed contemporary admixture between the two hemispheres (Chapter IV) and supported Grant & Bowen (2006) results. The distribution range of each nominal species is not well known, particularly at low latitudes, which led previous authors to assume a distributional gap between the two distribution areas, separated by a putative "temperature barrier" to dispersal. However, there are few records in Philippines and Indonesian waters (Whitehead *et al.*, 1988; Froese & Pauly, 2011), indicating a possible stepping-stone path or even permanent less abundant populations.

CONCLUDING REMARKS AND FUTURE PERSPECTIVES

Small pelagic fishes represent about 20 - 25 % of the total annual world fisheries catch (Alheit *et al.*, 2012) and are extremely vulnerable to environmental changes and human pressures (Grant & Bowen, 1998). Overfishing not only has a major impact on species abundance, but also contributes to the reduction of the genetic diversity (Pinsky & Palumbi, 2014). It is therefore pivotal a better understanding of the biology and ecology of these species. In the last 30 years, the European anchovy was highly studied in terms of its genetic background, mostly focused on population structure and historical demography, but restricted to small

geographic areas. In this thesis, we extended previous sampling areas and used both nuclear and mitochondrial markers to shed light on patterns and processes that shaped the genetic variation of the European anchovy along its distribution range. Our data suggests that the European anchovy: 1) responds to climate changes by massively tracking suitable environmental conditions to prevent population bottlenecks and reductions on genetic diversity; 2) mtDNA is under positive selection having impact on overall performance of dispersal with consequences to population structure; 3) life-history traits likely prevent speciation within the Atlantic Ocean, even in partially isolated populations such as the western Atlantic or in restricted sea basins such as the Baltic Sea or the Mediterranean Sea.

For a better understanding how anchovies will respond both to climate changes and human pressures, it is essential to study in detail both historical and contemporary patterns and mechanisms that shape the European anchovy genetic architecture. More sampling locations are needed, especially in the western Atlantic and all around Africa, including the eastern coast where some records were found, but little is known about the presence of the species. The application of new molecular techniques and high-throughput sequencing will provide a better resolution on population structure, migration patterns, genetic diversity and demography. Besides this, protein expression essays will likely be the next step to unveil the response of the European anchovy clade B to natural selection, in particular how genes are expressed at different temperature regimes. For a solid revision of the OWA, both morphologic and genetic characterizations need to be performed, including specimens from putative *E. albidus*. In terms of molecular characterization, multilocus phylogenies could provide reliable results to resolve the number of taxonomic units within this complex and the relationship between them.

REFERENCES

Alheit, J., Pohlmann, T., Casini, M., Greve, W., Hinrichs, R., Mathis, M., O'Driscoll, K., Vorberg, R. & Wagner, C. (2012) Climate variability drives anchovies and sardines into the North and Baltic Seas. *Progress In Oceanography*, **96**, 128—139.

Borrell, Y.J., Piñera, J.A., Sánchez Prado, J.A. & Blanco, G. (2012) Mitochondrial DNA and microsatellite genetic differentiation in the European anchovy *Engraulis encrasicolus* L. *ICES Journal of Marine Science*, **69**, 1357–1371.

Borsa, P., Collet, A. & Durand, J.D. (2004) Nuclear-DNA markers confirm the presence of two anchovy species in the Mediterranean. *Comptes Rendus Biologies*, **327**, 1113–1123.

Bouchenak-Khelladi, Y., Durand, J.D., Magoulas, A. & Borsa, P. (2008) Geographic structure of European anchovy: a nuclear-DNA study. *Journal of Sea Research*, **59**, 269–278.

Casper, J.K. (2010) *Global warming cycles: ice ages and glacial retreat*, 1st edn. Infobase Publishing, New York, USA.

Froese, R. & Pauly, D. (2011) *FishBase*. Available at: http://www.fishbase.org (accessed 04-12-2013 2013).

Grant, W.S. (2005) A second look at mitochondrial DNA variability in European anchovy (*Engraulis encrasicolus*): assessing models of population structure and the Black Sea isolation hypothesis. *Genetica*, **125**, 293–309.

Grant, W.S. & Bowen, B.W. (1998) Shallow population histories in deep evolutionary lineages of marine fishes: insights from sardines and anchovies and lessons for conservation. *The Journal of Heredity*, **89**, 415-426.

Grant, W.S. & Bowen, B.W. (2006) Living in a tilted world: climate change and geography limit speciation in Old World anchovies (*Engraulis*; Engraulidae). *Biological Journal of the Linnean Society*, **88**, 673–689.

Grant, W.S., Leslie, R.W. & Bowen, B.W. (2005) Molecular genetic assessment of bipolarity in the anchovy genus *Engraulis*. *Journal of Fish Biology*, **67**, 1242–1265.

Grant, W.S., Liu, M., Gao, T. & Yanagimoto, T. (2012) Limits of Bayesian skyline plot analysis of mtDNA sequences to infer historical demographies in Pacific herring (and other species). *Molecular Phylogenetics and Evolution*, **65**, 203–212.

Irwin, D.E. (2012) Local adaptation along smooth ecological gradients causes phylogeographic breaks and phenotypic clustering. *American Naturalist*, **180**, 35–49.

Lessios, H.A., Kessing, B.D., Robertson, D.R. & Paulay, G. (1999) Phylogeography of the pantropical sea urchin *Eucidaris* in relation to land barriers and ocean currents. *Evolution*, **53**, 806—817.

Maggs, C.A., Castilho, R., Foltz, D., Henzler, C., Jolly, M.T., Kelly, J., Olsen, J., Perez, K.E., Stam, W., Vainola, R., Viard, F. & Wares, J. (2008) Evaluating signatures of glacial refugia for North Atlantic benthic marine taxa. *Ecology*, **89**, S108–S122.

Magoulas, A., Tsimenides, N. & Zouros, E. (1996) Mitochondrial DNA phylogeny and the reconstruction of the population history of a species: the case of the European anchovy (*Engraulis encrasicolus*). *Molecular Biology and Evolution*, **13**, 178–190.

Magoulas, A., Castilho, R., Caetano, S., Marcato, S. & Patarnello, T. (2006) Mitochondrial DNA reveals a mosaic pattern of phylogeographical structure in Atlantic and Mediterranean populations of anchovy (*Engraulis encrasicolus*). *Molecular Phylogenetics and Evolution*, **3**, 734–746.

Pinsky, M.L. & Palumbi, S.R. (2014) Meta-analysis reveals lower genetic diversity in overfished populations. *Molecular Ecology*, **23**, 29—39.

Sanz, N., Garcia-Marín, J.L., Viñas, J., Roldán, M. & Pla, C. (2008) Spawning groups of European anchovy: population structure and management implications. *ICES Journal of Marine Science*, **65**, 1635–1644.

Spalding, M.D., Fox, H.E., Allen, G.R., Davidson, N., Ferdaña, Z.A., Finlayson, M., Halpern, B.S., Jorge, M.A., Lombana, A., Lourie, S.A., Martin, K.D., McManus, E., Molnar, J., Recchia, C.A. & Robertson, J. (2007) Marine ecoregions of the world: a bioregionalization of coastal and shelf areas. *BioScience*, **57**, 573–583.

Viñas, J., Sanz, N., Peñarrubia, L., Araguas, R.-M., García-Marín, J.-L., Roldán, M.-I. & Pla, C. (2013) Genetic population structure of European anchovy in the Mediterranean Sea and the northeast Atlantic Ocean using sequence analysis of the mitochondrial DNA control region. *ICES Journal of Marine Science*,

Whitehead, P.J., Nelson, W.S. & Wongratana, T. (1988) *Clupeoid fishes of the World (Suborder Clupeoidei)*. FAO, Rome.

Zarraonaindia, I., Pardo, M.A., Iriondo, M., Manzano, C. & Estonba, A. (2009) Microsatellite variability in European anchovy (*Engraulis encrasicolus*) calls for further investigation of its genetic structure and biogeography. *ICES Journal of Marine Science*, **66**, 2176–2182.

Zarraonaindia, I., Iriondo, M., Albaina, A., Pardo, M.A., Manzano, C., Grant, W.S., Irigoien, X. & Estonba, A. (2012) Multiple SNP markers reveal fine-scale population and deep phylogeographic structure in European anchovy (*Engraulis encrasicolus* L.). *PLoS ONE*, **7**, e42201.