Universidad de Concepción
Dirección de Postgrado
Facultad de Ciencias Naturales y Ocanográficas
Programa de Doctorado en Sistemática y Biodiversidad

Evaluando los factores ambientales que determinan la diversificación a escala microevolutiva y macroevolutiva: *Sprattus fuegensis*, género *Sprattus*, y Orden Clupeiformes como modelos de estudio

CRISTIAN BRAULIO CANALES AGUIRRE
CONCEPCIÓN-CHILE
2014

Profesor Guía: Dr. Cristián Hernández Ulloa
Dpto. de Zoología, Facultad de Ciencias Naturales y Ocenográficas

Profesor Co-Guía: Dr. Ricardo Galleguillos Gonzalez
Dpto. de Oceanografía, Facultad de Ciencias Naturales y Ocenográficas

Universidad de Concepción

Acknowledgements

Primero que todo quisiera agradecer a mi familia quienes a pesar de preguntase porque sigo estudiando entendieron que me gusta y me apoyaron en todo momento.

A mi profesor guía Cristian E. Hernández y co-guia Ricardo Galleguillos de quienes agradezco desde al 2005 por el apoyo y confianza que me otorgaron, y además por las enseñanzas y conversaciones motivacionales tenidas. Muchas de esas me dieron ánimos para seguir continuando este camino.

A los miembros de mi comité, Daniel Gomez-Uchida, y Antonio Brante que se dieron en revisar la tesis en un corto plazo de tiempo y por lo buenos comentarios y preguntas realizadas en la defensa publica que sin duda mejoraran las versiones posteriores de los manuscritos.

Al laboratorio de Genética y Acuicultura a sus miembros pasados y actuales (Vicky) y especialmente a Sandra Ferrada quien ha sido un apoyo importante en desenvolverme como científico desde mis inicios.

Al laboratorio de Ecología Evolutiva y Filoinformática, especialmente a Jorge Avaria (Tinti), Oscar Inostroza, Reinaldo Rivera por su apoyo y colaboración en esta última etapa de la tesis, a Sandra Sampertegui y Enrique Rodríguez por siempre darme ánimos, apoyo y buenos comentarios.

A Christian Muñoz (Charlie) compañero magallánico que brindo apoyo y buenas conversaciones.

A Peter Ritchie, por darme la oportunidad de pasar tres meses en su laboratorio de Victoria University of Wellington, Nueva Zelanda, a Jeff Shima, y Abigail Powell del Victoria University Coastal Ecology Laboratory (VUCEL), a Tom Trnski del Auckand War Memorial Museum en Auckland, Nueva Zelanda, y a Malcolm Francis del National Institute of Water and Atmospheric Research (NIWA).

A Sebastián Hernández con quien compartimos y conversamos muchas veces en mi estadía en Nueva Zelanda. Además por todos los contactos realizados para conseguir muestras de *Sprattus*.

A mis amigos, Edwin von Jentschyk, Paulina Arancibia, Carlos Jara y Sebastian Muñoz, por su apoyo.

Finalmente agradecer a Fernanda Oyarzun que sin duda, su constante apoyo, cariño y confianza me dieron aun más energía para terminar la tesis. Agradecerle además por las muchas conversaciones referente a su experiencia en posgrado y por la revisión del inglés de cada manuscrito.

Financiamiento: Fondo de Investigación Pesquera (FIP 2010-17) (PI Ricardo Galleguillos), FONDECYT 11080110 (PI Cristián Hernández) y a las Becas de pasantía doctoral al extranjero CONICYT.

Dedication

A mi hijo Luciano Canales

Universidad de Concepción

Abstract

Underlying marine environmental factors driving diversity patterns at several levels of their taxonomic hierarchy: *Sprattus fuegensis* species, *Sprattus* Genus and, Clupeiformes Order as study models

Cristian Braulio Canales Aguirre

Mentors:
Dr. Cristián E. Hernández
Dr. Ricardo Galleguillos

En esta tesis doctoral se evaluó como factores ambientales oceanográficos actuales y pasados están asociados con la diversidad genética, estructuración poblacional, demografía histórica y tasas de diversificación evolutiva a una escala microevolutiva y macroevolutiva. Se identificó un problema en la sistemática del genero *Sprattus* evaluando sus relaciones filogenéticas y biogeografía histórica. Se incorporó por primera vez en conjunto, análisis que incorporan el método comparativo filogenético para evaluar la señal filogenética histórica de variables y su relación con las tasa de diversificación de la especie *S. fuegensis* y del Orden Clupeiformes. De esta manera se construyó una metodología robusta para responder la pregunta si variables ambientales juegan un rol importante a diferentes niveles taxonómicos en peces marinos, tomando como modelos el Orden Clupeiformes y la especie *S. fuegensis*. Las variables ambientales que mostraron asociación con características genéticas de la especie fueron: temperatura, salinidad, oxigeno, pH, clorofila y disminución pasada del nivel del mar. Se observó que en diferentes sub-familias y familias las tasas de diversificación fueron variables, sugiriendo que no todos lo grupos taxonómicos han respondido de igual manera a los cambios u oscilaciones de factores ambientales oceanográficos. Finalmente y dado los resultados obtenidos en los diferentes capítulos de esta tesis, se demostró claramente que factores ambientales oceanográficos juegan un rol importante en la historia evolutiva desde especies hasta niveles taxonómicos mayores, y expande nuestro entendimiento de las variables ambientales a diferentes categorías taxonómicas.

Table of Contents

List of Figures

List of Tables

General Introduction (Spanish)

Los ambientes marinos presentan una alta riqueza de especies, sin embargo es difícil evaluar cuales son los factores que explican a esta diversidad debido a la ausencia de barreras geográficas "duras" que tiendan a promover especiación (Klanten 2003). La especiación es un proceso continuo desde población a especie, en donde la diferenciación genética resultante entre poblaciones aisladas puede contribuir a la diversidad dentro de especies (e.g. cambios en frecuencias alélicas o haplotipos), resultando potencialmente a largo plazo en generación de nuevas especies (e.g. cambios a nivel genómico que produzcan aislamiento reproductivo) (Grant et al. 2010). La diversificación en un contexto macroevolutivo puede definirse como especiación menos extinción, donde la tasa de diversificación dependerá de las tasas de especiación y extinción (Liow et al. 2011). La diversificación puede ser promovida por factores bióticos (Red Queen Hypothesis) o por factores abióticos (Court Jester Hypothesis), o una mezcla de ambas (Benton 2009). La diversificación de un grupo taxonómico puede estar asociado a la interacción de individuos tanto a nivel intra como inter-específico (Benton 2009, Liow et al. 2011), y a factores ambientales (Liow et al. 2011).

A nivel intraespecífico la diversificación genética estaría dada por la aparición o desaparición de haplotipos o alelos dentro de una población. De esta manera, cambios en frecuencias genotípicas o haplotipícas pueden generar diferencias poblacionales y así una diversificación genética intraespecífica, esto potencialmente ligado a la aparición de barreras geográficas presentes en el área de distribución (e.g. factores ambientales físicos o químicos), así como la a capacidad de dispersión que pueden tener las especies. Por otro lado, a nivel interespecífico, el conocimiento de cómo ha sido la dinámica de diversificación en diferentes niveles de organización biológica (especies, géneros, familias, etc.) producto de factores ambientales en ambientes marinos, nos permitiría entender el origen de la biodiversidad que hoy muestran estos ambientes. Además, este conocimiento puede ayudar a determinar si los factores ambientales que actúan a nivel

macroevolutivo, son los mismos que promueven la diversificación a nivel microevolutivo (diversificación genética y diferenciación poblacional).

En organismos marinos, el Orden Clupeiformes, que lo componen peces pelágicos como anchovetas y sardinas, presenta 364 especies (Whitehead 1988). Para el caso de los géneros que presenta el Orden Clupeiformes se sabe que su diribución diferentes distribuciones geográficas, distribución en el océano Pacífico, Atlántico e Indico y algunos mundialmente (Whitehead 1988). Algunos de ellos presentan distribución anti-tropical, donde no están claras las rutas de dispersión que tuvieron que sortear para colonizar nuevas áreas y por ende diversificar. Además, los Clupeiformes presentan especies que actualmente se distribuyen en áreas que han sido afectadas fuertemente por cambios ambientales drásticos como los ocurridos en el Pleistoceno (e.g. *Clupea pallasi*, *Sprattus sprattus*, *Sprattus fuegensis*) (Debes et al. 2008, Liu et al. 2011). Estas características dan cuenta de que el Orden Clupeiformes es un buen modelo de estudio para evaluar cuáles son los principales factores ambientales que dan cuenta de la diversidad del grupo. Además, a nivel intra-genérico, el género *Sprattus* es un buen modelo para evaluar rutas de dispersión y posterior diversificación genética intra-genérica. Por último, el caso de la especie *S. fuegensis* es un buen modelo para evaluar el efecto de factores ambientales, históricos (glaciaciones pleistocénicas) y contemporáneos (parámetros ambientales actuales) sobre parámetros genéticos intraespecíficos (e.g. linajes genéticos y divergencia poblacional).

Características del género *Sprattus* y *S. fuegensis*

Dentro de la subfamilia Clupeinae, las especies del género *Sprattus* son las que presentan una mayor distribución geográfica, con representantes en costas de cuatro continentes. El género *Sprattus* presenta peces pelágicos que forman cardúmenes, habitan cerca de la costa y pueden tolerar aguas de baja salinidad. Sus miembros presentan una distribución anti-tropical, encontrándose en el hemisferio Norte (Europa) y en el

hemisferio Sur (Australia, Nueva Zelanda y el cono sur de Chile y Argentina). Las especies que componen el género *Sprattus* son: *S. sprattus* (Linnaeus, 1758), *S. antipodum* (Hector, 1872), *S. muelleri* (Klunzinger, 1880), *S. novaehollandiae* (Valenciennes, 1847), y *S. fuegensis* (Jenyns, 1842). La especie *S. sprattus* es la especie que presenta mayor amplitud en su distribución geográfica, esta en el océano Atlántico Noreste, desde el mar del Norte y Báltico, al sur de Marruecos, también en el mar Mediterráneo, el mar Adriático, y el mar Negro. La especie *S. antipodum* se distribuye en las Costas de Nueva Zelanda, específicamente en la costa este de la Isla del Norte, el Estrecho de Cook, y al parecer, toda la costa del sur de la Isla Sur, de la Isla Stewart y el Estrecho Foveaux. La especie *S. muelleri* se distribuye en las costas de Nueva Zelanda (al este y oeste de las costas del norte y las islas del sur, posiblemente al sur del Estrecho de Foveaux e incluso en la isla Aukland. La especie *S. novaehollandiae* se distribuye en las costas de Tasmania, el Estrecho de Bass, sudeste de Australia al norte de Sydney. Por último, la especie *S. fuegensis* se distribuye en el océano Atlántico Suroeste (alrededor de 40°S hasta Tierra del Fuego, también en las islas Falkland / Malvinas), y por el océano Pacífico Suroriental desde el mar interior de Chiloé hacia Tierra del fuego (Whitehead, 1988; Aranis et al. 2006, 2007; Niklitschek et al. 2009). A nivel genérico, las relaciones filogenéticas entre las especies que componen el género *Sprattus* son desconocidas. Por otro lado, para el género *Sprattus*, mediante datos genéticos solo hay registros de diferenciación de dos especies del género mediante alozimas, *S. muelleri* (Klunzinger, 1880*)* y *S. antipodum* (Hector, 1872*)* (Whitehead et al. 1985).

El modelo de estudio a nivel microevolutivo es *S. fuegensis*, es una especie de importancia comercial en Chile y Argentina, con desembarques en Chile de 52602 toneladas el 2009 (Sernapesca 2010), aunque para Argentina no se conoce estadística pesquera. Es un desovante parcial con tres pulsos reproductivos por estación, con una longitud de madurez dentro del rango de 11 - 12 cm de largo total (Aranis et al. 2006, Niklitschek et al. 2009) y una edad de primera madurez entre 2 a 3 años (Hansen 1999). *S. fuegensis* realiza migraciones verticales circadianas; durante el día presenta agregaciones compactas tipo cardumen a mayor profundidad, que tienden a dispersarse

durante la noche hacia la superficie, en agregaciones de distintas densidades (Aranis et al. 2006). Es una especie pelágica y zooplanctófaga a lo largo de todo su ciclo vital, distribuida en las zonas costeras sur-australes de Chile y Argentina, principalmente en los primeros 50 metros de la columna de agua (Madirolas & Hansen 2000, Sabatini et al. 2001, Aranis et al. 2006).

Efecto de factores abióticos sobre la diversificación del Orden Clupeiformes y origen biogeográfico del género *Sprattus*

Los Clupeiformes están distribuidos a nivel mundial con 364 especies (Nelson, 2006). Muchos de ellas habitan en aguas marinas costeras tropicales y subtropicales, además de presentar varios grupos euralinos y anádromo (Whitehead, 1985). Por otro lado, también se pueden encontrar linajes de Clupeidos exclusivamente de ambientes dulceacuícolas (Whitehead, 1985; Whitehead et al., 1988), lo cual permitiría estudiar los mecanismos de transición dulceacuícola/marino y sus adaptaciones fisiológicas (Lavoué et al. 2007), así como los cambios en factores abióticos que pudieron derivar en esta diversificación específica dentro del orden. En un contexto macroevolutivo y a nivel taxonómico mayor, el Orden Clupeiformes corresponde a un grupo natural (Lavoué et al. 2007), en el cual algunas de las familias que la componen corresponden a grupos monofiléticos (Li & Ortí 2007; Lavoué et al. 2007). Se ha estimado, basado en la hipótesis del reloj molecular que el tiempo de divergencia del grupo Clupeidae oscila entre los 0.12 a 13.45 millones de años (Ma) (Cheng & Lu 2006). Esta familia esta distribuida en cuatro sub-familias: Dussumieriinae, Clupeinae, Pellonulinae, Alosinae y Dorosomatinae, las cuales presentan caracteres diagnósticos distintivos (Whitehead 1988). La sub-familia Clupeinae es la que presenta un mayor número de especies, distribuidas en 16 géneros: *Clupeonella*, *Sardina*, *Harengula*, *Herklotsichthys*, *Amblygaster*, *Sardinella*, *Escualosa*, *Platanichthys*, *Ramnogaster*, *Rhinosardinia*, *Lile*, *Strangomera, Sardinops*, *Opisthonema*, *Clupea y Sprattus* (Whitehead 1988). Si bien, hay estudios que resuelven en parte las relaciones filogenéticas del Orden y las familias que lo componen (Li & Ortí 2007; Lavoué et al. 2007), estos no dan cuenta del origen y

los factores que podrían estar asociados a la diversificación de este grupo. Por otro lado, se ha sugerido que el débil soporte de los principales linajes de Clupeiformes podría ser efecto de un corto periodo de diversificación entre ellos (Lavoué et al. 2008). Diversificación que podría estar asociada a cambios en características abióticas del ambiente donde se distribuyen.

A nivel genérico, se ha indicado que *Sprattus* es el género hermano de *Clupea* (Jérôme et al. 2003, Lavoué et al. 2007), formando un clado monofilético muy divergente dentro del grupo de los Clupeidae (Cheng & Lu, 2006). Este clado es el grupo hermano de los géneros *Alosa*, *Sardinops*, *Sardina*, *Nematalosa*, *Dorosoma*, *Sardinella*, *Ethmalosa*, *Odaxothrissa* y *Pellonula* (Lavoué et al. 2007). Aunque el tiempo de divergencia estimado entre *Sprattus* y *Clupea* dependerá de método de calibración que se utilice, este se encuentra entre los 2.66 y 6.75 Ma (Jérôme et al. 2003, Cheng & Lu 2006), periodo en donde ocurrieron cambios ambientales producto de las glaciaciones Pleistocénicas (e.g. Ruzzante et al. 2008), además de cambios geográficos oceanográficos importantes (e.g. cierre istmo de Panamá, Farris et al 2011; formación del Sistema de corrientes de Humboldt, Camus et al. 2001). Antecedentes en cuanto al origen y diversificación de algunos otros miembros en Clupeiformes son escasos. De los pocos estudios disponibles, se ha propuesto que algunos miembros del Orden presentan un origen antiguo en el Océano Atlántico (i.e. *Engraulis*: Nelson 1985), o producto de la separación de Norteamérica y Europa (*Alosa*: Faria et al. 2006), o producto de cambios en factores abióticos en el último máximo glacial (UMG) (*Sardina* y *Sardinops*: Parrish et al. 1989). El origen, diversificación y rutas de dispersión del género *Sprattus* es aún incierto, y dilucidarlo podría dar luces sobre los mecanismos que pudieron colaborar en la diversificación de las especies que componen el genero, lo cual podría estar asociado a cambios en factores ambientales de su hábitat.

Efecto de las glaciaciones pleistocénicas sobre la diversificación de linajes genéticos en *S. fueguensis*

En Sudamérica es escaso el conocimiento del efecto de los ciclos glaciares y los cambios ambientales asociados (factores abióticos) sobre las especies marinas que se distribuyeron en sus costas (Ruzzante et al. 2008). A la fecha se ha estimado que el Último Máximo Glacial (UMG), data entre 23 – 25 mil años (ka) (Sudgen et al. 2005) y que la influencia de estos cambios ambientales, como los ocurridos en el periodo Cuaternario dan cuenta de alteraciones en la distribución geográfica, y la estructuración genética de las especies en ambientes terrestres (Ruzzante et al. 2011). Sin embargo, es poco conocido el efecto de los ciclos glaciares sobre la fauna marina, el grado en que las poblaciones han disminuido su tamaño o las rutas a través del cual recolonizaron (Ruzzante et al. 2008). Especies marinas que se encuentran distribuidas actualmente en áreas geográficas que han sido afectadas por glaciares pueden dar cuenta de patrones genéticos diferenciales, y de esta forma, los patrones observados a nivel intra-específico o inter-específico de divergencia podrían estar ligados a cambios en el ambiente, producto de estos eventos glaciares durante el Pleistoceno (Ruzzante et al. 2006). En consecuencia, estos cambios ambientales dejarían huellas genéticas en los individuos, influenciando su historia demográfica (Avise 2000), así como la diversificación de linajes genéticos. Dentro de los efectos demográficos históricos producto de estas oscilaciones climáticas podemos encontrar, disminuciones del tamaño efectivo poblacional (N_E), poblaciones estables que han mantenido su N_E en el tiempo, cuellos de botella recientes, eventos fundadores, cuellos de botella seguidos de un rápido crecimiento poblacional y acumulación de mutaciones, y fuertes cuellos de botella en poblaciones grandes y estables o contacto secundarios (Grant & Bowen 1998).

Cambios demográficos en una especie pueden modificar el número y frecuencia de los haplotipos presentes en la distribución geográfica afectada por las glaciaciones (e.g. Ruzzante et al. 2006), lo cual puede ser visto como una diversificación de linajes genéticos. Esto último, tomando en cuenta que producto de cambios en el ambiente se pueden extinguir o generar nuevos haplotipos o modificar su frecuencia,. Estos cambios demográficos, tienen una fuerte repercusión en la diversificación de linajes genéticos, por

ejemplo, disminuciones en el N_E, y fuertes cuellos de botella o colonización de nuevas áreas, favorecen a la deriva génica, la cual contribuye al aislamiento reproductivo, incluso en presencia de flujo génico (Uyeda et al. 2009), lo que posteriormente podría generar nuevas especies (Butlin et al. 2012).

La costa de Chile desde los 41ºS hacia el sur fue una zona afectada directamente por el avance y retroceso de los hielos, así esta zona estuvo cubierta de hielo, los cuales posteriormente retrocedieron y formaron los fiordos y canales que actualmente observamos (Paskoff 2010). Este fenómeno pudo tener un importante efecto en las poblaciones de organismos acuáticos que se distribuyen en ese sector, como ocurrió en peces dulceacuícolas *Galaxias, Percichthys, y Percilia* (Zemlak et al. 2011, Ruzzante et al. 2006, Ruzzante et al. 2008), el molusco gastropodo *Concholepas concholepas* (Cárdenas et al. 2009), el alga marina *Durvillaea antarctica* (Frazer et al. 2010), y el pez marino *Eleginops maclovinus* (Ceballos et al. 2011).

En Clupeidos, se ha establecido el efecto de las glaciaciones sobre los linajes genéticos de estos, sugiriéndose que tanto sardinas como anchovetas son muy vulnerables a fluctuaciones poblacionales relacionadas con cambios climáticos (Chavez et al. 2003). No obstante, la mayoría de los trabajos se han realizado en especies de Clupeidos del hemisferio norte. Por ejemplo, en *Engraulis japonicus* y *Engraulis australis*, se sugiere que las poblaciones tuvieron una expansión poblacional en el Pleistoceno tardío, producto de oscilaciones climáticas (Liu et al. 2006). Aunque sólo en *E. australis* se encontró más de un linaje genético (Liu et al. 2006). Así mismo, este periodo afectó a otras especies como *Engraulis mordax* y *Sardinops sagax*, en donde se ha indicado una disminución poblacional severa y un evento fundador en la corriente de California durante el Pleistoceno tardío, respectivamente (Lecomte et al. 2004). Por otro lado, en *Sardina pilcharus*, se confirma la presencia de dos posibles subespecies: (i.e. *S. p. pilcharus* y *S. p. sardina),* así como un linaje divergente, el cual puede estar relacionado con un aislamiento y la acción de la deriva génica, producto de un cuello de botella reciente

ocurrido en el periodo pleistocénico (Atarhouch et al. 2006). En *C. pallasii*, tres linajes evolutivos se han detectado, y dos de ellos pudieron ser originados por una separación vicariante asociada a ciclos glaciares (Liu et al. 2011). Finalmente en *S. sprattus*, dos linajes se detectaron en donde la deglaciación producto de un aumento en la temperatura pudo jugar un rol importante en la generación de estos linajes (Debes et al. 2008). Para todos los casos, la diversificación de linajes genéticos esta asociada a la aparición de eventos vicariantes, disminución de flujo génico y cambios demográficos derivados de la última glaciación, así como los cambios climáticos ocurridos en el Pleistoceno en general.

Efecto de factores abióticos contemporáneos sobre la diversificación poblacional de *S. fuegensis*

Por un lado, tenemos en cuenta que barreras históricas, asociadas a eventos vicariantes en el pasado (en escala microevolutiva), podrían afectar la conectividad genética, y de esta manera la historia demográfica de las especies, generando diversificación de linajes genéticos. Por otro lado, factores ambientales actuales (e.g. factores físicos) del paisaje de la distribución de una especie también podrían actuar como barreras al flujo génico, pudiendo generar cambios en las frecuencias alélicas entre poblaciones de una especie, diversificando a nivel microevolutivo. Por esta razón, no solo es importante la detección de discontinuidades genéticas, sino también su potencial correlación con características ambientales presentes en la distribución de la especie, es decir, es necesario entender cómo las características del paisaje pueden influir en el movimiento de individuos o incluso gametos, determinando la estructura genética de una población (Manel et al. 2003).

En el ambiente marino la ausencia de barreras físicas aparentes da cuenta de una débil o nula diferenciación genética (e.g. Canales-Aguirre et al. 2010). No obstante, aunque la ausencia de barreras sea aparente, se han reportado casos de poblaciones genéticamente diferenciadas, que en algunos casos esta explicada por características del

paisaje, tales como, giros oceanográficos, corrientes marinas, frentes oceanográficos, batimetría, salinidad y temperatura (Nielsen et al. 2004, Jørgensen et al. 2005, Shaw et al. 2004, Teske et al. 2008, Galarza et al. 2009, Hemmer-Hansen et al. 2007, Knutsen et al. 2009). Específicamente, en peces Clupeidos se ha encontrado correlación entre características de su hábitat y diferenciación poblacional en *Clupea harengus* (Jørgensen et al. 2005, Bekkevold et al. 2005, André et al. 2011) y *S. sprattus* (Limborg et al. 2009, Glover et al. 2011). En general, en ambientes marinos, las características comúnmente correlacionadas con una diferenciación poblacional y que actúan como barrera al flujo génico son la temperatura y salinidad (Limborg et al. 2009). Particularmente, dado que las características del paisaje (e.g. parámetros físicos y/o químicos) en donde habita *S. fuegensis* (i.e. zona de canales y fiordos) en el océano Pacífico sur oriental, son bastante heterogéneos (Sievers 2006; Sievers & Prado, 1994; Silva *et al*., 1997, Sievers & Silva, 2006; Silva et al. 1998, Heusser 1990; Delgado, 2004; Rodrigo, 2006). Estas podrían jugar un rol importante como barreras al flujo génico, conduciendo a una diversificación intra-específica de la especie, generando cambios en las frecuencias alélicas y produciendo diferencias poblacionales locales.

Los antecedentes tanto a nivel macro como microevolutivo permiten proponer que la diversificación es promovida por múltiples factores (Benton, 2009), los cuales pueden afectar diferencialmente según la escala a la cual actúen. De esta manera conocer cual o cuales de los factores abióticos, que comúnmente son citados en literatura (e.g. temperatura, salinidad, etc.) pueden estar promoviendo esta diversificación es de gran ayuda para explicar la biodiversidad de los Clupeiformes. Actualmente, a nivel macroevolutivo, para el Orden Clupeiforme la información se restringe solamente a las relaciones filogenéticas del grupo, dejando un vacío en cuanto al origen y diversificación de las especies, además de los factores que podrían estar asociados a la diversificación de estas. A niveles taxonómicos inferiores hay algunas propuestas basadas en estudios poblacionales; sin embargo, no hay evaluaciones explícitas de los factores que promoverían la diversificación de este grupo en general. Por lo tanto, conocer la dinámica

de origen y diversificación de *Sprattus*, así como la diversificación de linajes o la estructuración poblacional (i.e. cambio en frecuencias alélicas) de *S. fueguensis* daría nuevas luces sobre los factores que han determinado su biodiversidad actual. Así mismo determinar si los factores ambientales que actúan a nivel macroevolutivo son los mismos que actúan a nivel microevolutivo, seria de gran ayuda para comprender como actúa la diversificación a diferentes escalas taxonómicas, y si los cambios ambientales actuales podrían tener fuertes impactos, no solo sobre la población, sino también sobre la biodiversidad en general.

Chapter 1 – Seascape genetics in Patagonian fjords: a small pelagic fish's point of view

Cristian B. Canales-Aguirre[1,2], Sandra Ferrada-Fuentes[1,2], Ricardo Galleguillos[1], Cristián E. Hernández[2]

[1]Laboratorio de Genética y Acuicultura, Departamento de Oceanografía, Facultad de Ciencias Naturales y Oceanográficas, Universidad de Concepción, Concepción, Chile. Casilla 160-C.
[2]Laboratorio de Ecologia Evolutiva y Filoinformática, Departamento de Zoología, Facultad de Ciencias Naturales y Oceanográficas, Universidad de Concepción, Concepción, Chile. Casilla 160-C.

Manuscript in submitted to Estuarine, Coastal and Shelf Science

Section 1: Abstract

Marine environmental variables can play an important role in promoting population genetic differentiation in marine organisms. Although fjords ecosystems have attracted much attention due to the great oscillation of environmental variables that produce heterogeneous habitats, species inhabiting this kind of ecosystem have received less attention. In this study, we used *Sprattus fuegensis*— a small pelagic species that populate inland waters of the continental shelf, channels and fjords from Chilean Patagonia and Argentina— as model species to test whether environmental variables of fjords relate to population genetic structure. A total of 282 individual were analyzed from Chilean Patagonia with eight microsatellite loci. Bayesian and non-Bayesian analyses were conducted to describe its genetic variability and whether it shows spatial genetic structure. Bayesian and non-Bayesian analyses showed two well-differentiated genetic clusters along Chilean Patagonia distribution, but no isolation by distance pattern was found with a Mantel test analysis. Temperature and Nitrate were highly correlated to the expected heterozygosities, and explained the genetic variation of data in the redundancy analyses. None of the tested environmental variables was better than the null model in the performed Bayesian analyses. Overall, our results support that temperature and nitrate could explain the allelic frequency variation but not the genetic structure in *S. fuegensis*.

Section 2: Introduction

Marine environmental landscape parameters play an important role in promoting population genetic differentiation in marine organisms (Selkoe et al. 2008). Consequently, identifying environmental parameters that promote population genetic differentiation is a major focus of study in evolutionary biology (Hansen & Hemmer-Hansen 2007; Selkoe et al. 200). Most research on the effects of the environmental marine landscape on the genetics of population structure has been qualitative (e.g. Jørgensen et al. 2005; Fontaine et al. 2007; Galindo et al. 2010). However, this descriptive interpretation may not always be completely successful in identifying the factors that are responsible of the observed genetic structure of natural populations, and most importantly, they do not evaluate explicitly the environmental factors. In fact, few studies evaluate both: genetic and marine environmental data (Galindo et al. 2006). Manel *et al.* (2003) introduced the landscape genetic concept, which is able to explain spatial genetic patterns throughout landscape features (i.e. geographic, physic and chemical variables) and spatial statistics (Manel *et al.* 2003; Storfer *et al.* 2007). Although useful, most studies that used this approach have been performed in terrestrial organisms, leaving marine and freshwater organisms mostly unexplored (Storfer *et al.* 2010). Recently, concepts like seascape genetics or marine landscape genetics have started to appear in studies that evaluate how biotic and abiotic factors promote microevolutionary process in marine species (i.e. fishes, mollusk, crustaceous; Galindo *et al.* 2006; Selkoe *et al.* 2008; Riginos & Liggins 2013). Although different marine habitats could potentially affect the genetic diversity within species, (i.e. estuary, open sea, intertidal, pelagic, benthic), fjord habitats in particular have the potential to greatly affect population genetic diversity due to the complex scenario produced by its heterogeneous geography and environmental characteristics.

Fjords are high latitude estuaries and productive ecosystems that connect the open sea with freshwater from land drainage and ice melting (Pantoja *et al.* 2011; Landaeta *et al.* 2012). In addition, this ecosystem has been characterized mainly by strong fluctuations in salinity, temperature, pH, oxygen (Sievers 2006) and current patterns characteristics (Sievers & Silva 2006). These environmental characteristics could potentially affect the mechanisms that generate population differentiation (Jørgensen *et al.* 2005; Rijnsdorp *et al.* 2009; Pespeni *et al.* 2013; Coleman *et al.* 2013; Reusch 2014). For example, there is evidence of the effect of environmental oscillations on fjords' marine organisms at different organization levels: changes in composition of macrobenthic and zooplankton communities (Basedow *et al.* 2004; Willis *et al.* 2006; Beuchel *et al.* 2006), differences in mortality and growth (Kristoffersen & Salvanes 1998, 2009), abundance and search efficiency (Castellani *et al.* 2013). Environmental factors associated to fjords have been proposed as causes of trophic and reproductive adaptation (Goodson *et al.* 1995; Olsen *et al.* 2002; Kaartvedt *et al.* 2009), and transport and retention of larvae (Balbontín 2006; Bustos *et al.* 2007). Also, other studies have found population genetics differentiation between inner and outer fjords waters (Bradbury *et al.* 2009; Kristoffersen & Salvanes 2009; Fevolden *et al.* 2012).

The Chilean Patagonian fjords are one of the largest fjord regions in the world, extending from 41.5ºS (Reloncaví Fjord) to 55.9ºS (Cape Horn) and covering a total of 240,000 km^2 (Pantoja *et al.* 2011). The geographic landscape of this region includes channels, estuaries, fjords, peninsulas and islands (Pantoja *et al.* 2011). In addition, this ecosystem has been characterized mainly by strong fluctuations in salinity, temperature, pH, oxygen (Sievers 2006) and circulation patterns (Sievers & Silva 2006). *Sprattus fuegensis* or Patagonian sprat is a small pelagic marine fish of economic importance that inhabits from 41ºS, specifically in inland water and fjord in south of Chile to 40ºS in Argentina, including the Falkland Islands (Cousseau 1982; Whitehead 1985; Pequeño 1989; Aranis *et al.* 2007; Leal *et al.* 2011). This species reach a maximum age of 6 years (Cerna et al. 2014) and in a reproductive context is a partial spawner (Shirakova 1978a;

b; Sánchez *et al.* 1995; Leal *et al.* 2011) where female sprat maturates at an average length of 13.5 cm (Leal *et al.* 2011) and its eggs and larvae are pelagic (Bustos *et al.* 2008; Landaeta *et al.* 2011, 2012). These first developmental states are mainly abundant in inland waters from Chiloe, channels and fjord in Chile (Bustos et al., 2008; Landaeta et al., 2012, 2011) and in the Atlantic Ocean they have been reported nearby Santa Cruz, Argentina and southward to Falkland Island (Cousseau, 1982; Hansen, 1999).

Based on landscape marine environmental characteristic of Chilean fjord, we propose that the pelagic small fish *S. fuegensis* has population genetic differentiation. Given the geography of the area, we expected to find at least two genetic clusters: one group from inner water from Chiloé and fjord close to Aysén, and finally another group in the most distant locality from the Strait of Magellan. The goal of this study was to describe the genetic diversity and population structure of *Sprattus fuegensis* along the Chilean fjord and to evaluate the effect of marine environmental variables as a causal mechanism on population structure.

Section 3: Methods

Sample collection

A total of 282 individuals were collected from ten locations from the Chilean Patagonian fjords (Figure 1), including inland waters and fjord where *S. fuegensis* has been recorded. Locations were selected based on early studies from scientific cruises (Niklitschek *et al.* 2009; Landaeta *et al.* 2012) between 41º to 46ºS, except the most southern location (i.e. 53ºS), which was selected based on personal communication with artisanal fishermen from Punta Arenas. The collection of samples was done during *Sprattus fuegensis*' spawning season (September and December) because this season represents the most robust period to delineating population genetic structure (Glover *et al.* 2011). Muscular tissue was sampled from each individual and stored in 96% ethanol for further analyses.

Environmental database

The currently available global marine environmental databases (e.g. BioOracle, AquaMaps, MARSPEC) have large gaps on information from inland waters and fjords. We then compiled an environmental marine database based on published literature and oceanographic research cruises. We recorded maximum, minimum, average and range of seven marine environmental variables: temperature, salinity, pH, oxygen, phosphate, and nitrate (Supplementary Table 1.6; Silva *et al.* (1997, 1998), Guzmán & Silva (2002), Sievers *et al.* (2002), Silva & Calvete (2002), Valdenegro & Silva (2003), Silva & Guzmán (2006), Silva & Valdenegro (2008), and Carrasco & Silva (2010)) Oceanographic environmental collections obtained from literature are not identical to our sample locations; therefore, we used the nearest site to obtain marine environmental data.

Genetic database

Total genomic DNA was isolated using NucleoSpin tissue Kit (Machery-Nagel) and carried out according to the manufacturers' recommendations. The quality and quantity of DNA purification were measured in an Eppendorf biophotometer® and the template DNA was diluted to 20 ng/µL for the PCR amplifications. We used eight tetranucleotide microsatellites loci described to *S. fuegensis* by (Ferrada-Fuentes *et al.* 2014) (i.e. Spfu_6, Spfu_9, Spfu_29, Spfu_30, Spfu_42, Spfu_44, Spfu_45, and Spfu_48). These loci were amplified following the protocol described previously by Ferrada-Fuentes *et al.* (2014) and in PCR procedures we included both, positive and negative controls. The PCR products were analyzed on an ABI 3330 DNA sequencer. Alleles were scored using Peak Scanner™ software v1.0 (www.appliedbiosystems.com/peakscanner), with GS500 as the internal size standard.

Genetic variability

Because large individual samples are expected to have more alleles than small individual samples, and the number of individuals per locality was not homogeneous, we

conducted a rarefaction analyses to estimate how many individuals we need to detect all alleles present in a population (i.e. allelic richness, A_R) in HP-RARE (Kalinowski 2004, 2005). Outputs of allelic richness obtained from rarefaction analyses indicated that the average expected number of alleles in our standardized sample size (i.e. n=12) was less than our smallest sample size obtained in field (i.e. n=24 to ZH and ZL), therefore our number of individuals was well-suited to further analyses (Table 1.1). To evaluate the quality of the genetic database, we estimated the presence of genotyping errors, like drop-out alleles, stutter bands, and possible presence of null alleles. These analyses were conducted in the MICRO-CHECKER v2.2.3 software (van Oosterhout *et al.* 2004). Accordingly to MICRO-CHECKER several loci showed that the general excess of homozygotes is distributed across most allele size classes yielding possible Hardy & Weinberg deviations and presence of null alleles (van Oosterhout *et al.* 2004). Taking into account that null alleles presence could have an impact on the estimation of population differentiation (Chapuis & Estoup 2007), and in order to avoid decrease power in further analyses (Girard & Angers 2008), we employed model-based clustering and Bayesian assignment methods (Corander *et al.* 2003; Guillot *et al.* 2008; Carlsson 2008). This takes into account null alleles and significantly improves their accuracy like in GENELAND software (Guillot *et al.* 2008). In addition, simulations including datasets that include the presence of null alleles have demonstrated that genetic clustering outputs do not show more gene pools than there are in reality (Chapuis *et al.* 2008) and that they improve significantly their precision in determining genetic clusters (Guillot *et al.* 2008). Therefore we used the raw microsatellite data set without any correction for null alleles to infer the number of population clusters in *S. fuegensis*. The total number of alleles (N_A), expected (H_E) and observed (H_O) heterozygosity were estimated to determine the genetic variability of the samples; these parameters were calculated for each locus and locality using GENALEX v6.5 software (Peakall & Smouse 2012). To determine whether localities had significant deviations to Hardy-Weinberg equilibrium and linkage-disequilibrium we conducted analyses in ARLEQUIN v3.1 (Excoffier & Lischer 2010) and GENEPOP 3.1 (Raymond & Rousset 1995; Rousset 2008), respectively. Finally, a

pairwise F_{ST} comparison between sampling locations were obtained from ARLEQUIN where p-value was obtained after 10,100 permutations. Sequential Bonferroni correction (Rice 1989) for multiple comparisons was applied when necessary.

Number of genetic clusters

To infer genetic cluster number (K) in our sample set, we used three Bayesian approaches based on clustering method which differed in that they: a) incorporate or not a null allele model, and b) use non-spatial or spatial algorithm. First, we used STRUCTURE v.2.3.3 (Pritchard *et al.* 2000; Falush *et al.* 2003), which does not incorporate a null allele model, use a non-spatial model based on a clustering method and it is able to quantify individual genome proportion from each inferred population. Previous ran were carryout to define what ancestry models (i.e. no admixture model and admixture model) and allele frequency models (i.e. correlated and uncorrelated allele frequency model) fit our data set. All these previous running were conducted with locality information priors to improve the detection of structure when this could be weak (Hubisz *et al.* 2009). These previous simulations were run testing K = 1 – 10 (i.e. 1 indicate non genetic population difference and 10 indicate alike cluster number), with a 5.000 iterations of burn-in and a run length of 50.000, all these ran were replicated 5 times. Before choosing models to run our data set we evaluate ΔK Evanno's index (Evanno *et al.* 2005), to identify whether different models yielded different K values, implemented in STRUCTURE HARVESTER (Earl & VonHoldt 2012). Finally, to choose the best model to run our data we compare marginal likelihood of each model evaluated using the Bayes Factor (BF). The best model was no admixture with correlated frequency incorporating its sampling locations like an informative prior (Table 1.2). The final simulations were ran testing k = 2, with a 500,000 iterations of burn-in and a run length of 1,000,000, all these replicated ten times independently. Second, we used GENELAND v.0.3 (Guillot *et al.* 2005b), which incorporates geographic information (i.e. coordinates) in a spatial model in order to detect spatial discontinuities among populations with possible uncertainty in spatial coordinates (Guillot *et al.* 2005a) and a

null allele model that improves significantly their accuracy to inferences (Guillot *et al.* 2008). Similarly to the methods with STRUCTURE described above, we previously ran short analyses to determine what model (i.e. correlated or uncorrelated frequency model) fit our data set. All runs were performed using the "null allele model" setting given that it may have been present in our data. Previous simulations were run testing K = 1 – 10, using 1,000,000 MCMC iterations, 10,000 thinning and all these ran were replicated five times each one. We integrated the Evanno's method to GENELAND output, in order to estimate the true K in each models combination conducted in this software. As well as above, the selected of the best models was evaluated using BF. The best model was using correlated frequency model (Table 1.2). The final simulations were ran testing K = 2, using 10,000,000 MCMC iterations, 10,000 thinning and all these ran were replicated ten times each. To identify the number of genetic cluster present in our data we made a graphic with density probability, per each K, per iteration. Finally, we plotted a posterior probability map distributed in our sampling area.

Landscape analyses

To identify patterns of population genetic variation that derive from spatially limited gene flow (i.e. Isolation by distance, IBD), we conducted a Mantel test using a genetic matrix (i.e. F_{ST} and R_{ST}) and geographic distance (i.e. Log). Pearson correlation coefficients (i.e. r) was calculated in the R package VEGAN (Oksanen *et al.* 2013), and p-values were calculated on 10,000 permutations. To identify average genetic diversity parameters (i.e. N_A, H_O, H_E) that show correlation with average environmental variables (i.e. Temperature, Salinity, pH, Oxygen, Phosphate and Nitrate), we conducted a correlation analyses in the R package VEGAN.

Environmental factors that promote changes at the microevolutionary level (i.e. population genetic structure) were estimated using hierarchical Bayesian models. We conducted analyses in GESTE v2 (Foll & Gaggiotti 2006), in order to evaluate whether variables from our marine environmental data set explain patterns of population genetics structure (specific factors and data set used in GESTE were described above). Explicitly,

GESTE relates F_{ST} values with environmental factors using a generalized linear model (GLM). We ran ten pilot runs (burn-in period) to have priors of mean and variance of alpha's parameters distribution (alpha is the vector of regression coefficients that correspond to environmental data). After pilot running, we ran 10,000 MCMC iterations with 100 thinning and all these ran were replicated five times each one. In all combinations of marine environmental variables were considered and evaluated using estimates of posterior probability, the 95% highest probability density interval (HPDI). In order to identify whether environmental variables could explain variation in allele frequencies among locations we conducted a redundancy analysis (RDA) in the R package VEGAN. Specifically, we identify the relative contribution of each environmental variable on the allelic frequency variation using a forward stepwise selection (i.e. *ordistep* function) using a Akaike's information criterion in VEGAN. P-values were estimate based on 10,000 permutations. Finally, we plot only environmental variables that more explain the data variability by *ordistep* function.

Section 4: Results

Overall, high genetic variability at all microsatellite loci were found for *S. fuegensis* samples, where the N_A per locus ranged from 13 to 23, H_E ranged from 0.801 to 0.828, and H_O ranged between 0.699 and 0.847 (Table 1.1). Significant deviations from Hardy-Weinberg equilibrium were found at loci for some locations samples due to heterozygote excess as indicated for MICROCHECKER outcomes (Table 1.1). No pairwise comparison locus seems to be in linkage disequilibrium ($P > 0.05$). Pairwise F_{ST} and R_{ST} indices showed highly significant difference in comparison between Zone_D and the remainder locations (Table 1.3).

Bayesian approaches based on clustering method were congruent among them, despite STRUCTURE does not included a model that incorporated locus with possible null allele as GENELAND. The smallest values of K that capture the major structure of the data were 2 in all of case (Figure 1.2). One genetic cluster (i.e. the major one, called

largest cluster, LC hereafter) includes almost all sampling localities (i.e. ZA, ZB, ZE, ZH, ZI, ZJ, ZK, ZL, ZN). Another genetic cluster includes only ZD sampling locality (called smallest cluster SC hereafter). Outcomes plot from STRUCTURE revealed six individuals sampling from CL2 show a high genome proportion (>60%to multilocus genotype) from CL1 (Figure 1.2A). Maps of posterior probabilities of population membership obtained from GENELAND to the CL1 (Figure 1.2B), and CL2 (Figure 1.2C) showed the highest-probability lines (i.e > 0.8 posterior probability) indicating the potential spatial position of genetic discontinuities between CL1 and CL2. In addition, outcomes from GENELAND do not present any ghosts populations at not sampled areas.

Transformed pairwise genetic distances between locations ((F_{ST}/(1-F_{ST}) and R_{ST}/(1-R_{ST})) and log natural of geographic distance did not reveal any association of genetic distances with geography in the Mantel tests: low, non-significant negative correlations between distance matrices (i.e. F_{ST} (r= -0.1682, P= 0.7811) and R_{ST} (r= -0.1426, P= 0.6973) were inferred, indicating the absence of isolation by distance. Coefficient of determination (R^2) between genetic diversity indices showed values ranged 0.03 to 0.48. The relationship between H_O and Temperature average was the highest correlation with a Pearson correlation coefficient (r) of 0.69.

The marine environmental factors that showed the highest sum of posterior probability included in the analyses were nitrate average and, minimum; oxygen maximum; temperature maximum, and range; and finally phosphate average, minimum, and range (Table 1.4). All that factors seems to be important to described genetic variation, notwithstanding none environmental variable showed a high sum of posterior probability, the null model always explained more than 44% of the genetic structure among locations in each dataset (Table 1.5). Consequently, it means that any single environmental factor tested could explain the genetic structure observed. The RDA incorporating the significant environmental variables ordered by AIC (Supplementary Table 1.7) indicated that the minimum values of nitrate, and range values of temperature explain the allelic frequency variation in out data set (P < 0.039) (Figure 1.3B).

Section 5: Discussion

Marine environmental landscape parameters play an important role in promoting population genetic differentiation in marine organisms. Taking into account that landscape features could drive population differentiation, species that live in heterogeneous environments could be susceptible to diverge. Our results showed two genetic clusters in *Sprattus fuegensis* Patagonian fjords, genetic structuration, which is not related with environmental variable tested here, however temperature and nitrate are related with genetic diversity and it might explain the allelic frequency variation that we found

Genetic diversity along inland waters and fjords

The genetic variability of *S. fuegensis* is remarkably similar to the variability obtain with microsatellite in marine, anadromous and freshwater fishes by DeWoody & Avise (2000) . Compared to other marine organisms, the heterozigosity (H_E = 0.80-0.89) in this study was similar to that of the congeneric *S. sprattus* (H_E = 0.82-0.89 (Glover *et al.* (2011)), it was higher than the values of their relatives *Clupea pallasii* (H_E = 0.7-0.95, Semenova *et al.* (2014)) and *C. harengus* (H_E = 0.71-0.78, Jørgensen *et al.* (2005), and H_E 0.85-0.84, Mariani *et al.* (2005)).

Individuals from the zone with low variability (SC) showed less allele number and privative alleles respect to the zone with high variability (LC) (Table 1.1). Although this genetic pattern may be related to unequal sample size in each cluster, small populations can be highly affected by genetic drifts, changing allele frequencies trough the time and therefore fixing alleles in a population, which may decreasing the genetic variation like showed by SC. With the exception of private allele 306 at Spfu_29 locus from SC, all others private alleles showed low frequency. In addition, some alleles from SC despite were shared with remain locations they showed a 2 or 3 higher fold frequency than remain locations.

Genetic structure: biological characteristics and environmental features

Although none of the predicted clusters were found, two well-structured clusters were founded by STRUCTURE and GENELAND software, providing strong support to two genetic populations of *S. fuegensis* along its Chilean Patagonia distribution. Moreover F_{ST} and R_{ST} indices are consistent with Bayesian analyses. Despite of that, our result is in contrast with the genetic homogeneity founded in other marine species at the same geographic area (*Genypterus blacodes*; Canales-Aguirre *et al.* 2010)), our F_{ST} and R_{ST} outcomes are quite comparable to fixation indices at neutral loci obtained for relatives to Patagonian sprat, *S. sprattus* (Limborg *et al.* 2009; Glover *et al.* 2011), *Clupea harengus* (Shaw *et al.* 1999; Jørgensen *et al.* 2005, 2008; Teacher *et al.* 2013) and *Clupea pallasi* (Sugaya *et al.* 2008; Wildes *et al.* 2010; Semenova *et al.* 2014).

The largest cluster (LC) (Figure 1.2) may be explained by life history characteristic like dispersal stage or by marine environmental features. The latter because the localities that form the LC showed large environmental variability in its entire distributional range. On the other hand, the smallest cluster (SC) may be explained by local-scale larval retention. At present, specific spawning grounds of *S. fuegensis* have not been identified in its Chilean Patagonia distribution, nonetheless mature adults have been identified in inner waters from Chiloé (Aranis *et al.* 2007). Moreover, presence of juveniles has been detected in inner waters from Chiloé and fjord close to Aysén (i.e. between Puerto Aguirre and Estero Elefante) (Cárdenas 2009; Niklitschek *et al.* 2009). In numerous locations adjacent to the Strait of Magellan —Pacific and Atlantic ocean on Magellanic shelf— eggs presence have been found (Sánchez *et al.* 1995; Landaeta *et al.* 2011, 2012). Using otolith microchemistry from juvenile *S. fuegensis* individuals, Galleguillos *et al.* (2012) showed that three different nursery grounds (inner water sea from Chiloé, and fjord close to Aysén and in the Strait of Magellan can be found). In the area of the Strait of Magellan and channels adjacent to Atlantic Ocean, Sánchez *et al.* (1995) identify the largest nursery ground of *S. fuegensis* species along the Argentinian Patagonian coast with a juvenile production of 1.3×10^9 individuals. Incorporating this

valuable information to our results, we hypothesize that the lack of genetic differentiation found in LC should be explained by the abundance and distribution of larvae, eggs and juveniles from nursery grounds or by the close proximity of spawning grounds along the Chilean geographic range. Similarly, a probable explanation to the none genetic differences found in *Genypterus blacodes* along inland waters, channels and fjords in Chilean Patagonia was the close proximity of spawning grounds in the same study area (Canales-Aguirre *et al.* 2010). Adults migration in *S. fuegensis* has not been recorded to date, however, indirect evidence (i.e. microchemistry of otoliths and parasites tags) has pointed out that there must exist active dispersal of adults between inner seawater from Chiloé, and fjord close to Aysén (Galleguillos *et al.* 2012). Similarly, the same mechanism has been proposed in *Engraulis ringens* (Ferrada *et al.* 2002) and *Strangomera bentinki* (Galleguillos *et al.* 1997), two small pelagic marine fishes distributed along the continental shelf. Furthermore, taking into account the broad distribution of eggs, larvae and juvenile that has been recorded for this species (Sánchez *et al.* 1995; Cárdenas 2009; Niklitschek *et al.* 2009; Landaeta *et al.* 2011, 2012), we can suggest that passive dispersal migration might be playing a key role in the lack of genetic structure found within LC. Water bodies and circulation patterns can give support to the role played by dispersal passive migration (Silva *et al.* 1998). Sievers & Silva (2006) recorded the directionality of different bodies of water along the Patagonian Chilean sea. They described in the superficial level (i.e. 0-30 m) a narrow estuarine water layer with low salinity that lead into Boca del Guafo (Sievers & Silva 2006). At the middle level (i.e. 30-150 m) —depth where mainly *S. fuegensis* can be recorded— they described a broad subantartic body of water that goes into Boca del Guafo and then divide to northward in Inner sea water from Chiloé and to southward to fjord and waterway close to Aysén (Sievers & Silva 2006). These suggest that in the LC the circulation pattern of bodies of water may be the underlying mechanism that drives the null genetic differences among localities.

The SC was an unexpected outcome considering that, based on previous environmental information; we expected to find genetic differences between the more isolated areas. The SC showed highly significant differences, giving a strong support for its existence. Based in GESTE analyses, none of the tested environmental variables, physical (i.e. temperature) or chemical (salinity, pH, oxygen, phosphate and nitrate) of the data sets incorporated in this study were better than the null model (Table 1.5). The RDA showed similar results; however, minimum nitrate and range temperature were variables that explain the allelic frequency variation in the two clusters found in this study (Figure 3). The SC is localized within Chiloense ecoregion (Spalding *et al.* 2007), ecoregion that has been described has having an upwelling system where mesoscaled process like eddies, fronts and plumes increase the retention of phytoplankton (Alvarez *et al.* 2010). Therefore, it has been suggested that the Chilean fjord region is highly productive area (Iriarte *et al.* 2007, 2010), indirectly relating to nitrate measures. Unfortunately, we cannot incorporate primary productivity as a variable in our analyses because there is no information available for all the locations used in this study. The high marine fishes biodiversity that have been found in this area may be related to this primary productivity (Balbontín & Bernal 1997; Bernal & Balbontín 1999). Therefore, high phytoplankton and zooplankton aggregations and kelp forests provide feeding and refuge, respectively, to diverse fishes and invertebrate communities (Alvarez *et al.* 2010).

In conclusion, our results suggest two genetic clusters for *S. fuegensis* in its Chilean distribution, where the nitrate and temperature might explain the allelic frequency observed but not its genetics structure. These environmental features together with local oceanographic processes could be postulated as causal mechanisms of genetic geographic differentiation. Further analyses should be carrying out to confirm this spatial genetic pattern, whether this pattern is stable in long terms and also whether environmental features not tested now are able to explain better population genetic structure of this species.

Section 6: Tables

Table 1.1: Genetic diversity parameters per sampling location and genetic cluster in microsatellite loci of *Sprattus fuegensis.*

N: sample size, Lat: Latitude, Lon: Longitude, N_A: allele number, H_O: observed heterozygosity, H_E: expected heterozygosity, F: fixation index, A_R: allelic richness. F_{ST}(hap): p-value per locus to F_{ST} index estimation from haplotype frequency, F_{ST} and R_{ST}: p-values per locus to F_{ST} index estimation from distance matrix.

	N	Lat	Lon	Parameter	Spfu_6	Spfu_9	Spfu_29	Spfu_30	Spfu_42	Spfu_44	Spfu_45	Spfu_48	Mean
Zone_A	32	-41.793	-73.286	N_A	10	17	12	14	14	21	15	12	14.4
				H_O	0.679	0.679	0.429	0.656	0.806	0.636	0.563	0.656	0.638
				H_E	0.864	0.915	0.852	0.887	0.905	0.937	0.899	0.878	0.892
				F	0.214	0.258	0.497	0.26	0.109	0.321	0.374	0.253	
				A_R	9	15	10	12	13	19	13	11	
Zone_B	28	-42.063	-72.860	N_A	12	16	12	16	15	23	14	9	14.6
				H_O	0.636	0.679	0.148	0.852	0.75	0.857	0.75	0.571	0.655
				H_E	0.879	0.821	0.848	0.909	0.899	0.943	0.889	0.851	0.880
				F	0.276	0.174	0.825	0.063	0.166	0.091	0.156	0.329	
				A_R	11	12	11	14	13	19	12	8	
Zone_D	28	-43.860	-72.968	N_A	10	15	12	12	12	19	9	13	12.8
				H_O	0.593	0.815	0.63	0.56	0.667	0.852	0.704	0.741	0.695
				H_E	0.636	0.767	0.826	0.834	0.861	0.855	0.785	0.862	0.803
				F	0.068	-0.062	0.238	0.329	0.225	0.004	0.104	0.141	
				A_R	8	11	10	11	11	15	8	11	
Zone_E	28	-43.095	-73.675	N_A	14	16	10	21	14	25	16	11	15.9
				H_O	0.75	0.667	0.481	0.692	0.786	0.654	0.643	0.643	0.665
				H_E	0.87	0.91	0.813	0.936	0.904	0.942	0.89	0.888	0.894
				F	0.138	0.267	0.408	0.26	0.131	0.306	0.277	0.276	
				A_R	12	14	9	18	13	21	13	11	
Zone_H	24	-44.091	-73.796	N_A	12	16	8	13	15	20	12	10	13.3
				H_O	0.87	0.739	0.5	0.792	0.765	0.5	0.667	0.625	0.682
				H_E	0.887	0.879	0.665	0.903	0.917	0.936	0.868	0.814	0.859
				F	0.019	0.159	0.248	0.123	0.166	0.466	0.232	0.232	
				A_R	11	14	7	12	15	18	11	9	
Zone_I	30	-44.688	-72.980	N_A	10	15	11	16	15	26	14	11	14.8
				H_O	0.759	0.667	0.3	0.867	0.8	0.828	0.552	0.533	0.663
				H_E	0.862	0.899	0.861	0.922	0.913	0.955	0.897	0.865	0.897
				F	0.12	0.259	0.651	0.06	0.124	0.133	0.385	0.383	
				A_R	9	13	10	14	13	22	12	10	
Zone_J	30	-45.306	-73.801	N_A	11	23	10	13	16	24	14	11	15.3
				H_O	0.767	0.621	0.4	0.767	0.815	0.69	0.679	0.69	0.679
				H_E	0.853	0.868	0.836	0.891	0.906	0.937	0.883	0.845	0.877
				F	0.101	0.285	0.521	0.14	0.101	0.264	0.231	0.184	
				A_R	10	17	9	11	14	19	12	10	
Zone_K	28	-45.767	-73.598	N_A	10	17	7	15	17	26	12	11	14.4
				H_O	0.565	0.778	0.259	0.821	0.741	0.577	0.731	0.654	0.641
				H_E	0.852	0.901	0.798	0.901	0.892	0.942	0.886	0.845	0.877
				F	0.336	0.137	0.675	0.088	0.17	0.387	0.175	0.227	
				A_R	9	14	7	13	14	21	11	10	
Zone_L	24	-45.181	-73.847	N_A	15	11	7	17	14	25	12	8	13.6
				H_O	0.75	0.435	0.25	0.591	0.957	0.75	0.87	0.75	0.669
				H_E	0.883	0.868	0.802	0.907	0.905	0.95	0.886	0.797	0.875
				F	0.15	0.499	0.688	0.349	-0.057	0.21	0.018	0.059	
				A_R	13	11	7	15	13	21	11	7	

Table 1.1: Genetic diversity parameters per sampling location and genetic cluster in microsatellite loci of *Sprattus fuegensis*.

N: sample size, Lat: Latitude, Lon: Longitude, N_A: allele number, H_O: observed heterozygosity, H_E: expected heterozygosity, F: fixation index, AR: allelic richness. F_{ST}(hap): p-value per locus to F_{ST} index estimation from haplotype frequency, F_{ST} and R_{ST}: p-values per locus to F_{ST} index estimation from distance matrix.

Zone_J	30	-45.306	-73.801	N_A	11	23	10	13	16	24	14	11	15.3
				H_O	0.767	0.621	0.4	0.767	0.815	0.69	0.679	0.69	0.679
				H_E	0.853	0.868	0.836	0.891	0.906	0.937	0.883	0.845	0.877
				F	0.101	0.285	0.521	0.14	0.101	0.264	0.231	0.184	
				A_R	10	17	9	11	14	19	12	10	
Zone_K	28	-45.767	-73.598	N_A	10	17	7	15	17	26	12	11	14.4
				H_O	0.565	0.778	0.259	0.821	0.741	0.577	0.731	0.654	0.641
				H_E	0.852	0.901	0.798	0.901	0.892	0.942	0.886	0.845	0.877
				F	0.336	0.137	0.675	0.088	0.17	0.387	0.175	0.227	
				A_R	9	14	7	13	14	21	11	10	
Zone_L	24	-45.181	-73.847	N_A	15	11	7	17	14	25	12	8	13.6
				H_O	0.75	0.435	0.25	0.591	0.957	0.75	0.87	0.75	0.669
				H_E	0.883	0.868	0.802	0.907	0.905	0.95	0.886	0.797	0.875
				F	0.15	0.499	0.688	0.349	-0.057	0.21	0.018	0.059	
				A_R	13	11	7	15	13	21	11	7	
Zone_N	30	-53.613	-70.923	N_A	13	15	13	16	18	21	15	12	15.4
				H_O	0.759	0.5	0.185	0.759	0.893	0.435	0.69	0.6	0.603
				H_E	0.885	0.893	0.877	0.9	0.927	0.933	0.908	0.872	0.899
				F	0.142	0.44	0.789	0.157	0.036	0.534	0.241	0.312	
				A_R	12	13	11	13	16	19	13	10	
Cluster 1	254	-	-	N_A	17	38	17	24	22	46	20	18	25.3
				H_O	0.729	0.642	0.326	0.758	0.812	0.667	0.676	0.633	0.655
				H_E	0.889	0.910	0.860	0.919	0.929	0.963	0.905	0.870	0.906
				F	0.180	0.295	0.621	0.175	0.127	0.308	0.253	0.272	
				A_R	12	13	11	13	16	19	13	10	
Cluster 2	28	-	-	N_A	10	15	12	12	12	19	9	13	12.8
				H_O	0.593	0.815	0.630	0.560	0.667	0.852	0.704	0.741	0.695
				H_E	0.636	0.767	0.826	0.834	0.861	0.855	0.785	0.862	0.803
				F	0.068	-0.062	0.238	0.329	0.225	0.004	0.104	0.141	
				A_R	12	13	11	13	16	19	13	10	
			F_{ST}(hap)	p-value	0.000	0.000	0.000	0.001	0.000	0.000	0.002	0.002	
			F_{ST}	p-value	0.000	0.000	0.000	0.048	0.003	0.004	0.092	0.069	
			R_{ST}	p-value	0.000	0.099	0.000	0.000	0.003	0.004	0.000	0.810	

Table 1.2: Bayes factor comparison among different models incorporated in STRUCTURE and GENELAND software.

AdUn: admixture and uncorrelated model, AdCo: admixture and correlated model, NAdUn: no admixture and uncorrelated model, NAdCo: no admixture and correlated model.

Program	Ancestry model	Allele Frequency model	Best K	ln P(model | data)	S.E.	AdUn	AdCo	NAdUn	NAdCo
STRUCTURE	Admixture	Uncorrelated	2	-11266.032	+/- 0.341	-	0.00E+00	7.07E+04	0.00E+00
STRUCTURE	Admixture	Correlated	2	-11238.516	+/- 0.272	8.91E+14	-	6.30E+13	0.00E+00
STRUCTURE	No Admixture	Uncorrelated	2	-11270.29	+/- 0.309	0.014	0.00E+00	-	0.00E+00
STRUCTURE	No Admixture	Correlated	2	-11222.583	+/- 0.239	7.41E+18	8.31E+09	5.23E+20	-
GENELAND	No Admixture	Uncorrelated	2	-10532.236	+/- ∞			0	-
GENELAND	No Admixture	Correlated	2	-10493.699	+/- ∞			-	5.45081E+16

Table 1.3: Pairwise F_{ST} and R_{ST} indices between sampling locations.

Bold values represent P-values less than 0.05. ** = Significant comparison

		Zone_A	Zone_B	Zone_D	Zone_E	Zone_H	Zone_I	Zone_J	Zone_K	Zone_L	Zone_N
F_{ST}	Zone_A	-		**							
	Zone_B	0.000	-	**							
	Zone_D	**0.048**	**0.056**	-	**	**	**	**	**	**	**
	Zone_E	0.001	0.006	**0.051**	-						
	Zone_H	0.004	0.002	**0.057**	0.007	-					
	Zone_I	0.000	0.006	**0.054**	0.005	0.005	-				
	Zone_J	0.003	0.003	**0.057**	0.010	0.009	0.003	-			
	Zone_K	0.006	0.009	**0.052**	0.004	0.008	0.003	0.010	-		
	Zone_L	0.000	0.000	**0.064**	0.005	0.005	0.004	0.012	0.003	-	
	Zone_N	0.001	-0.002	**0.057**	0.004	0.006	0.001	0.004	0.009	0.003	-
R_{ST}	Zone_A	-		**							
	Zone_B	-0.012	-	**				**			
	Zone_D	**0.086**	**0.063**	-	**	**	**	**	**	**	**
	Zone_E	0.009	0.061	**0.119**	-						
	Zone_H	-0.024	0.038	**0.116**	0.009	-					
	Zone_I	-0.029	0.031	**0.087**	-0.002	0.007	-				
	Zone_J	0.027	**0.100**	**0.128**	-0.013	0.013	0.018	-			
	Zone_K	-0.033	0.028	**0.084**	0.021	0.010	0.005	0.050	-		
	Zone_L	0.039	0.079	**0.144**	0.002	0.027	0.007	0.019	0.040	-	
	Zone_N	0.013	0.028	**0.120**	0.011	-0.047	0.002	0.023	0.007	0.021	-

Table 1.4: Sum of posterior probabilities of models that include a given factor.

GESTE analyses included all 6 factors. Bold value indicates factor with highest score

	Sum posterior probabilities			
Factor	Average	Maximum	Minimum	Range
Temperature	0.075	**0.103**	0.066	**0.185**
Salinity	0.088	0.078	0.106	0.134
Oxygen	0.089	**0.115**	0.076	0.147
pH	0.119	0.091	0.090	0.143
Phosphate	**0.131**	0.085	**0.181**	**0.200**
Nitrate	**0.126**	0.088	**0.130**	0.123

Table 1.5: Posterior probabilities of the three most probable models for the analyses including the all factors tested.

Models are listed in decreasing order of posterior probabilities. Pr: Probability

Data set	Pr	Factor included
Average	0.54	Null
	0.07	Null, pH
	0.06	Null, Phosphate
Maximum	0.58	Null
	0.06	Null, Oxygen
	0.06	Null, Temperature
Minimum	0.53	Null
	0.08	Null, Phosphate
	0.06	Null, Salinity
Range	0.44	Null
	0.07	Null, Temperature
	0.04	Null, Salinity

Section 7: Figures

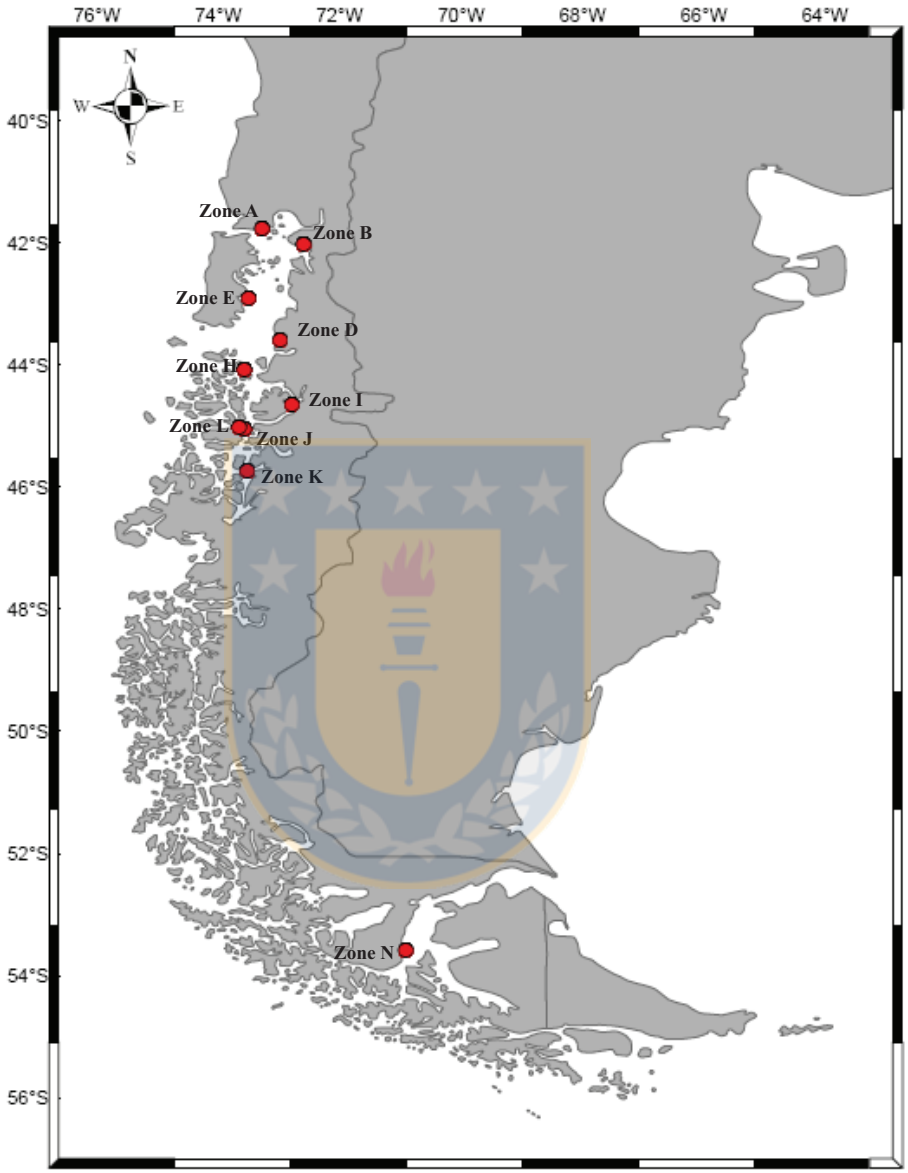


Figure 1.1: Map showing sample locations of *Sprattus fuegensis*.

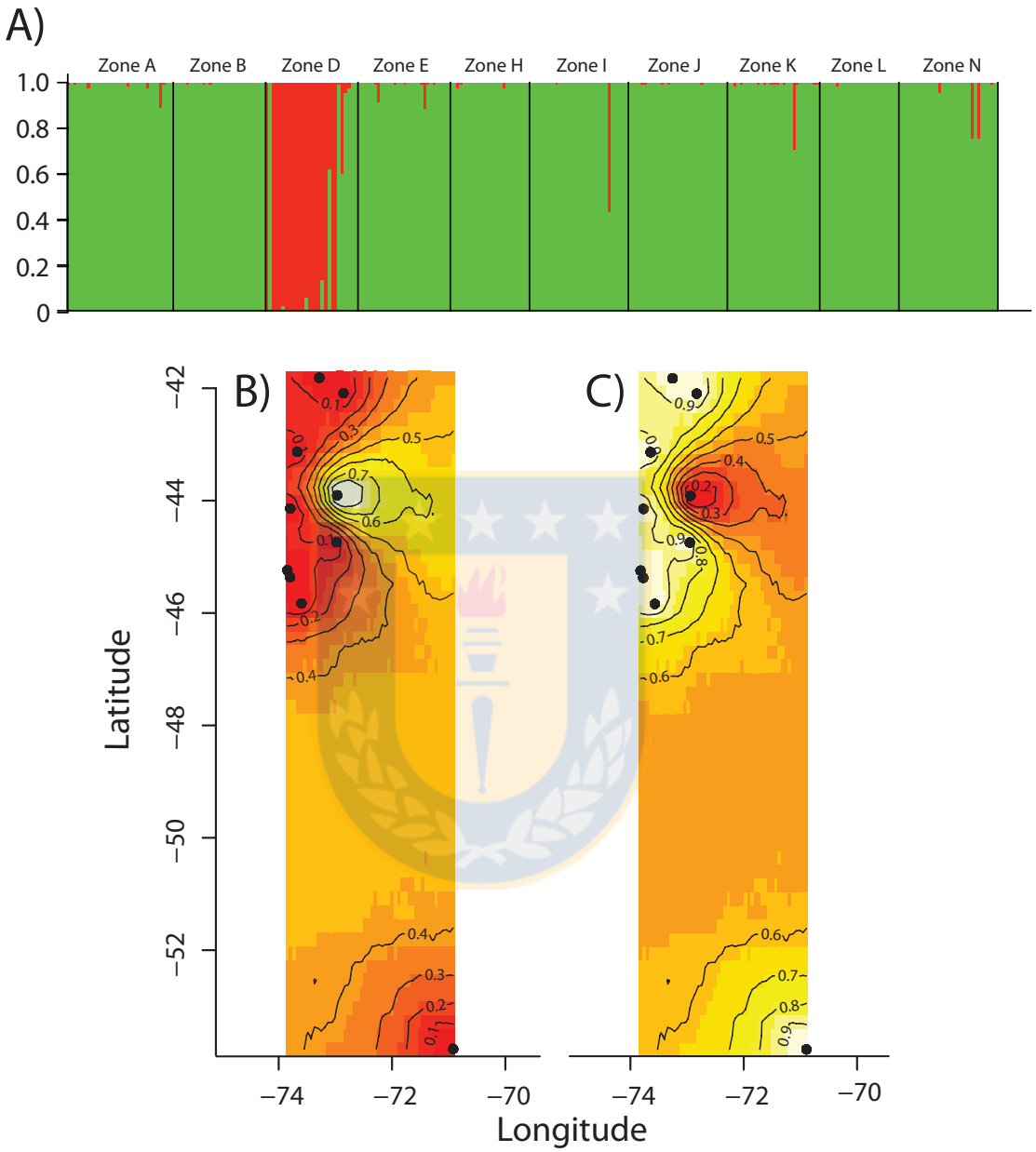


Figure 1.2: Bayesian clustering results from STRUCTURE and GENELAND.

A) Plot is shown the most likely number of clusters for dataset. GENELAND analyses with posterior probability isoclines denoting the extent of genetic landscapes. Clusters indicated by GENELAND: B) Largest cluster (LC) and C) Smallest Cluster (SC). Black dots represent localities analyzed in this study and regions with the greatest probability of inclusion are indicated by white, whereas diminishing probabilities of inclusion are proportional to the degree of colouring.

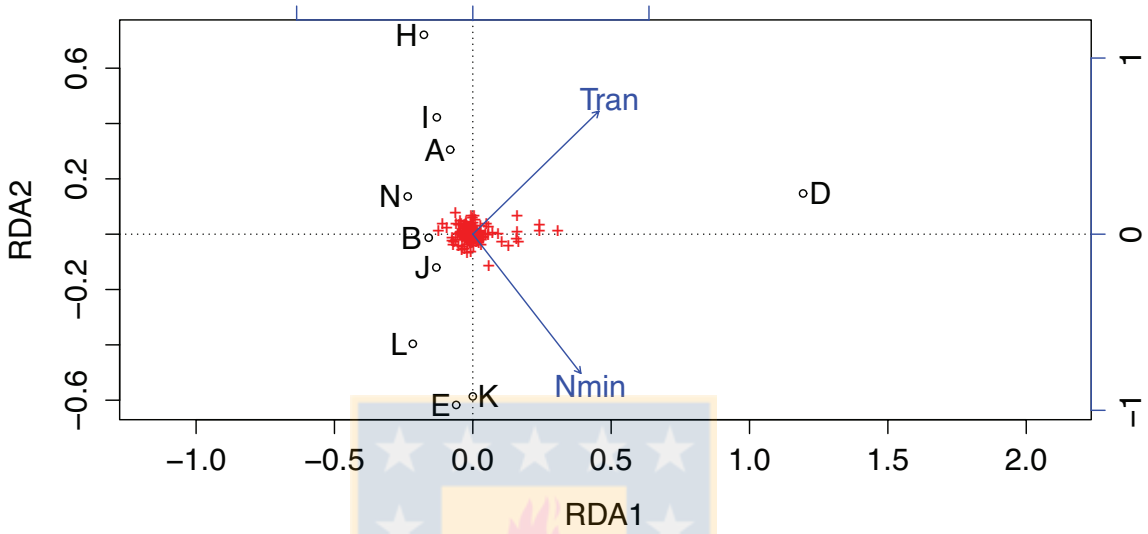


Figure 1.3: Redundancy analyses.

Based on only factors that show less Akaike value form *ordistep* analyses. Significant values less than 0.05 thresholds. Tran: Temperature range; Nmin: Nitrate minimum

Section 8: Supplementary tables

Table 1.6: Environmental variables used in our analyses.

Ave: average, Rang: Range, Max: Maximum, Min: Minimum

Variable	Parameter	Zone_A	Zone_B	Zone_D	Zone_E	Zone_H	Zone_I	Zone_J	Zone_K	Zone_L	Zone_N
Temperature (ºC)	Ave	10.675	10.550	10.587	9.875	10.063	9.813	9.688	9.438	9.750	7.063
	Rang	1.5	1.0	2.8	0.2	1.5	2.0	0.5	0.5	0.5	0.5
	Max	11.5	11.0	11.9	10.0	11.0	11.0	10.0	9.5	10.0	7.5
	Min	10.0	10.0	9.1	9.8	9.5	9.0	9.5	9.0	9.5	7.0
Salinity (psu)	Ave	31.563	31.563	32.696	32.850	31.786	29.063	30.250	29.000	30.250	30.563
	Rang	2.5	2.5	2.3	0.2	3.0	13.5	5.0	2.0	5.0	0.5
	Max	32.5	32.5	33.6	33.0	33.0	33.5	33.0	30.0	33.0	31.0
	Min	30.0	30.0	31.3	32.8	30.0	20.0	28.0	28.0	28.0	30.5
Oxygen (ml/L)	Ave	6.750	6.125	5.424	5.950	6.625	5.500	5.500	5.875	5.500	6.500
	Rang	3.0	2.0	3.3	0.2	3.0	4.0	2.0	0.5	2.0	0.0
	Max	8.0	7.0	7.1	6.0	8.0	7.0	6.0	6.0	6.0	6.5
	Min	5.0	5.0	3.8	5.8	5.0	3.0	4.0	5.5	4.0	6.5
pH	Ave	7.863	7.763	7.83	7.770	7.863	7.788	7.688	7.725	7.700	7.750
	Rang	0.3	0.2	0.2	0.0	0.4	0.5	0.2	0.1	0.2	0.1
	Max	8.0	7.9	7.9	7.8	8.1	8.0	7.8	7.8	7.8	7.8
	Min	7.7	7.7	7.7	7.8	7.7	7.5	7.6	7.7	7.6	7.7
Phosphate (µM)	Ave	1.300	1.400	2.15	1.550	0.950	1.200	1.500	1.600	1.500	0.650
	Rang	1.0	1.0	0.4	0.4	1.2	2.4	0.8	0.0	0.8	0.4
	Max	1.8	1.8	2.4	1.6	1.6	2.4	2.0	1.6	2.0	1.0
	Min	0.8	0.8	2.0	1.2	0.4	0.0	1.2	1.6	1.2	0.6
Nitrate (µM)	Ave	11.000	13.500	22.5	16.000	9.000	10.000	15.000	16.000	15.000	4.500
	Rang	16.0	12.0	4.0	0.0	20.0	24.0	8.0	0.0	8.0	4.0
	Max	20.0	20.0	24	16.0	20.0	24.0	20.0	16.0	20.0	8.0
	Min	4.0	8.0	20	16.0	0.0	0.0	12.0	16.0	12.0	4.0

Table 1.7: Relative contribution of each environmental variable tested using Akaike's information criterion.

Phos: Phosphate, Lat: Latitude, Long: Longitude, Nit: Nitrate, Oxy: Oxygen, Sal: Salinity, Tem: Temperature. Ave: average, Rang: Range, Max: Maximum, Min: Minimum. Bold values show significant p-values. Variable kept means environmental variables that explain variation in allele frequencies among locations

	All environmental factors			Excluding Nmin			Excluding Nmin and Tran		
	AIC	F	Pr(>F)	AIC	F	Pr(>F)	AIC	F	Pr(>F)
Nave	-12.62	20.82	**0.04**	-11.51	10.27	0.35	-12.94	10.12	0.57
Nmax	-11.40	0.93	0.37	-11.40	0.94	0.48	-12.96	10.30	0.53
Nmin	-12.14	16.16	**0.03**			Variable kept			
Nran	-11.28	0.82	0.67	-11.40	0.94	0.46	-12.96	10.30	0.52
Oave	-11.60	11.04	0.33	-11.19	0.78	0.75	-13.02	10.73	0.44
Omax	-11.24	0.79	0.75	-12.56	19.16	**0.02**	-12.82	0.93	0.62
Omin	-11.27	0.81	0.66	-11.33	0.88	0.61	-13.04	10.88	0.41
Oran	-11.24	0.79	0.72	-12.00	14.30	0.13	-12.83	0.94	0.61
Pave	-12.68	21.50	**0.03**	-11.49	10.09	0.37	-12.86	0.96	0.64
pHave	-11.52	10.38	0.42	-12.57	19.21	**0.02**	-12.99	10.51	0.45
pHmax	-11.14	0.70	0.90	-12.30	16.83	0.07	-12.95	10.19	0.51
pHmin	-11.24	0.79	0.61	-11.22	0.80	0.73	-13.08	11.09	0.38
pHran	-11.08	0.65	0.90	-11.82	12.76	0.18	-13.24	12.31	0.23
Pmax	-11.67	11.72	0.27	-11.43	0.96	0.45	-13.02	10.73	0.42
Pmin	-12.37	18.41	**0.05**	-11.46	0.98	0.43	-12.59	0.77	0.77
Pran	-11.20	0.75	0.67	-11.10	0.70	0.85	-12.95	10.18	0.51
Save	-11.87	13.55	0.18	-11.72	11.98	0.26	-13.13	11.46	0.34
Smax	-11.26	0.80	0.60	-11.38	0.92	0.52	-12.65	0.81	0.75
Smin	-11.34	0.87	0.35	-11.30	0.85	0.62	-13.27	12.48	0.22
Sran	-11.14	0.70	0.75	-11.20	0.78	0.74	-13.18	11.83	0.29
Tave	-11.20	0.75	0.60	-11.02	0.64	0.93	-12.63	0.80	0.79
Tmax	-11.52	10.38	0.25	-11.56	10.69	0.33	-12.61	0.78	0.81
Tmin	-10.91	0.50	1.00	-10.92	0.56	0.97	-12.61	0.78	0.81
Tran	-12.56	20.26	0.08	-13.38	26.74	**0.00**	Variable kept		

Chapter 2 – Consequences of past environmental variables on historic demography and intraspecific diversification of Patagonian sprat (*Sprattus fuegensis*)

Section 1: Abstract

Oceanographic environmental variation plays a major role in intraspecific demographic processes and has been defined as a driver of evolutionary change. Specifically, it has been hypothesized that the actual effective population size and genetic diversity of a species has been historically driven by ancient environmental changes. However, most genetic studies that investigate associations between genetic parameters and ancient environmental oscillations have used a qualitative approach. In these studies, little attention has been put on incorporating explicit paleocenographic data, or a proxy through phylogenetic signals of present-day environmental factors, to evaluate the relationship of environmental factors with microevolutionary diversification patterns. In this study we evaluate the role of historic environmental changes on the effective population size and genetic diversity of a model small pelagic marine fish, *Sprattus fuegensis*. This species inhabits the continental shelf of Chilean Patagonian fjord and Argentina, distribution which was highly impacted during the Pleistocene glaciation. To identify population structure we sequenced the Control Region mtDNA of a total of 335 individuals and we calculated the genetic diversity and differentiation indexes. Historical demography was evaluated using Bayesian Skyride Plot (BSRP) reconstruction, and then historical demography data was correlated with a paleoceanographic data set (i.e. temperature, salinity and sea-level). Also present-day oceanographic factors that showed phylogenetic signal were used to estimate its impact on patterns of intraspecific diversification rate. The results showed no population structure in *S. fuegensis* and BSRP supported two increments of effective population size through time. Results from BSRP

were strongly correlated with environmental oscillations of ancient temperature, salinity and sea level. Three present day oceanographic factors that have phylogenetic signal (i.e. salinity, phosphate and nitrate), were related to the microevolutionary diversification rate of the haplotypes; haplotype losses rate was constant and haplotype accumulation rate decreased trough time. Our outcomes suggest that ancient oceanographic environmental factors have played an important role in shaping the current population dynamic of *S. fuegensis*. We conclude that both, paleoceanographic proxies as well as historical signals of current oceanographic factors are key to disentangle historical population dynamics of this species.

Section 2: Introduction

Understanding the effects of environmental variation on microevolutionary patterns is a challenging issue in evolutionary biology (Losos et al. 2008). Evidence suggests that environmental variation plays a key role in intraspecific processes such as population genetic divergence, changes in effective population size, and population genetic variability (Kozak et al., 2008), and can act as a driver of evolutionary change (Erwin, 2009). However, most phylogeographic studies have not explicitly incorporated environmental data on its analyses, and its outcomes are based on posterior qualitatively interpretation of historical environmental oscillations (i.e. Pleistocene glaciations; Atarhouch *et al.* 2006; Cárdenas *et al.* 2009; González-Wevar *et al.* 2011; Liu *et al.* 2011; Ceballos *et al.* 2011). Furthermore, little effort has been made on correlating paleoenvironmental variables proxies and phylogeographic outcomes (e.g. Porobić *et al.* 2013), or on using present-day environmental factors to explicitly infer its relationship with microevolutionary diversification patterns (Kozak et al., 2008).

Currently, the available information about past climate changes offers the possibility to test how past environmental oscillations affected the species that occur in glaciated localities (e.g. South America, Vimeux *et al.* 2009). During the Pleistocene,

Southern Patagonia was heavily impacted by the glaciations (e.g. McCulloch *et al.* 2000; Hulton *et al.* 2002; Sugden *et al.* 2005) being the last one, the Last Glacial Maximum (LGM) which occurred between 19,000-26,500 years ago (Clark et al., 2009). This region was covered with ice between 36ºS and 55ºS (McCulloch et al., 2000), had an extension of about 1800 km (Hollin and Schilling, 1981), and reached the border of the continental shelf of the Pacific Ocean (DaSilva et al., 1997; Hulton et al., 2002). As a result of these environmental changes, the Sea Surface Temperature (SST) was cooler than is today (Lamy and Kaiser, 2009; Lamy et al., 2004) and the sea level decreased more than a hundred meters at high latitudes (Lambeck et al., 2002). Consequently, the last glaciation generated changes in habitat availability for marine species that inhabited the intertidal and near shore pelagic environments (e.g. González-Wevar *et al.* 2011; Ceballos *et al.* 2011), probably generating demographic changes (e.g. González-Wevar *et al.* 2011; Ceballos *et al.* 2011) and leaving others genetic footprints in marine populations (Avise, 2000).

Most studies that have evaluated the historical effects of the environment on the genetic diversity of a species in the Southern Patagonia have been performed on sessile organisms or in species with limited dispersal potential at the adult stage. Among these algae (Fraser et al., 2010, 2009; Macaya and Zuccarello, 2010; Montecinos et al., 2012) and invertebrates (Brante et al., 2012; Cárdenas et al., 2009a; Guzmán et al., 2011; Haye et al., 2014; Porobić et al., 2013; Sánchez et al., 2011a; Weis and Melzer, 2012) have received the most attention. Mobile species are less subject to local conditions than sessile species, therefore it is not expected that the effect of past environmental variables will act in a similar way in sessile or mobile organisms. However, studies in mobile organisms such as fishes are still lacking in this area (Canales-Aguirre et al., 2010; Ceballos et al., 2011). Marine fishes that inhabit shallow waters close to the coast, are good models to start exploring how environmental oscillation could affect the demography of a mobile species and if there is a recognizable genetic footprints.

The Patagonian sprat species, *Sprattus fuegensis*, is a small pelagic marine fish, of which there is scarce information on its population dynamics. This species inhabits the continental shelf, near to shore and their bathymetric distribution is shallow and narrow (0 – 50 m) (Cousseau, 1982; Whitehead, 1985). Its geographical distribution extends from 41ºS, specifically in inland water and fjord in south of Chile to 40ºS in Argentina, including the Falkland Islands (Aranis et al., 2007; Cousseau, 1982; Leal et al., 2011; Whitehead, 1985). *Sprattus fuegensis* has economic importance in Chile where 20,179 tones where harvested in 2010 (Sernapesca, 2012), while in Argentina is not considered an important economic resource and its catches are negligible (Hansen, 1999). Sprat is a partial spawner with a spawning season around spring and summer in their whole distribution (Leal et al., 2011; Shirakova, 1978a, 1978b). Female sprat mature at an average length of 13.5 cm (Leal et al., 2011) producing both pelagic eggs and larvae (Bustos et al., 2008; Landaeta et al., 2012, 2011). These first developmental states are mainly abundant in inland waters from Chiloe, channels and fjord in Chile (Bustos et al., 2008; Landaeta et al., 2012, 2011) and in the Atlantic Ocean they have been reported nearby Santa Cruz, Argentina and southward to Falkland Island (Cousseau, 1982; Hansen, 1999). Spawning grounds in the south of Chile have not been identified yet, but studies based on the presence of juveniles indicated that spawning could be occurring throughout all inland waters and fjords (Niklitschek et al., 2009). In contrasts, in their Atlantic distribution, spawning grounds have been described on the Patagonian shelf and near to the Falkland Islands (Sánchez et al., 1995; Shirakova, 1978a, 1978b).

Taking into account that currently *S. fuegensis* inhabits coastal and shallow marine waters in the Southeast Pacific and Southwest Atlantic Ocean, areas that were heavily impacted by both, glacial retreat and expansion that occurred in LGM, we hypothesize that the effective population size and genetic diversity of this species has been historically driven by ancient environmental changes. Therefore we predict that: (1) there is genetic differentiation between individuals from Southeast Pacific Ocean (i.e. impacted area by LGM) and Southwest Atlantic Ocean (non-impacted area by LGM). (2)

paleoenvironmental variables (i.e. paleotemperature, paleosalinity and historical sea level) show a correlation with historical effective population size (N_E) through the time due to the ancient environmental oscillations that occured during the Pleistocene's glaciations; (3) present-day environmental factors showed phylogenetic signals at the intraspecific level and these environmental variables are related to shifts in intraspecific evolutionary diversification rates.

The aim of this study was to determine whether *S. fuegensis* showed phylogeographic patterns along its distribution and whether ancient and present-day marine environmental variables explain the intraspecific evolutionary diversification rate of this species. To do this, we used mitochondrial DNA sequences (mtDNA), proxies from paleodatabases and comparative phylogenetic methods.

Section 3: Methods

Sampling locations

A total of 335 individual of *Sprattus fuegensis* were obtained from 12 locations distributed along the inland waters in Chilean Patagonia in Southeast Pacific Ocean and Southwest Atlantic Ocean (Figure 1). Nine locations were sampled in the Chilean Patagonian fjords, one location in the Strait of Magellan and two locations on the Argentinian shelf. A small piece (i.e. 50-100grs) of muscular tissue was taken to each individual and stored in ethanol 95% for subsequent analyses.

DNA extractions, PCR and mtDNA amplification

Total genomic DNA was isolated from tissue samples using NucleoSpin Tissue Kit ®, following manufacture conditions. The quality and quantity of the DNA was measured in an Eppendorf biophotometer® and the template DNA was diluted to 20

ng/µL for the PCR amplifications. The universal primers, L19 (Bernatchez et al., 1992) and 12SAR (Palumbi, 1996) were used to amplified a fragment of >1400 basepairs (bp) from mitochondrial DNA, that included almost all Control Region (CR), Phenylalanine tRNA (Phe), and a fragment of 12S ribosomal RNA (12S). mtDNA was amplified in 30 µL reactions containing 1X PCR buffer, 1.5 mM $MgCl_2$, 0.2 mM dNTP's, 0.2 µM of each primer, 0.24 U/µL Taq DNA polymerase (Invitrogen®) and 20 ng of genomic DNA template. PCR amplification was performed in a PTC-200 (MJ Research) thermal cycler with the following parameters: 94 °C for 50 s, followed by 32 cycles of 95 °C for 30 s, 54 °C for 120 s, 72 °C for 90 s, and a final extension at 72°C for 300 s. PCR products were visualized using ethidium bromide stain and posteriorly, both strands were sequenced, forward and reverse, using a ABI 3730xl BigDye Terminator Cycle Sequencing 3.1 (Applied Biosystems) standard protocol.

Phylogeographic analyses of mtDNA sequences data

Initial alignments were performed with the SEQUENCHER™ program (Gene Codes, Ann Arbor, MI), and final alignments were adjusted by eye. Due to that inherence of mitochondrial DNA is genome-complete we performed posterior analyses with whole sequences (CR+Phe+12S). Best evolution sequence model that fitted our dataset was evaluated using the Akaike Information Criterion (AIC) in Jmodeltest 2 software (Darriba et al., 2012). The best model was TrN+I+G (I=0.834, G=0.451). We conducted genetic diversity measures such as haplotype number (*H*), haplotype diversity (*h*), segregating sites (*S*) and nucleotide diversity (π), in DNAsp v5.0 software (Librado and Rozas, 2009).

We applied a Bonferroni correction previous to carrying out analyses of pairwise genetic differentiation, given that the use of multiple comparisons increases the risk of false positives (Type I error; Rice, 1989). The new threshold to alpha (*a'*) was obtained as [$a' = a / k$], where a is 0.05, and k is the number of comparisons performed (one locus * 12 locations), therefore $a' = 0.004$.

We used the pairwise F_{ST} index to determine genetic differences between pairwise locations. In addition, we conducted an Analyses of Molecular Variance (AMOVA) in order to quantify genetic variation among locations (F_{ST}), among locations within a group (F_{SC}) and among groups (F_{CT}). We tested as alternatives to the AMOVA, Panmixis throughout the whole study area, Atlantic Ocean samples versus Pacific Ocean samples, and finally Atlantic Ocean plus Magellan strait samples and other Pacific Ocean samples. Significance of F-statistics was achieved using 10,000 permutations. Both analyses, pairwise F_{ST} and AMOVA were conducted in ARLEQUIN 3.1 software (Excoffier et al., 2005).

Intraspecific phylogenetic analysis was conducted using a Bayesian Markov Chain Monte Carlo approach (BMCMC), which incorporates the uncertainty in the reconstruction of the phylogenetic tree. To BMCMC approach, we applied a general likelihood-based mixture model of gene-sequence evolution as described by Pagel & Meade (2004, 2005). The latter because the molecular marker used in this study included fragments of CR, Phe and 12S genes that could have several patterns and rates of nucleotide substitution. This mixture model, based on the general time-reversible (GTR) model (Rodríguez et al., 1990), accommodates cases in which different sites in the alignment evolved in qualitatively distinct ways, but does not require prior knowledge of these patterns or partitioning of the data. We used the mixture model, implemented in a BMCMC framework, to estimate the posterior probability of the intraspecific phylogenetic trees and to include this information in further comparative method analyses. The Reversible-Jump Markov Chain Monte Carlo procedure (Pagel and Meade, 2008, 2006) was used with the objective of integrating results of all patterns, and produced an mixture model that summarized the sequence evolution, using BayesPhylogenies v1.1 software (http://www.evolution.rdg.ac.uk/BayesPhy.html). This approach enabled us to explore the variety of possible models and parameters, converging towards the model that best fits the data in the posterior tree sample (Pagel and Meade, 2008). Independent BMCMC analyses were ran using 10,000,000 generations of phylogenetic trees, sampling every 10,000th trees to assure that successive

samples were independent. From this sample of trees, the first 10 % were removed to avoid including trees sampled before the convergence of the Markov Chain. We obtained a final sample of 910 trees, and these trees were used to obtain the phylogenetic consensus tree, and their posterior probabilities.

Deviation from equilibrium were tested using Fu's F_S (Fu, 1997) neutrality test, which was conducted in DNAsp v5.0. Negative values of Fu's F_S are expected under a model of sudden population expansion. In addition, a Bayesian Skyride Plot (BSRP) analysis (Minin et al., 2008) was used to evaluate temporal oscillation of the effective population size (N_E), incorporated in BEAST v1.7.2 software (Drummond et al., 2012). We evaluate each clock model implemented in BEAST, Strict clock model, Log Normal Relaxed clock model, Exponential Relaxed clock model, and Random Local clock model before performing the BSRP (Drummond et al., 2006; Lemey et al., 2010). To determine the best clock model that fitted our dataset we used Bayes Factor (BF), which measures the weight of the evidence in the proposed model against another candidate model (Goodman, 1999). The best clock model that fitted our dataset was an Exponential Relaxed clock model (Drummond et al., 2006). We ran 10 independent Marcov Chain Monte Carlo (MCMC) with a length chain of 10,000,000 and sampled each 10,000 iterations, and we used priors as well as frequency of each base, proportion invariant sites and proportions of each transitions and tranversions, in order to increase the effective sample size (ESS). The Effective Sample Size (ESS) of a parameter sampled from an MCMC (such as BEAST) is the number of effectively independent draws from the posterior distribution that the Markov chain. We used a mutation rate between $1.1x10^{-7}$ (McMillan and Palumbi, 1997) per site per year used for the Control Region of *Sardinops pilchardus* (Atarhouch et al., 2006). Finally, we combined each independent run and burnt first 250 iterations using LogCombiner software include in BEAST.

Paleoreconstruction of environmental parameters and effective population size

In order to evaluate the effect of marine paleoenvironmental variables proxies on effective population size (N_E) a Pearson correlation coefficient (ρ) was calculated to the following variables: paleotemperature, paleosalinity and historical sea levels. Paleotemperature database (i.e. sea surface temperature) was obtained from a calibrated alkenona reconstruction (Prahl *et al.* 1988; Pahnke & Sachs 2006). Paleosalinity database was obtained from $\delta^{18}O_{SW}$ record from core MD07-3128 (Caniupán et al., 2011). Historical sea level was obtained from the SPECMAP stack (Imbrie et al., 1984). The lowest point of the sea level was 130 m at 19,000 years before present (YBP) (Huybrechts, 2002). We estimated a Pearson correlation coefficient to paleodata set, previous to LGM, posterior to LGM using as threshold the 21,000 YBP, and in whole range of BSRP estimated. The latter due to Pearson correlation coefficient works well with lineal variables. All correlations were performed in PAST v2.17 (Hammer et al., 2001).

Present-day environmental variables, comparative methods and intraspecific diversification analyses

Contemporary physical and chemical marine environmental variables (i.e. Temperature, Salinity, pH, Oxygen, Nitrate, Silicate, and Phosphate) from Patagonian Chilean fjords were obtained from the literature (Sievers et al., 2002; Silva and Calvete, 2002; Silva et al., 1998, 1997; Valdenegro and Silva, 2003), as well as from the marine environmental database of the Southwest Atlantic Ocean from Argentinian Naval Hydrographic Service (http://www.hidro.gov.ar/ceado/ceado.asp; Supplementary Table 2.4).

To test the effect of ancient environmental variables on individual patterns of intraspecific diversification rates (i.e. gain of haplotype rate minus loss of haplotype rate), we first searched for possible different patterns of intraspecific diversification in the consensus ultrametric tree of *S. fuegensis* using BAMM software, and then we used the QuaSSE method to test if these different patterns of diversification were caused by the

past environmental variables inferred from actual data. We used only those oceanographic environmental variables that had phylogenetic signal in *S. fuegensis*.

BMCMC molecular phylogeny and estimates of divergence times

Given that our hypothesis-testing framework was based on phylogenetic relationship over time, we reconstructed the intraspecific molecular phylogenies for *S. fuegensis*. We simultaneously estimated phylogenetic relationships with aligned sequences, branch lengths, and divergence times for the group using BEAST 1.7.5 software (Drummond and Rambaut, 2007). This analysis was conducted using a BMCMC framework to estimate the posterior probability of phylogenetic trees. As prior information we used a GTR+Γ+I model of sequence evolution, the Yule process of speciation, and a rate mutation of $1.1x10^{-7}$ (McMillan and Palumbi, 1997) per site per year used for the Control Region of *Sardinops pilchardus* (Atarhouch et al., 2006). Analyses were based on four models of mutation rate: 1) a strict molecular clock; 2) an uncorrelated lognormal relaxed clock; 3) an uncorrelated exponential relaxed clock; and 4) a random local clock. One independent MCMC chain was ran for 21,000,000 generations (discarding the first phylogenetic trees, before the posterior probabilities distribution of the selected diversification model converged), with parameters sampled every 10,000 steps, and we resampled at a lag where the final sample of trees were not autocorrelated in its likelihood values. Examination of MCMC sample was performed using TRACER v. 1.4 software (Rambaut and Drummond, 2007) to evaluate that the chain was adequately sampling the probability distribution; and that effective sample size for all parameters of interest were greater than 500. In order to find the best molecular clock model we used BF (Gelman et al., 1995) to compare the four clock models, given that it is the soundest theoretical framework for model comparison in a Bayesian framework (Drummond and Rambaut, 2007). Finally, to obtain the consensus ultrametric tree from the Bayesian sample, we used the TreeAnnotator v1.7.5 using the maximum clade credibility tree, and we kept target heights to discard negative branch lengths.

Inferring patterns of intraspecific diversification rates over time and between clades in *S. fuegensis*.

With Bayesian statistical inference, we explored a variety of time-varying models that could account for diversification rate heterogeneity in time and among clades with BAMM software (Bayesian Analysis of Macroevolutionary Mixtures; Rabosky, 2014). BAMM uses rjMCMC to automatically explore a vast universe of candidate models of lineage splitting, finding locations for shifts to new time-varying or diversity-dependent evolutionary rate regimes that are maximally supported by the data, with no a priori specification as to where these shifts in dynamics might have occurred. These shifts can also be abrupt in time, and a mixture of models (i.e. many distinct combinations of evolutionary shift regimes) might have roughly equal posterior probabilities to explain the data.

Testing the phylogenetic signal in present-day environmental variables

We evaluated the extent to which the phylogeny correctly predicts patterns of similarity in oceanographic environmental variables (i.e. phylogenetic signal, see Table 2.3) using the phylogeny scaling parameter, λ. In this approach λ reveals whether the phylogeny fits the patterns of covariance among species for a given trait. This parameter evaluates whether one of the key assumptions underlying the use of the comparative method (i.e. that species are not independent), fits the data for a given phylogeny and trait, assessing the strength of the phylogenetic signal. Values close to zero indicate that there is no concordance between phylogeny and the trait values of species (phylogenetic independence). If traits are evolving as expected, given the tree topology and branch lengths, λ takes the value of 1. Intermediate values of λ, between 0 and 1, indicate different degrees of a phylogenetic signal. We inferred the maximum likelihood values of lambda for each oceanographic variable based on the ultrametric tree (obtained by BEAST) in BayesTrait v2.0 software (Pagel, 1999, 1997). To test the significance of the phylogenetic signal we compared the estimated value with a model of lambda forced to 0

by BF. Only those variables that had significant phylogenetic signal were used to evaluate its past effect on diversification patterns.

Testing the effect of past environmental variables on intraspecific diversification patterns.

We tested the hypothesis that differences in intraspecific diversification rates were influenced by past climate using QuaSSE algorithm (FitzJohn, 2010), that takes a time-phylogeny and a set of continuous trait measurements for the tips. In order to fit a series of birth–death models in which the intraspecific diversification rates are independent of trait evolution, or vary along branches as a function of a continuous trait that evolves according to a diffusion process, with or without an evolutionary tendency (in this case the environmental variables with phylogenetic signal), we fitted models in which intraspecific diversification rates were invariant with respect to past variables. In these models, rates varied as linear, sigmoidal, or hump-shaped function respect to past variables, with and without a general tendency in the evolution of climatic variables that lineages occupy over time. Hence, we compared seven models of intraspecific diversification rates: 1) a constant model of haplotype accumulation; 2) haplotype accumulation that varies as a linear function of environmental variables, evolving by a diffusion process; 3) haplotype accumulation that varies as a sigmoidal function of environmental variables, evolving by a diffusion process; 4) haplotype accumulation that varies as a hump function of environmental variables, evolving by a diffusion process; 5) haplotype accumulation that varies as a linear function of environmental variables, evolving by a diffusion process with a directional trend; 6) haplotype accumulation that varies as a sigmoidal function of environmental variables, evolving by a diffusion process with a directional trend; and 7) haplotype accumulation that varies as a hump function of environmental variables evolving by a diffusion process with a directional trend. In a linear model, the intraspecific diversification rates varies proportional to environmental variables, in a sigmoidal model species that inhabited low values of environmental variables had a low diversification rate compared to species of high values of

environmental variables, and in a hump model, species that occupied middle values of environmental variables had the highest speciation rate. These models have the following parameters: the haplotype gain accumulation and haplotype los accumulation rate parameters (λ, μ); the diffusion parameter (σ2), which is the expected squared rate of change and captures the stochastic elements of trait evolution; and the directional trend ‘‘drift’’ parameter (θ), which captures the deterministic or directional component of trait evolution, this is the expected rate of change of the character over time and may be due to selection or any other within-lineage process that has a directional tendency (FitzJohn, 2010). These models were implemented in QuaSSE algoritm, and the analyses were done using the Diversitree package of R software (FitzJohn, 2012).

Section 4: Results

Sequences information and genetic diversity indices

We obtained a sequence of 1438 bp that included 1046 bp of Control Region gene (CR), 69 pb of Phenylalanile (Phe) and 323 bp of 12S gene. Our CR fragment is the 97% of the whole CR of *Sprattus fuegensis*. The whole sequence had 1269 conserved positions, 169 variable positions and 69 singletons. The whole fragment sequence showed 304 haplotypes, Hd = 0.999, S = 169 and π = 0.726% (Table 2.1).

Population differentiation index and phylogenetic reconstruction

Pairwise F_{ST} showed only seven significant comparisons at p-value of 0.05. All this significant comparison were between ARG locations and Zones D, H, I, J and K (see Table 2.2). In the post application of Bonferroni’s test, only one comparison was significant (ARGA-Zone H) ($p<0.003$) (Table 2.2). AMOVA showed non-significant differences among hypotheses tested. Analysis based on Panmixia hypotheses showed an F_{ST} index -0.0013 (p=0.304), between Atlantic samples (ARGA-ARGB) and Pacific Samples (Fjord+Magellan) showed an F_{CT} index -0.0013 (p=0.304). Finally, Fjord samples and Atlantic+Magellan samples showed an F_{CT} index -0.0025 (p=0.382). The

BMCMC phylogenetic reconstruction showed a large population with a high posterior probability (Figure 2A,B). In addition we observed several groups in this tree but with a low posterior probabilities (< 0.5).

Demographic analyses

Fu's F_S test was significant in almost all localities, indicating departure from mutation-drift equilibrium, and possible population expansion (F_S=-685.8, P<0.001 to CR+12S) (Table 2.1). The BSRP analysis showed two effective population size expansions in *S. fuegensis* followed by constant population size after 20,000 YBP. The first increment pulse on effective population size occurred between 50,000 and 60,000 YBP and the second one between 20,000 and 30,000 YBP (Figure 2.3A).

Paleoreconstruction of environmental parameters and effective population size

All environmental paleoreconstructions showed a significant correlation with effective population size of *S. fuegensis*. Sea surface temperature (Figure 3B) showed a positive relationship (ρ = 0.8868, p-value = $1.50x10^{-10}$ and R^2 = 0.79) during the last 21,000 YBP and a negative relationship (ρ = -0.5658, p-value = $2.71x10^{-07}$, and R^2 = 0.32) between 70,000 to 21,000 YBP. Salinity (Figure 3C) showed a negative relationship (ρ = -0.9362, p-value = $7.51x10^{-13}$ and R^2 = 0.88) between 21,000 YBP to the present. Finally, ancient sea-level (Figure 3D) showed a negative relationship (ρ = 0.6642, p-value = $8.50x10^{-5}$ and R^2 = 0.44) between 21,000 YBP to the present and a positive relationship (ρ = 0.7261, p-value = $7.72x10^{-13}$, and R^2 = 0.53) between 70,000 to 21,000 YBP.

Inferring patterns of intraspecific diversification rates over time and between clades in *Sprattus fuegensis*.

The rjMCMC algorithm implemented in BAMM indicated that the intraspecific diversification of *S. fuegensis* was best described by one evolutionary patterns generated by an exponential decreased in haplotype accumulation rate in *S. fuegensis* history (Figure 4A,B). This pattern is consistent with a slowdown in intraspecific diversification over time showing a high haplotype accumulation rate (≈ 70 lineages per 100,000 YBP) early in its history (Figure 4B). Haplotype loss rate was constant and haplotype accumulation rate decreased trough time (Figure 4B).

Testing the effect of past environmental variables on intraspecific diversification patterns.

The variables that showed significant phylogenetic signal were maximum temperature; average and maximum salinity; range pH; minimum and range PO_4; maximum, minimum and range NO_3^- (Table 2.3), indicating that in *S. fuegensis* these current environmental variables have been determined by the environment in which the ancestral individuals inhabited.

The evolution of average and maximum salinity; range PO_4; maximum, minimum and range NO_3^- of the environment inhabited by *S. fuegensis* were associated to the intraspecific diversification patterns as follows: average salinity evolved with a positive directional trend, and was related lineally to haplotype accumulation rates with a negative slope; maximum salinity evolved with a positive directional trend, and was related lineally to haplotype accumulation rates with a negative slope; range PO_4 evolved with a positive directional trend, and was related sigmoidally to haplotype accumulation rates; maximum NO_3^- evolved with a positive directional trend, and was related to haplotype accumulation rates as a hump function; minimum NO_3^- evolved with a negative directional trend, and was related lineally to haplotype accumulation rates with a positive slope; finally range NO_3^- evolved with a positive directional trend, and was related sigmoidally to haplotype accumulation rates (Table 2.3).

Section 5: Discussion

Our results support the hypothesis that ancient marine environmental variables played a key role on the microevolutionary diversification rate of *Sprattus fuegensis*, and therefore environmental changes that occurred in the past resulted in changes on genetic diversity and effective population size of this species.

Genetic variability and population structure

Our analyses showed that *S. fuegensis* has high genetic diversity based on haplotype diversity and low nucleotide diversity (Table 2.1). Prior works in small pelagic marine fishes related to *S. fuegensis* and that inhabit other parts of the world, had also documented high genetic diversity. Among them, *Sprattus sprattus* (Debes et al., 2008), *Clupea pallasi* (Liu et al., 2011), *Sardina pilchardus* (Atarhouch et al., 2006), *Engraulis japonicus* and *Engraulis australis* (Liu et al., 2006). The high genetic diversity found in these species may be due to a high mutation rate, of the Control Region in teleost fishes (Lee et al., 1995; Vodolazhskii et al., 2008). A mechanism that have been suggested in sturgeons and that could explain the high polymorphism observed, may be related to frequent erroneous comparison of nucleotide bases in the region of repeats before the elongation is promoted by the stable secondary structures in the D-thread of mtDNA (Vodolazhskii et al., 2008). Evolutionary rates hypothesis has suggest that environmental factors (i.e. temperature) increase mutation rate in mtDNA in vertebrates (Evans and Gaston, 2005). However, this pattern has been well documented in species that inhabit warm waters which are high-energy areas (Evans and Gaston, 2005), and it might not apply to *S. fuegensis* which is associated to subantartic water (less-energy area). The metabolic rate theory which postulates an inverse relationship between mutation rate and body mass (Rand, 1994), and that has been used to explain the high mutation rate of small homoeothermic and poikilotherms organism (Martin and Palumbi, 1993; Penzo et

al., 1998). The possibility to accumulate DNA mutations in marine vertebrates with small body size is faster than in bigger organisms (Martin and Palumbi, 1993; Penzo et al., 1998; Rand, 1994). In *Sprattus sprattus* environmental temperature and body mass have a significant effect on the metabolic rate of individuals (Meskendahl et al., 2010). In addition, if we take into account that generational time in small pelagic marine fishes is relatively short (i.e. few years), this biological characteristic may increase the probability that new mutations will accumulate faster (Martin and Palumbi, 1993) and therefore the species will show high mutation rates. Finally, living in environments like fjords with extreme changes in oceanographic conditions (i.e. temperature and salinity) could yield physiological stress, which could potentially increase the mutation rate. However, at the moment there is no data on physiological stress in Clupeiformes that would allowed this hypothesis to be tested.

We found no genetic population structure *S. fuegensis*, and we observed a large panmictic population with a remarkably high genetic diversity, suggesting that there is no environmental effect on spatial genetic structure in this species. Different patterns of intraspecific lineage diversification have been found in the Order Clupeiformes (Atarhouch et al., 2006; Debes et al., 2008; Grant and Bowen, 1998; Grant et al., 2012; Liu et al., 2011, 2006). For instance, using mtDNA Liu et al. (2011) showed that *Clupea pallasii* had a strong population structure between Yellow Sea and Japan Sea, indicating that high sea temperature would be acting like a barrier. In *Sprattus sprattus*, Debes et al. (2008) found strong genetic divergence between individuals sampled from Mediterranean Sea and those sampled from the Baltic Sea, the North Sea, the Bay of Biscay, the Black Sea and the Bosporus. Intraspecific lineages of *S. sprattus* (Debes et al., 2008) could reflect different dispersal strategies or be the product of changes of historic key oceanographic parameters like salinity and temperature. For example, in *C. pallasi* the three mayor lineages coincided with rapid climate changes and oceanographic shifts that typified the mid-Pleistocene (Liu et al., 2011). Nonetheless, it is common to not find genetic differences in fishes along the Chilean distribution, what has been

explained by: close spawning areas (i.e. *Genypterus blacodes*; Canales-Aguirre *et al.* 2010), large population size (i.e. *Trachurus murphyi*; Cárdenas *et al.* 2009b; Galleguillos *et al.* 2012), trophic and reproductive migratory behavior (i.e. *Micromesistius australis*, Canales-Aguirre *et al*. in prep), synchronic spawning, few or none oceanographic barriers, egg and larvae drift by mean current (*Strangomera benktincki*; Galleguillos *et al.* 1997) and conservative genetic variability in nuclear genes (*Engraulis ringens*; Ferrada et al., 2002). Microchemistry analyses from the core of otoliths have shown two well differentiated juvenile nursery areas in Chilean Patagonia; however, microchemistry from edge of the otolith showed a mixed adult population (Galleguillos *et al*. 2012). Traditional and geometrics morphometric analyses have revealed differences in the same area; however a high mixed between both areas has been described as well (Galleguillos *et al*. 2012). These two latter approaches may be suggesting that adults of *S. fuegensis* are able to move between areas, thus homogenizing the population which supports the outcome of our study.

Correlation between marine paleoenvironmental data set and BSRP

Our results support an increase in effective population size through time in *S. fuegensis*, suggesting that the environmental changes played a key role in the evolution of this species. In fact, there are several variables that are associated to this demographic change. The main changes of effective population size were generated previous to LGM where posteriorly the N_E showed stabilization in *S. fuegensis*. Our analyses, demonstrated that this clear sudden increase in effective population size previous to LGM, is correlated to a decrease in temperature, salinity and sea level. Therefore, we suggest three scenarios that would explain the effect of environmental factor in the increase of N_E posterior to LGM : A) temperature scenario, B) salinity scenario and C) decreased sea-level scenario.

a) Temperature scenario

Based on water temperatures that currently inhabit *S. fuegensis* (i.e. 8-12ºC; Sánchez *et al.* 1995) and taking into account information of paleoceanographic

temperatures from southeast Pacific Ocean (Pahnke and Sachs, 2006) and our results from Figure 3B, we could hypothesized that low N_E around 70,000 YBP may be related to high temperatures. Posteriorly, decreasing water temperature promoted habitat expansion for cold-water species, and therefore an increase in N_E. Temperature is a key environmental factor that regulates demographic population parameters. For instance, the temperature is an important factor in setting an upper limit to the vertical sprat distribution (Stepputtis et al., 2011). Moreover, it plays a key role in egg development and hatching in *S. sprattus* (Petereit et al., 2008), where an optimal thermal window for growth and survival for eggs of 5–17 ºC, and for young early feeding larvae of 5–12 ºC, have been indicated (Peck et al., 2012a). This is similar to other related species such as *Clupea harengus*, which has a optimal thermal window that ranges between 7-13ºC (Peck et al., 2012b). Evidence shows that recruitment in *S. sprattus* is influenced by fluctuations in biotic and abiotic conditions (Petereit et al., 2009), among which temperature has been significantly positivily correlated with recruitment in the Baltic Sea (MacKenzie et al., 2008). Nevertheless, recruitment in pelagic fishes like sprat, cod and herrings is related to the spawning stock biomass and climatic forcing (i.e. temperature; Margonski et al., 2010). However, Peck et al. (2013) suggested that changes in temperature are important but only partially induce population levels shift in small pelagic fishes. Recently, Canales-Aguirre *et al.* (in prep) found no relationship between current SST and allelic frequency in microsatellite loci of *S. fuegensis*, suggesting that SST has no short term significant affect on N_E. In contrast, our results suggest that SST has a long-term affect on N_E (i.e. evolutionary scale, thousand years). Small pelagic fishes tend to search cooler and deeper waters when temperature is not favourable (e.g. Ñiquen & Bouchon 2004).

b) Salinity scenario

Salinity data showed a strong and highly significant correlation with N_E (Figure 3C), which could explain the halotolerancy showed by some *Sprattus* species in other studies (Landaeta et al., 2012; Petereit et al., 2009). Salinity has been related to spawning areas in *S. sprattus*, most probably because salinity can affect the density of fertilized or

incubated eggs, and also influence egg buoyancy patterns (Petereit et al. 2009). Recent studies in *S. fuegensis* larvae has shown that the otoliths growth index is not correlated with salinity, indicating that larvae grow up similarly in stratified and mixed waters (Landaeta et al., 2012). The high correlation between salinity and N_E (Figure 3C) suggests that the sprat ancestor may have acquired salinity tolerance. Salinity has been suggested as having its largest ecological impact during spawning, because it determines the vertical habitat where eggs and yolk sac larvae buoyance (Petereit et al., 2009). Ojaveer and Kalejs (2010) showed that changes in abundance in *S. sprattus* are related to freshwater discharges and variation in water volume of its habitat . Difference in salinity have also been suggested as key reproductive barrier between the North Sea and the Baltic Sea (Limborg et al., 2009). This is due to a stronger correlation between genetic distance and difference in salinity on spawning sites in *S. sprattus*. By contrast, Glover et al. (2011) showed that in *S. sprattus* from Norwegian fjords, genetic differences might be related to limited connectivity between Norwegian fjord and sea-going populations.

C) Decreased sea-level scenario

We found that the decrease in sea level was also related to N_E. This pattern has also been shown in other species. For instance, Porobić et al. (2013) suggested that the drop in sea level promoted more habitat availability in the seamount and therefore increase on N_E in the Juan Fernandez' rock lobster,*Jasus frontalis*. Conversely, in *Nacella concinna* sea level decrease could have affected dramatically habitat availability due to glacial cycles, causing a decrease in its genetic diversity and isolating them or generating changes in its distribution (González-Wevar et al., 2011). In the congeneric species *S. sprattus,* Debes et al. (2008) proposed that a bottleneck in the Adriatic population was the result of a decreased in sea level. In *S. fuegensis,* as well as others small pelagic fishes that live close to the coast and on continental shelf (Aranis et al., 2007; Cousseau, 1982; Leal et al., 2011; Whitehead, 1985), the decrease in habitat that would have been produced by the extensive icesheet (DaSilva et al., 1997; Hulton et al., 2002) should have decreased Ne. However, our results showed that before the LGM in

the Chilean Patagonia there was an increase in N_E associated to a decrease in the sea level (Figure 3D). We propose that *S. fuegensis* may have migrated to the extensive continental shelf in the southwest Atlantic, thus avoiding mortality from adverse environmental conditions.

Historical marine environmental data and intraspecific diversification

This is the first time that diversification analyses at the intraspecific levels are evaluated together with historical oceanographic variables. Based in BAMM we could detect an exponential decrease in the intraspecific evolutionary diversification rate (Figure 4B), and a decrease in the number of haplotype through time. In relation to the historical oceanographic variables, we found that nitrate and salinity are related to the intraspecific evolutionary diversification rate of the species (Table 2.3).

On the one hand, our data suggests that nitrate had a positive effect on the intraspecific diversification rate of this species (Table 2.3). Nitrate (i.e. NO_3^-) concentrations have been pointed out as the major limiting factor in marine ecosystem (Frank et al., 2005) by affecting the basal links in marine food webs. At a macroevolutionary level, there is a correlation between genus origination rates and nutrients availability (Cárdenas and Harries, 2010). Thus, variation in nutrients may have led to the formation, persistence and differentiation of isolated populations, and then to speciation events (Allmon, 2001). In a microevolutionary context our results support a relationship between nutrients and intraspecific diversification rates of *S. fuegensis*. Nutrients' input from continental waters to the ocean have been proposed as having critical evolutionary impact over species taxon diversification (Martin, 1996), and at a intraspecific level could have play a key role in generating new lineages and on maintaining the intraspecific evolutionary diversification pattern.

On the other hand, salinity showed a negative effect in relation to the diversification rate of this species (Table 2.3). Euryhalinity species has been described as having a key innovation that would confer them with an evolutionary advantage to adapt and diversify (Schultz and McCormick, 2013). This key innovation, would allow

euryhaline species to colonize new habitat or increase it geographical range promoting population divergence (Schultz and McCormick, 2013). Congeneric *Sprattus* species are associated to marine habitats, and their capacity to inhabit estuarine and fjords make them highly tolerant to salinity changes (Landaeta et al., 2012; Limborg et al., 2009; Petereit et al., 2009). For instance, *S. sprattus and S. fuegensis* have being reported to have the capacity to tolerate low salinities (Petereit et al., 2009)(Landaeta et al., 2012). In intraspecific analyses, salinity gradient has been associated to population divergence in *S. sprattus* (Limborg et al., 2009) and *C. harengus* (Bekkevold et al., 2005; Jørgensen et al., 2005). Previous studies at the interspecific (Lavoué *et al.* 2013) and intraspecific level (e.g. Jørgensen *et al.* 2005; Bekkevold *et al.* 2005; Limborg *et al.* 2009), in addition to our results on the phylogenetic signal related to paleosalinity data (Table 2.3, Figure 2.3C), suggests that salinity has played an important historical role in this group, at both the species and population level.

Two bigger models has been applied to explain diversity at higher taxonomic level, the Red Queen hypothesis which is related to biological interactions that drive diversification (VanValen, 1973) and the Court Jester hypothesis where environment plays a key role in the evolutionary diversification patterns (Barnosky, 2001). These hypotheses are not mutually exclusive (Benton, 2009), because taxon evolutionary history is the product of the interactions of species diversity, ecology and climate (e.g. Ezard *et al.* 2011). Although these hypotheses were developed to be applied at the interspecific level, here we demonstrated that we can applied this conceptual framework to intraspecific level, because individuals from populations also interact with biological and environmental factors that could be able to shape its evolutionary patterns. Our results in *S. fuegensis* are in agreement with the Court Jester Hypothesis (Barnosky, 2001) due to the high influence of current environmental factors on this species (Canales-Aguirre *et al.* in prep), as well as ancient environmental factors based on paleodata from proxies (i.e. paleotemperature, paleosalinty, ancient sea level) and historical signal obtained using comparative methods approach.

Finally, our use of the comparative phylogenetic methods and the quantification of intraspecific diversification rates showed an effect of several environmental factors on the historical demography the *Sprattus fuegensis*, thus disentangling the role of environmental oceanographic variables at the intraspecific level. This study shows that incorporating the conceptual framework of diversification rate (i.e. speciation and extinction) into a microevolutionary level (i.e. loss and gain haplotype) can help to understand how biotic or abiotic factors promote speciation

Section 6: Tables

Table 2.1: Molecular diversity indices for mitochondrial variation at the sampling locations (Zones) considered.

N = sample size. *h* = number of haplotypes; *S*= number of segregates sites; *Hd* = haplotype diversity; π = nucleotide diversity; F_S = Fu's test and F_S Fu's p-value

Locality	N	h	S	Hd	π	Fs Fu	p-value
Zone A	12	12	46	1.000	0.0085	-4.21	0.019
Zone B	18	18	47	1.000	0.0068	-10.51	0.000
Zone D	8	7	27	0.964	0.0060	-0.74	0.273
Zone E	16	15	36	0.992	0.0064	-6.27	0.007
Zone H	18	16	40	0.987	0.0067	-5.63	0.014
Zone I	17	16	43	0.993	0.0075	-6.20	0.007
Zone J	17	16	46	0.993	0.0074	-6.28	0.008
Zone K	18	17	39	0.993	0.0074	-7.11	0.004
Zone L	9	9	37	1.000	0.0084	-2.36	0.070
Zone N	9	9	40	1.000	0.0091	-2.19	0.075
Arg A	124	118	109	0.999	0.0073	-24.33	0.000
Arg B	69	65	81	0.998	0.0067	-24.61	0.000
All localities	335	304	169	0.999	0.0073	-685.8	0.000

Table 2.2: Pairwise location comparisons of all samples.

The lower triangular matrix shows pairwise F_{ST} values and the upper triangular matrix shows p-values from the pairwise heterogeneity comparisons. Values in bold type show significant comparisons after correction for multiple tests was non-significant ($p \leq 0.004$). All p-values based on 10,100 permutations.

	Zone A	Zone B	Zone D	Zone E	Zone H	Zone I	Zone J	Zone K	Zone L	Zone N	Arg A	Arg B
Zone A	*	1.000	0.147	0.501	0.245	0.489	0.505	0.500	1.000	1.000	0.639	0.399
Zone B	0.000	*	0.087	0.204	0.234	0.236	0.231	0.493	1.000	1.000	0.530	0.308
Zone D	0.017	0.016	*	0.286	0.088	0.104	0.112	0.102	0.209	0.209	0.022	0.015
Zone E	0.004	0.004	0.013	*	0.355	0.178	0.390	0.168	0.519	0.522	0.114	0.291
Zone H	0.007	0.007	0.023	0.004	*	0.107	0.106	0.114	0.369	0.360	**0.003**	0.025
Zone I	0.004	0.004	0.020	0.008	0.010	*	0.233	0.183	0.539	0.527	0.024	0.084
Zone J	0.004	0.004	0.020	0.004	0.010	0.007	*	0.182	0.528	0.523	0.060	0.043
Zone K	0.003	0.003	0.020	0.007	0.010	0.007	0.007	*	0.542	0.541	0.027	0.086
Zone L	0.000	0.000	0.018	0.004	0.007	0.004	0.004	0.003	*	1.000	0.728	0.463
Zone N	-0.009	0.000	0.018	0.004	0.007	0.004	0.004	0.003	0.000	*	0.724	0.467
Arg A	0.000	0.000	0.014	0.002	0.007	0.004	0.003	0.004	0.000	0.000	*	0.137
Arg B	0.001	0.001	0.017	0.001	0.006	0.004	0.005	0.003	0.001	0.001	0.001	*

Table 2.3: Environmental variables used to test phylogenetic signal (λ) and to test an association with diversification rate.

Ave = average value; *Max* = maximum value; *Min*= minimum value; *Ran* = Range of the value

		(λ)	Diversification	Model	Slope
Temperature	Ave	no	-	-	-
	Max	**yes**	no	-	-
	Min	no	-	-	-
	Ran	no	-	-	-
Salinity	Ave	**yes**	**yes**	Lineal	**-4,99E+03**
	Max	**yes**	**yes**	Lineal	-5,88E+03
	Min	**no**	-	-	
	Ran	**no**	-	-	
Oxygen	Ave	no	-	-	-
	Max	no	-	-	-
	Min	no	-	-	-
	Ran	no	-	-	-
pH	Ave	**no**	-	-	
	Max	**no**	-	-	
	Min	**no**	-	-	
	Ran	**yes**	no	-	
PO4	Ave	no	-	-	-
	Max	no	-	-	-
	Min	**yes**	no	-	-
	Ran	**yes**	**yes**	Sigmoidal	-
NO3	Ave	**no**	-	-	
	Max	**yes**	**yes**	Hump	
	Min	**yes**	**yes**	Lineal	**3,40E+01**
	Ran	**yes**	**yes**	Sigmoidal	
Si	Ave	no	-	-	-
	Max	no	-	-	-
	Min	no	-	-	-
	Ran	no	-	-	-

Section 7: Figures

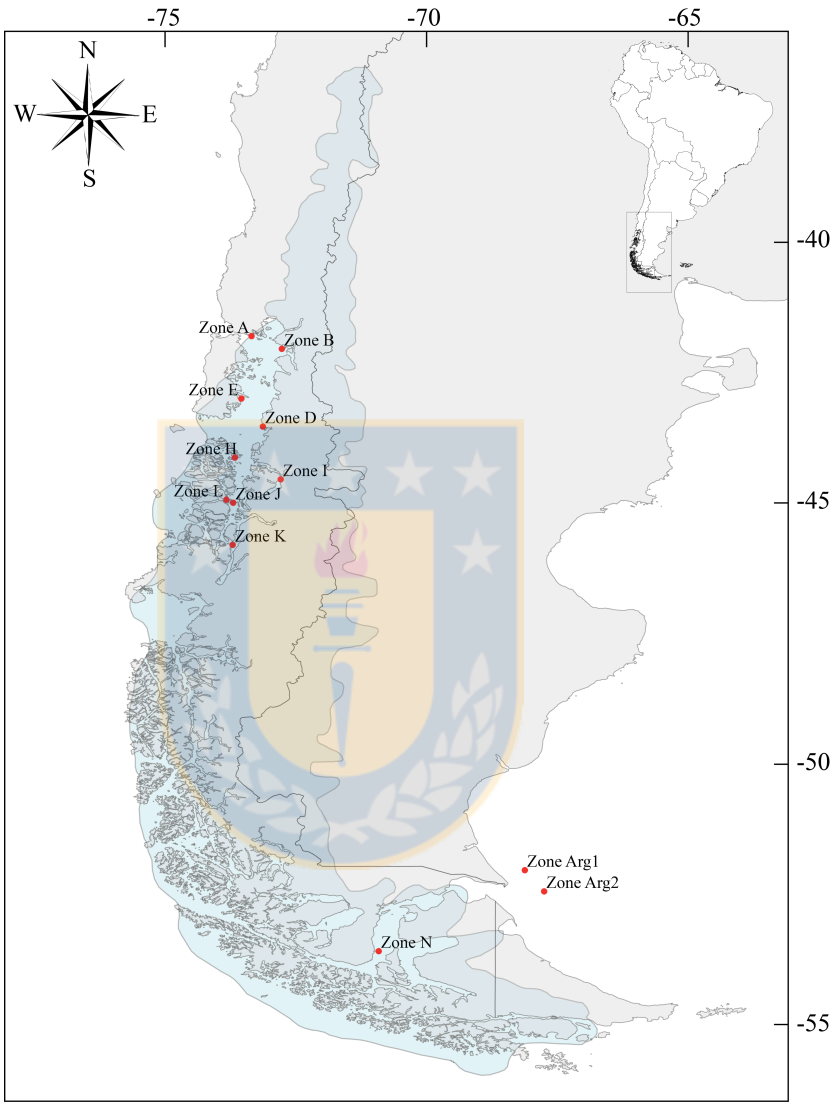


Figure 2.1: Map indicating sampling locations from Patagonian fjords where *Sprattus fuegensis* are distributed.

Red dots showed sampling locations. Light blue layer showed the maximum coverage extension of Last Glacial Maximum.

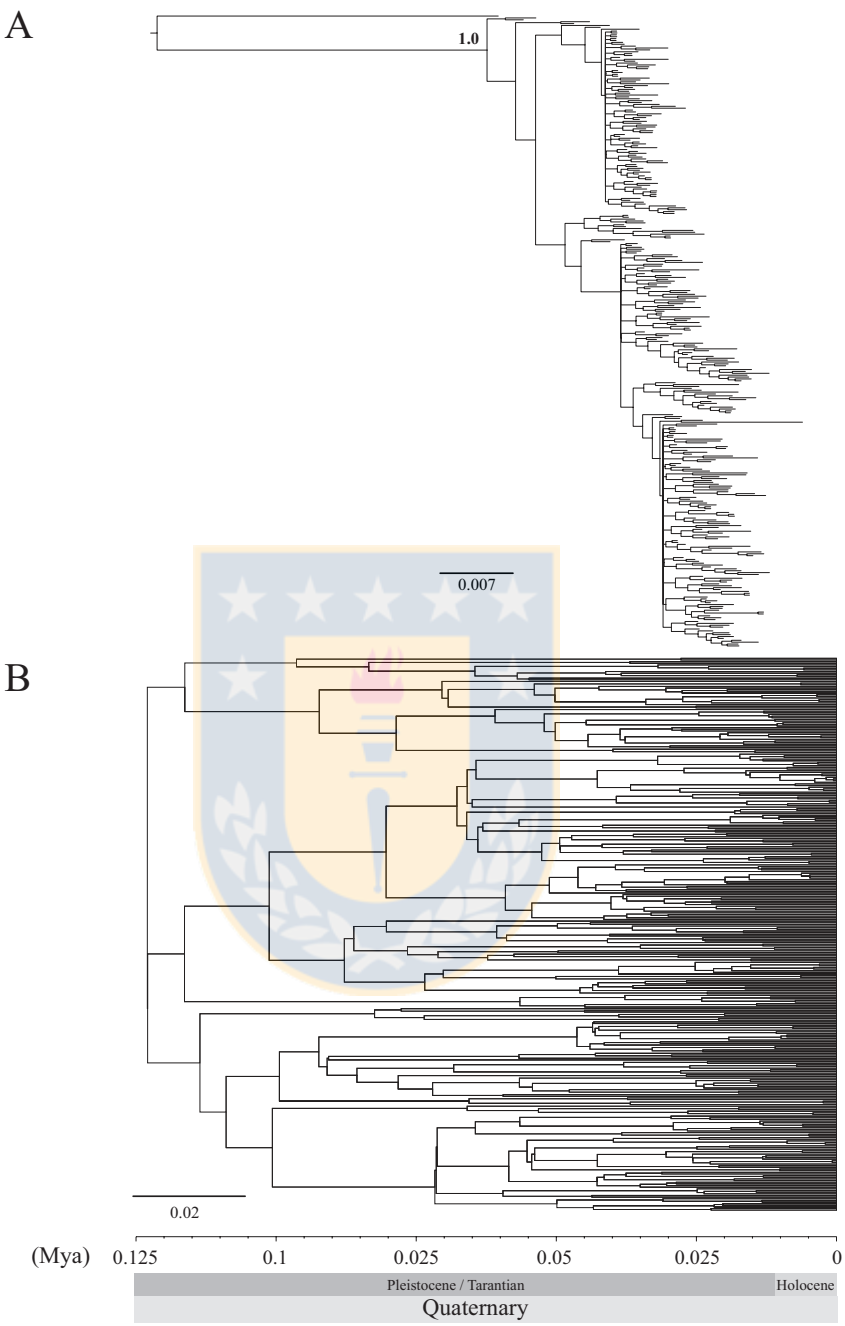
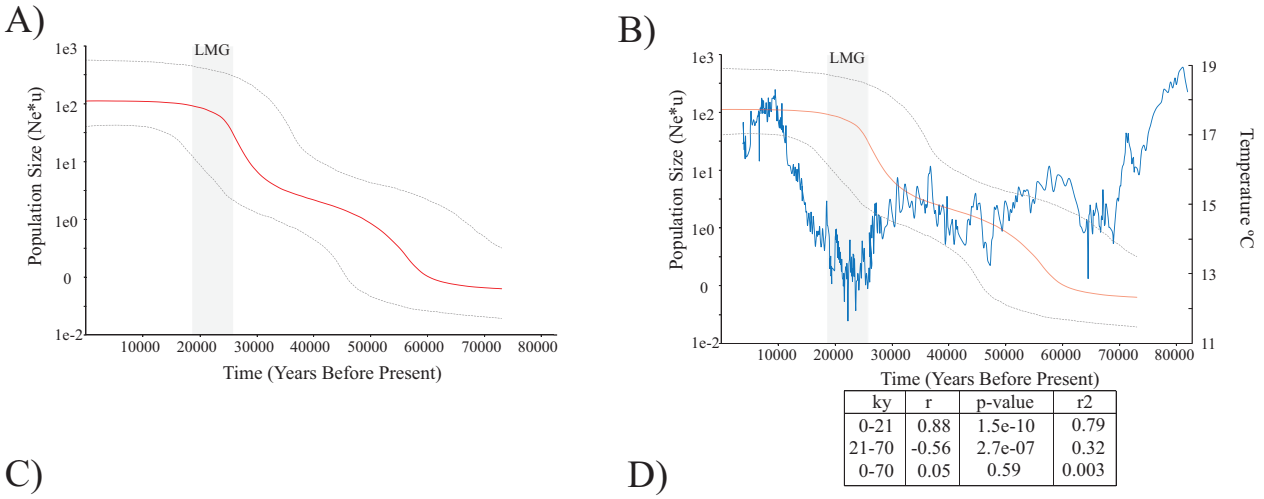


Figure 2.2: Bayesian phylogenetic reconstructions of intraspecific tree.

A) Intraspecific Bayesian consensus tree of *Sprattus fuegensis*. B) Bayesian consensus tree obtained from ultrametric trees based on an uncorrelated exponential relaxed clock.



ky	r	p-value	r2
0-21	0.88	1.5e-10	0.79
21-70	-0.56	2.7e-07	0.32
0-70	0.05	0.59	0.003

ky	r	p-value	r2
0-21	-0.93	7.5e-13	0.88
-	-	-	-
0-25	-0.69	4.5e-06	0.48

ky	r	p-value	r2
0-21	-0.66	8.5e-05	0.44
21-70	0.72	7.7e-13	0.53
0-70	-0.30	0.002	0.09

Figure 2.3: Correlations between effective population size (N_E) and paleoenvironmental variables during the last 80,000 YBP.

Correlations were carried out previous and posterior data of Last Glacial Maximum and whole data set. A) Bayesian Skyride Plot (red line), B) Temperature (blue line), C) Salinity (yellow line), D) Sea level (green line). Gray zone indicate period where LGM occurs.

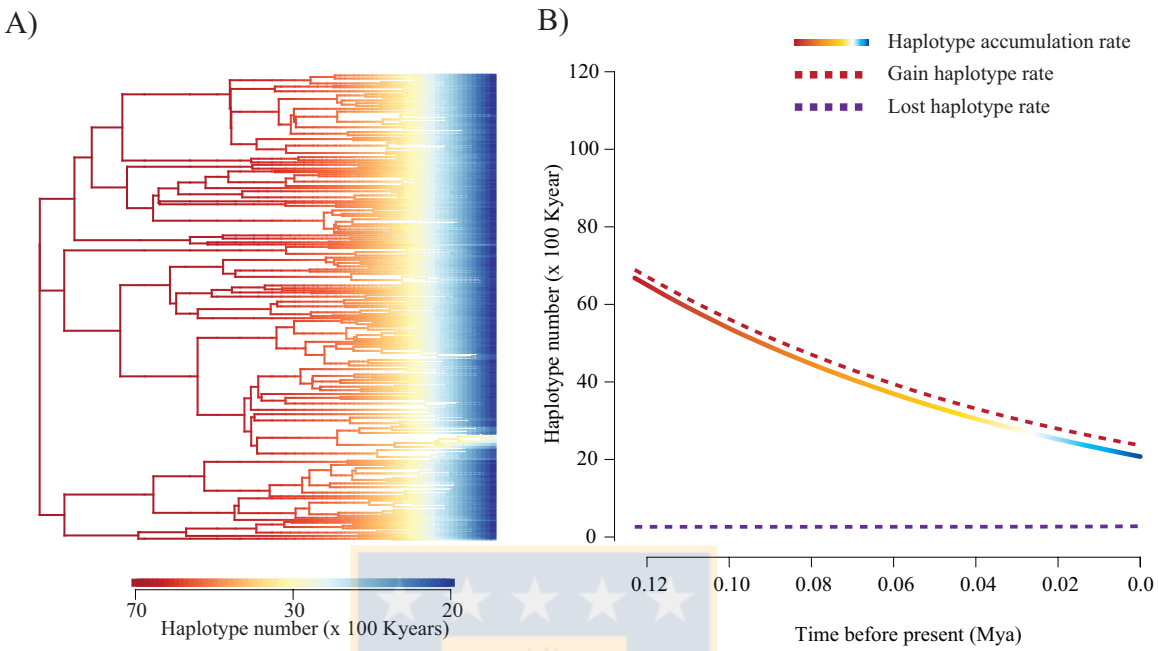


Figure 2.4: Intraspecific diversification rate analyses.

A) Phylorate plot showing haplotypes accumulation rate rates along each branch of the intraspecific phylogeny of *Sprattus fuegensis*. Cool colors = slow, and warm colors = fast. B) Haplotypes accumulation rate (gradient color line), haplotype gain rate (red dashed line) and haplotype loss rate (violet dashed line) through the time.

Section 8: Supplementary tables

Table 2.4: Environmental variables used in our analyses.

Ave: average, Rang: Range, Max: Maximum, Min: Minimum

Variable	Parameter	Zone_A	Zone_B	Zone_D	Zone_E	Zone_H	Zone_I	Zone_J	Zone_K	Zone_L	Zone_N	Arg1	Arg2
Temperature (ºC)	Ave	10.7	10.6	10.6	9.9	10.1	9.8	9.7	9.4	9.8	7.1	6.5	8.2
	Rang	1.5	1.0	2.8	0.2	1.5	2.0	0.5	0.5	0.5	0.5	5.4	3.2
	Max	11.5	11.0	11.9	10.0	11.0	11.0	10.0	9.5	10.0	7.5	9.8	10.0
	Min	10.0	10.0	9.1	9.8	9.5	9.0	9.5	9.0	9.5	7.0	4.4	6.8
Salinity (psu)	Ave	31.6	31.6	32.7	32.9	31.8	29.1	30.3	29.0	30.3	30.6	32.8	32.6
	Rang	2.5	2.5	2.3	0.2	3.0	13.5	5.0	2.0	5.0	0.5	1.3	1.0
	Max	32.5	32.5	33.6	33.0	33.0	33.5	33.0	30.0	33.0	31.0	33.5	33.0
	Min	30.0	30.0	31.3	32.8	30.0	20.0	28.0	28.0	28.0	30.5	32.2	32.0
Oxygen (ml/L)	Ave	6.8	6.1	5.4	6.0	6.6	5.5	5.5	5.9	5.5	6.5	6.5	5.8
	Rang	3.0	2.0	3.3	0.2	3.0	4.0	2.0	0.5	2.0	0.0	2.5	2.5
	Max	8.0	7.0	7.1	6.0	8.0	7.0	6.0	6.0	6.0	6.5	7.3	6.6
	Min	5.0	5.0	3.8	5.8	5.0	3.0	4.0	5.5	4.0	6.5	4.8	4.1
pH	Ave	7.9	7.8	7.8	7.8	7.9	7.8	7.7	7.7	7.7	7.8	8.1	-
	Rang	0.3	0.2	0.2	0.0	0.4	0.5	0.2	0.1	0.2	0.1	0.2	-
	Max	8.0	7.9	7.9	7.8	8.1	8.0	7.8	7.8	7.8	7.8	8.2	-
	Min	7.7	7.7	7.7	7.8	7.7	7.5	7.6	7.7	7.6	7.7	8.0	-
Phosphate (µM)	Ave	1.3	1.4	2.2	1.6	1.0	1.2	1.5	1.6	1.5	0.7	1.4	1.0
	Rang	1.0	1.0	0.4	0.4	1.2	2.4	0.8	0.0	0.8	0.4	1.0	1.1
	Max	1.8	1.8	2.4	1.6	1.6	2.4	2.0	1.6	2.0	1.0	1.7	1.7
	Min	0.8	0.8	2.0	1.2	0.4	0.0	1.2	1.6	1.2	0.6	0.7	0.6
Nitrate (µM)	Ave	11.0	13.5	22.5	16.0	9.0	10.0	15.0	16.0	15.0	4.5	12.4	-
	Rang	16.0	12.0	4.0	0.0	20.0	24.0	8.0	0.0	8.0	4.0	9.4	-
	Max	20.0	20.0	24.0	16.0	20.0	24.0	20.0	16.0	20.0	8.0	16.5	-
	Min	4.0	8.0	20.0	16.0	0.0	0.0	12.0	16.0	12.0	4.0	7.1	-

Chapter 3 – Origen and historical biogeography of the genus *Sprattus*

Section 1: Abstract

The genus *Sprattus* comprises five species of marine pelagic fishes distributed worldwide, antitropically in temperate waters. Their particular distributions suggest an ancient origin during a cold period of the Earth's history. In this study, we evaluated this hypothesis and the monophyly of the genus *Sprattus* using a phylogenetic approach based on three-mitochondrial DNA fragments and historical biogeography. We found that the genus *Sprattus* is not monophyletic, and that *S. sprattus* is more related to members of the genus *Clupea* than to other *Sprattus* relatives. In addition, the phylogenetic tree showed two well-supported clades that correspond to the species distributed in each hemisphere. Time-calibrated phylogenetic analyses showed an ancient divergence between north and south hemispheres at 17.6 MYBP, followed by an ancient diversification in the northern hemisphere clade (i.e. 16.4 MYBP) and a recent diversification in southern hemisphere clade (i.e. 5.7 MYBP). Historical biogeography analyses showed that the most common ancestor likely inhabits the Atlantic Ocean in Northern Hemisphere. Our results, based on the molecular clock calibration, suggest that *Sprattus* dispersed through the tropics to the Southern Hemisphere when the temperature decreased during the Miocene. Finally, the non-monophyly found in the genus *Sprattus* suggests that the systematics of this genus should be reevaluated to disentangle the history and relationship between *S. sprattus* and the genus *Clupea* and its relatives.

Section 2: Introduction

The five species currently assigned to the genus *Sprattus* (Clupeiforme Order: Clupeinae subfamily) are small marine pelagic fishes that inhabit coastal areas and are well known for their schooling behavior (Whitehead, 1985). These species occur in cooler waters and have an antitropical distribution in Europe, South America and Oceania (Whitehead 1985). *Sprattus sprattus* (Linnaeus, 1758) is the most widely distributed species and it is the only species in the genus found in the northern hemisphere, mainly around the coasts of Europe, and it is most common in the Northeast Atlantic (from North Sea and Baltic south to Morocco; also Mediterranean, Adriatic, Black Sea) (Whitehead, 1985). *Sprattus fuegensis* (Jenyns, 1842) is found in the South American coast, mainly in the Patagonian shelf from both oceans, Pacific and Atlantic including Falkland Island (Aranis et al., 2007; Whitehead, 1985). The other three species are found in Oceania, including: *Sprattus novaehollandiae* (Valenciennes, 1847) in Tasmania, Bass Strait, southeastern Australia between Adelaide Melbourne and Sydney (Whitehead, 1985); *Sprattus antipodum* (Hector, 1872) present in New Zealand coasts (east coast of the North Island, Cook Strait, apparently all coasts of the South Island south to Stewart Island and Foveaux Strait) (Whitehead, 1985; Whitehead et al., 1985); and *S. muelleri* (Klunzinger, 1880) in the North and the South Islands of New Zealand, possibly occurring south to Foveaux Strait and even to Auckland Island, nearly 51°S (Whitehead, 1985; Whitehead et al., 1985).

Molecular analyses on large phylogenies show that *Sprattus* is the sister group of *Clupea* (Cheng and Lu, 2006; Jérôme et al., 2003; Lavoué et al., 2007; Li and Ortí, 2007) which are morphologically distinctive within Clupeinae (Nelson, 1967). It has been estimated that this *Sprattus+Clupea* group diversified during the Messinian (upper Miocene and the Pliocene Epoch) 2.66 – 6.75 million years ago (MYA) (Cheng and Lu, 2006; Jérôme et al., 2003), which is consist with the fossil records of the genus *Clupea* from the Miocene. Moreover, the current *Clupea* species are thought to have radiated in

Pliocene period (3.3 and 3.5 MYA) (Grant, 1986; Wilson et al., 2008), which is when *Sprattus* is thought to have diverged. Recently, Lavoué et al. (2013) evaluated the origin of Clupeoidei based on large fossil-calibrated phylogenies, and found that the *Sprattus* genus was not monophyletic, where *S. sprattus* would be more closely related to *Clupea* spp. than to its relatives in the southern hemisphere (Lavoué et al. 2013). In addition, Bloom and Lovejoy (2014) testing the evolutionary origins of diadromy inferred from a time-calibrated phylogeny for Clupeiformes reinforce the paraphyly of this group.

Currently, no study has addressed the biogeographic origin of the antitropical distribution of *Sprattus*. However, several studies looked at members of the Clupeiformes (e.g. *Engraulis* spp, *Sardinops* spp.), which have the same distribution pattern (Bowen and Grant, 1997; Grant and Bowen, 1998; Grant and Leslie, 1996; Grant et al., 2010, 2005; Okazaki et al., 1996). Moreover, the antitropical biogeographic pattern it is not exclusive to Clupeiformes members (See (Burridge, 2002). For examples, antitropical studies of *Sardinops* showed a recent diversification event, between 0.2 – 2 (MYA) (Bowen and Grant, 1997; Grant and Bowen, 1998; Grant and Leslie, 1996; Okazaki et al., 1996) and *Engraulis* diversified between 5-10 MYA (Grant et al., 2005). When considering marine species that have an antitropical distribution, tropical waters appear to act as a barrier to long-distance dispersal, restricting gene flow between the northern and southern hemispheres (Grant et al., 2010). In fact, studies of thermal tolerance in Clupeiformes (e.g. *Clupea harengus* and *Sardinops sagax*) have showed that warm tropic-like temperatures can potentially restrict migration (Martínez-Porchas et al., 2009; Peck et al., 2012b).

The current antitropical distribution pattern of the genus *Sprattus* suggests that the lower sea temperatures of the tropics during the cooler glacial time might have provided *Sprattus* with the opportunity to disperse to the Southern Hemisphere. This particular distribution suggests an ancient origin during some cold period of the earth. Previous studies of the close phylogenetic relationship of the genera *Sprattus* and *Clupea* suggest

that *S. sprattus* was the earliest diverged species of the taxa, and that the ancestral *Sprattus* species occupied the North Hemisphere (NH). Genetic data and the phylogenetic approach have played a crucial role in studies that assess species boundaries, timing of species divergence, phylogenetic relationships, and the historical biogeographic processes that influence the evolution of to marine species (Palumbi, 1996).

In this study we evaluate the hypotheses on the origin and the monophyly of the genus *Sprattus* using phylogenetic approach based on three-mitochondrial DNA fragment, and historical biogeography to determine the pattern and timing of species diversification.

Section 3: Methods

Sample collection

Tissue samples from five *Sprattus* species were collected (Figure 3.1): *S. fuegensis* (n = 7) from Chilean fjords in Southeast Pacific Ocean, *S. sprattus* (n = 5) from Norweigan fjords in Northeast Atlantic Ocean, *S. muelleri* (n = 4) from Auckland harbor (New Zealand) and *S. antipodum* (n = 1) from Wellington harbor (New Zealand). Tissue or fin-clipped samples were fixed in 95% ethanol and stored at 4°C.

DNA extraction, PCR and DNA sequencing

Total genomic DNA was dissolved in a buffer containing proteinase K and SDS detergent, and then extracted using a standard phenol-chloroform protocol (Sambroock & Russell 2001). DNA was precipitated in 70% ethanol and resuspended in 50 μL of TE buffer. DNA was quantified using a NanoDrop ND-1000 spectrophotometer and diluted to a concentration of 20 ng/μL.

Three pairs primers were designed from complete mitochondrial genomes sequences deposited in Genbank: *S. sprattus* (NC009593), *S. muelleri* (NC16669) and *S.*

antipodum (NC16673). The primers amplified included Cytochrome b (Cytb), NADH dehydrogenase subunit 2 (ND2), and NADH dehydrogenase subunit 3 (ND3). PCRs were conducted in 10 µL volumes containing 67 mM Tris-HCl pH 8.8, 16 mM $(NH_4)_2SO_4$, 3 mM $MgCl_2$, 0.2 mM of each dNTP's, 0.6 µM of each primer, 0.4 mg/mL of Bovine Serum Albumin (BSA), 1.5 U/µL Taq DNA polymerase (Fisher®), and 20 ng/µL of genomic DNA. Thermal cycling was performed in a Eppendorf Mastercycler ep gradient S thermocycler with the following parameters: 95 °C for 300 s, followed by 35 cycles of 94 °C for 30 s, 55 °C for 30 s, 74 °C for 60 s, and a final extension at 74 °C for 600 s. CR was amplified using 2 mM $MgCl_2$ and the thermo cycling parameters: 94 °C for 300 s, followed by 35 cycles of 94 °C for 30 s, 54 °C for 60 s, 74 °C for 90 s, and a final extension at 74 °C for 600 s. PCR product were purified with ExoSAP-IT® following manufacturer's guidelines and sequenced in both directions using an ABI 3730xl Genetic Analyzer (Massey University Genome Sequencing Service). Initial alignment was performed in Geneious® 6.0.5 (Drummond et al. 2013), and the final alignment was adjusted by eye. *Clupea pallasi* (NC009578), *Clupea harengus* (KC193777) and *Ethmidium maculatum* (NC016710) were used as outgroup because they are the species more closely related to the genus *Sprattus*. In addition, sequences of *S. sprattus* (NC009593), *S. muelleri* (NC16669) and *S. antipodum* (NC16673) from GenBank where included in our final data set for further analyses. Genetic diversity and phylogenetic analyses were carried out per gene, while divergence time and historical biogeography analyses were conducted with concatenated alignment datasets.

Phylogenetic analyses and divergence time

Before conducting the phylogenetic analyses we performed Xia's test (Xia et al., 2003) implemented in DAMBE v5 (Xia, 2013) to evaluate whether the sequences we used displayed saturation by substitution, and hence could be useful in the phylogenetic analyses. This is an entropy-based index that estimates a substitution saturation index

(Iss) and compared it to a critical substitution saturation index (Iss.c) via a randomization process with 95% confidence intervals (Xia, 2013; Xia et al., 2003). Xia's test presented low saturation, where the critical index of substitution saturation values (Iss.c = 0.793) was significantly higher than the observed index of substitution saturation values (Iss = 0.155; $p < 0.0001$), therefore, the sequences were suitable for performing phylogenetic analyses.

Phylogenetic analyses were conducted using a Bayesian Markov Chain Monte Carlo approach (BMCMC), which incorporates the uncertainty in the reconstruction of the phylogenetic tree. To BMCMC approach, we applied a general likelihood-based mixture model (hereafter MM) of gene-sequence evolution as described by Pagel and Meade, (2005, 2004). We used this approach because the molecular marker used in this study comes from different species and strains that have diversified rapidly, so their patterns and rates of nucleotide substitution are potentially different. This model, based on the general time-reversible (GTR) model (see Rodríguez et al., 1990), accommodates cases in which different sites in the alignment evolved in qualitatively distinct ways, but does not require prior knowledge of these patterns or partitioning of the data. The MM model, implemented in a BMCMC framework, was used to estimate the posterior probability of the phylogenetic trees to include this information in the comparative method analyses used in this study. The Reversible-Jump Markov Chain Monte Carlo (hereafter RJMCMC) procedure (Pagel and Meade, 2008, 2006) was used with the objective of integrating results of all patterns, and produces an MM that summarizes the sequence evolution, using BayesPhylogenies v1.1 software (http://www.evolution.rdg.ac.uk/BayesPhy.html). This approach enables researchers to explore the variety of possible models and parameters, converging towards the model that best fits the data in the posterior tree sample (Pagel and Meade, 2008). Five independent BMCMC analyses were run using 10,000,000 generations of phylogenetic trees, sampling every 10,000th tree to assure that successive samples were independent. From this sample of trees, first 25 % were removed to avoid including trees sampled before the

convergence of the Markov Chain, and we obtained a final sample of 750 trees. These trees were used to obtain the phylogenetic consensus tree and posterior probabilities.

Approximate divergence times among *Sprattus* species were estimated using a Bayesian Approach included in BEAST software v1.8 (Drummond and Rambaut, 2007; Drummond et al., 2012). To estimate divergence we first estimated the best molecular clock model that fitted our concatenated dataset. We first evaluated the different molecular clock model: a strict clock model (SCM), Log-Normal relaxed clock model (LNCM), Exponential Relaxed Clock model (ECM) and random local clock model (RLCM) (Drummond and Suchard, 2010; Drummond et al., 2006). We ran each model five times using the most complex sequence evolution model, GTR+I+G, with 5,000,000 generations sampling each 10,000 generations. Outputs of each four ran per model were combined in LogCombiner in order to increase effective sample size (ESS), at least >200. The Effective Sample Size (ESS) of a parameter sampled from an MCMC (such as BEAST) is the number of effectively independent draws from the posterior distribution that the Markov chain. To determinate the best molecular clock models we used a Bayes Factor (BF) criterion, which measures the weight of the evidence in the proposed model against another candidate model (Goodman, 1999). The best clock model that fitted our dataset was the Exponential Relaxed Clock model (Table 3.1) (Drummond et al., 2006).

To obtain the posterior distribution of the estimated divergence time, one calibration point was applied as prior to constrain the age in the *Clupea* species node (i.e. *Clupea* sp. fossil; www.paleobiology.org). Prior age distribution of this clade follow a lognormal distribution with a minimal age equal to the minimum age of the geological stage from which the fossil was excavated and the maximum was equal to maximum age of the stage (i.e. 95% credibility interval). The age of this *Clupea* sp. fossil was dated at 15.97 MYA, www.paleobiology.org), thus we set an offset of 15.97 and a mean of 0.1. Posteriorly, we used the model chosen and previous set parameters to ran 10 independent Markov Chain Monte Carlo (MCMC) with a length chain of 100,000,000 generations sampling each 10,000 generation and used priors as well as frequency of each base,

proportion invariant sites and proportions of each transition and transversion, in order to increase the ESS. These individual runs were combined using LogCombiner v1.8 burning 250 trees per each sample. Finally, a maximum clade credibility trees were created in TreeAnnotator v1.8 to visualize a summary tree in FigTree v1.4.

Historical biogeography

Before evaluating the ancestral area of the genus *Sprattus* we estimated the phylogenetic signal of the distributional traits. The phylogenetic signal could be defined as the tendency of more closely related species to have more similar traits than expected by chance from the same tree (Hernández et al., 2013; Münkemüller et al., 2012). Phylogenetic signal was evaluated using the association index (AI, Wang et al., 2001) incorporate in BaTS v1.0 software (http://evolve.zoo.ox.ac.uk/evolve/BaTS.html) (Parker et al., 2008). In our case the traits evaluated were distributional characters: 1) Hemisphere where the species are present, North Hemisphere (NH) versus South Hemisphere (SH), and 2) Ocean that they inhabit, Atlantic Ocean (AO) versus Pacific Ocean (PO). To reconstruct the ancestral states of the geographic distribution we used BayesMultiState methods implemented in BayesTraits software (Pagel & Meade 2009). This method is used to reconstruct how a finite number of discrete states (traits) evolve on phylogenetic trees and it can be applied to traits that adopt two or more discrete states (Pagel et al. 2004). We used tree samples obtained in BMCMC phylogenetic analyses and a traits matrix (hemispheres or ocean). We ran 10,000,000 generations sampled every 10,000 generations to obtain a parameters sample. Posteriorly we burned 25 % to avoid including parameters sampled before the convergence of the Markov Chain, and we obtained a final sample of 750 parameters sampled.

Section 4: Results

Phylogenetic reconstructions per each dataset (i.e. each mtDNA locus independently) showed similar results (Supplementary Figure 3.5). Each *Sprattus* species

incorporated into the Bayesian phylogenetic reconstructions were recovered as monophyletic (Figures 3.2, Figure supplementary 3.5). However, the genus *Sprattus* was not monophyletic: *S. sprattus* is closely related to the genus *Clupea* (Figures 3.2, Supplementary Figure 3.5). Two geographic clades representing each hemisphere were recovered (Figures 3.2, Supplementary Figure 3.5). The Northern hemisphere clade included the species *C. harengus*, *C. pallasi* and *S. sprattus*, and the Southern hemisphere clade included *S. fuegensis*, S, *antipodum* and *S. muelleri*.

Time-calibrated phylogenetic analyses showed an ancient divergence between north and south hemispheres at 17.6 MYBP, followed by an ancient diversification event in the species from the north hemisphere clade (i.e. 16.4 MYBP) and a recent diversification event in the species from the south hemisphere clade (i.e. 5.7 MYBP) (Figure 3.3). Species of *Clupea* genus diverge around 2 million years ago and *S. sprattus* around 1.5 millions years ago. Among the species of the south hemisphere clade, *S. fuegensis* showed an origin at 1.36 millions year ago (Figure 3.3). The most common ancestor of *S. antipodum* and *S. muelleri* diverged around 2.2 millions year ago and each species was originated around 500,000 and 400,000 years ago, respectively (Figure 3.3). Reconstruction of the ancestral area of the genus *Sprattus* indicate that the most recent common ancestor likely inhabit the north hemisphere and it was distributed in the Atlantic Ocean (Figure 3.4).

Section 5: Discussion

Based on mtDNA, the genus *Sprattus* appear to be a paraphyletic group, where *S. sprattus* are more closely related to *Clupea* genera members than its relatives, *S. fuegensis*, *S. antipodum* and *S. muelleri*. This phylogenetic result provides support for two biogeographic clades, one for south hemisphere (i.e. *S. fuegenesis*, *S. antipodum*, *S. muelleri*) and other one for the north hemisphere (*C. harengus*, *C. pallasi* and *S. sprattus*). Time-calibrated phylogenetic tree support a divergence between hemispheres

of 17.6 MYBP, an ancient diversification to species of north hemisphere clade (i.e. 16.4 MYBP) and a recent diversification in south hemisphere clade (i.e. 5.7 MYBP). Finally, the more likely origin to this genus was the north hemisphere in the Atlantic Ocean, which suggest a posterior historical dispersion to the southern hemisphere.

Non-monophyletic *Sprattus* genus

Our results show that the genus *Sprattus* was not monophyletic, challenging the taxonomic status of *Sprattus* species currently extant (Figure 3.2, Supplementary Figure 3.5). Nonetheless, we found two well differentiate geographic clades including species distributed from the North Hemisphere (*Sprattus*+*Clupea*) and South hemisphere (only *Sprattus*) (Figure 3.2). We suggest that the *Sprattus* genus is an endemic monophyletic group of the South Hemisphere, and *S. sprattus* is phylogenetically a *Clupea* species, which are endemic to the northern hemisphere. Actually, using mtDNA we provide support for a paraphyletic nominal *Sprattus* genus. Previous work that include both *Clupea* species and *S. sprattus* have showed that they are closely related; however the study only used three species (Cheng and Lu, 2006; Jérôme et al., 2003; Lavoué et al., 2007; Li and Ortí, 2007). Recently, two studies included to *S. antipodum* and *S. muelleri* in a large phylogeny for the Clupeiformes Order expose this systematic incongruence (Bloom and Lovejoy, 2014; Lavoué et al., 2013). For instance, based on mitogenomic evidence, Lavoué et al. (2013) found that *S. sprattus* is the sister group of *Clupea* species. A recent study, that included nuclear loci and total evidence analyses, showed that the *Ramnogaster* genus is closely related to the New Zealand's sprats, and that the clade *S. sprattus* and the genus *Clupea* are sister groups of the former (Bloom and Lovejoy, 2014). In this latter study, although the authors incorporated four genes (two nuclear, Rag1 and Rag2 and two mitochondrial, 16S and CytB), they do not considered all species of the genus *Sprattus* and also their data set do not consider homologous character for all genes (i.e. *Ramnogaster* sp. only CytB, *Sprattus* species only mitochondrial genes and *Clupea* species all genes). Consequently, phylogenetic reconstruction could be bias only to homologous nucleotide positions in genes shared by all the species analyzed. All the

previously mentioned studies, including our study, provide compelling evidence that *Sprattus* seems to not be a monophyletic group.

The genera *Sprattus* and *Clupea* show similarities that have formerly placed *Sprattus* species into the *Clupea* genus (see details in Whitehead, 1985; Whitehead et al., 1985). Nonetheless, Mathews, (1884) provided a clear distinction between both *Clupea* and *Sprattus* genus, which is the absent of pterotic bullae in *Sprattus* species (see details in Whitehead, 1985, 1964; Whitehead et al., 1985); this feature is now a diagnostic trait. Whitehead et al., (1985) proposed that the absence of pterotic bullae could be the result of a single loss event within the genus *Sprattus*. Other ,morphological traits in addition to the pterotic bullae have been used to differentiate *Sprattus* from *Clupea*. For instance, the number of pelvic finrays is frequently between i6 – i7 in *Sprattus* and i8 in *Clupea* (Whitehead et al., 1985), and the pelvic fin insertion occurs below or in front the dorsal fin in *Sprattus* and behind the dorsal fin in *Clupea* (Whitehead, 1985). In addition, reproductively, the species of the genus *Sprattus* produce pelagic eggs, instead *Clupea* species produce demersal eggs which they attached to the seabed or to the marine vegetation (Whitehead, 1985).

The discrepancies that we found between the species tree and the gene tree could be explained by: incomplete sorting lineage, retain ancestral polymorphism, introgression, misclassification of species or convergence in morphological traits. When incomplete sorting lineage or retain ancestral polymorphism occur, two species should share haplotypes. Therefore, these two explanations could be virtually ruled out in this case because none of the species of this study showed shared haplotypes. In addition, introgression between *S. sprattus* and *Clupea* species should also be ruled out because they have different reproductive strategies, pelagic and demersal eggs respectively (Whitehead, 1985). Moreover, our results suggest that *Sprattus* and *Clupea* diverged deep in the phylogenetic tree, thus hypothesis of introgression would not be supported. In the present study we can not discard misclassification of species or convergence in morphological traits. Although now there is good morphological evidence that separates *S. sprattus* and *Clupea* species (Mathews, 1884; Whitehead, 1985, 1964; Whitehead et

al., 1985), in our study we found evidence that *S. sprattus* is more genetically related to other *Clupea* species than to the rest of the species of *Sprattus.* We also found two clades reciprocally monophyletic (i.e. North and South Hemisphere), which suggest that there are two phylogenetic species. In the past *S. sprattus* was considered a member of the genus *Clupea*, actually, the species was first described as *Clupea sprattus* by Linnaeus (1758). Posteriorly, the genus *Sprattus* was described by Gingensohn (1846) based on the species *S. haleciformis;* however that species was later synonymized to *S. sprattus*. Given that *S. sprattus* was previously described as *C. sprattus*, and that in our studies appears to be more related to other species of the genus Clupea than to the rest of the *Sprattus,* we propose that the genus *Sprattus* should be invalidated. Further analyses should focuss on the genomics of these species (i.e. nuclear gene unlinked) to disentangle the history and relationship between *S. sprattus* and the genus *Clupea* and its relatives, and the systematics of the group should be reevaluated following the International Code of Zoological Nomenclature.

Divergence time and historical biogeography

Based on fossil calibration we found that both antitropical clades diverged at 17.6 YBP (Figure 3.3) at early Miocene. The species of the northern hemisphere clade diverged first in the phylogenetic tree than the species from south hemisphere (16.4 and 5.7 MYBP, respectively; Figure 3.3).

Previous studies incorporating *S. sprattus* and *C. harengus and C. pallasii* dated a splitting divergence event between these two genera around 2.66 – 6.75 MYBP (Cheng and Lu, 2006; Jérôme et al., 2003), which is in disagreement with the ancient divergence found in our study. The molecular dating calibrations done by Cheng and Lu (2006) and Jérôme et al. (2003) were done based on a standard gene substitution rate, contrarily to our divergence time based on fossil data. Different calibrating methods are used to determine the age of a clade (see Hipsley and Müller, 2014), and although any method can determine the true age of a clade, estimations based on substitution rates may arise problems of overestimating divergence times (Hipsley and Müller, 2014; Ho et al., 2011;

Phillips, 2009). Contrarily to previous rate substitution calibration, a current time calibrated phylogeny reconstructions in Clupeiformes (i.e. incorporated more species of the genus *Sprattus*) have calibrated the divergence of the genera around the Miocene and Oligocene (Lavoué et al., 2013), which falls closer to our estimation (Figure 3.3). In addition, studies on other small pelagic marine fishes have also suggest a divergence event around the Miocene (e.g. Grant et al., 2005; Grant and Bowen 1998). For instance, data suggests that *Sardinops* and *Sardina* genera split around 20 MYBP and diversification likely follow close to the Tethys Sea, where dispersal across tropical waters which also resulted in an antitropical distribution of *Sardinops* species (Grant and Bowen, 1998). In the genus *Engraulis* the divergence time has been estimated around 5 – 10 MYBP (Grant and Bowen, 1998; Grant et al., 2005). The most common recent ancestor in the Clupeiformes has currently been estimated in Cretaceous, around 119 MYBP (Lavoué et al., 2013), therefore the taxa has has a long time to diversify into the more than 300 species that now exist, with the variation in morphology and habitat preferences (Nelson, 2006). These studies on small pelagic marine fishes, in addition to our results suggest an ancient origin and recent diversification in the Clupeiformes.

Lavoué et al., (2013) indicated that the Clupeiformes' ancestor could have been originated and diversified from a tropical marine precursor distributed in Indo-West Pacific region. From this tropical region some lineages dispersed to both, the north and south hemispheres, where four lineages would have originated, including *Sprattus* and *Clupea* members which likely dispersed southward (Lavoué et al.2013). However, the fossil record of the genus *Clupea* of the Northern Hemisphere does not support the southward pathway described by Lavoué et al. (2013). To date there are no fossil records of the genus *Sprattus*, but there are six extinct species described for the genus *Clupea* (i.e. *C. testis*, *C. longimana*, *C. sardinites*, *C. voinovi* and two *Clupea* sp.), all of them found in the Northern Hemisphere (i.e. Romania, Germany, Czech Republic and North Korea, see www.paleobase.org). Then, the fossil record supports the hypothesis of a Northern Hemisphere origin of the *Sprattus – Clupea* clade.

In our study, Bayesian analyses that evaluated the biogeographic historic origin of the surveyed species suggest that the most recent common ancestor likely originated in the North hemisphere (Figure 3.4A) in the Atlantic Ocean (Figure 3.4B). Previous historical biogeographic analyses have pointed out that the ancestor of the Order Clupeiformes may have originated in the Indian Ocean around the Cretaceous (Lavoué et al., 2013). This ancestor would have been adapted to warm temperatures (i.e. > 25ºC, Lavoué et al., 2013) and to the marine environment (Bloom and Lovejoy, 2014; Lavoué et al., 2013). Extant members of *Sprattus* and *Clupea* genera are distributed antitropically in temperate waters (Lavoué et al., 2013; Whitehead et al., 1985) and mainly inhabit marine environments (Bloom and Lovejoy, 2014), although they can also inhabit areas with highly variable oceanographic features, like fjord (e.g. Aranis et al., 2007; Glover et al., 2011).Then, *Sprattus* and *Clupea* species have departed from the ancestral habitat and they have adapted to colder waters, while simultaneously expanding their tolerance to salinity fluctuations which allowed them to also colonize fjords habitat.

Overall, researches that have surveyed marine fishes with antitropical distribution have usually explained this geographic pattern under the point of view of dispersalist and vicariant (Burridge and White, 2000; Burridge, 2002; Grant and Bowen, 1998; Grant et al., 2005; Halvorsen et al., 2012; Le Port et al., 2013; Stepien and Rosenblatt, 1996). Dispersalist examples invoke tropical submergence and migrations across tropical areas as a mechanism to explain antitropical distribution of species (Burridge, 2002; Lindberg, 1991). The tropical submergence, refers to the possibility that marine organisms distributed in cool or temperate areas are able to disperse across the tropical region through deeper colder tropical waters (Burridge, 2002; Lindberg, 1991; Sverdrup et al., 1942) . These explanation has been frequently invoke as the main mechanisms that would explain the antitropical distribution of pelagic fishes that occurred in the Pleistocene (e.g. Burridge and White, 2000; Burridge, 2002; Grant and Bowen, 1998; Grant et al., 2005; Halvorsen et al., 2012; Le Port et al., 2013; Liu et al., 2006; Stepien and Rosenblatt,

1996). Vicariant mechanisms like plate tectonic, relictual distribution and equatorial isolation by climatic change or biological interactions have been advocated by some researchers (Stepien and Rosenblatt, 1996). However, mechanisms associated to plate tectonic are not supported in our data because splitting and divergence time among *Sprattus* and *Clupea* nominal species would have occurred during the Miocene (Figure 3.3). The present continental configuration closely resembles the configuration of the continents during the Miocene, Other studies on marine pelagic fishes have supported dispersalist mechanism to explain its antiropical distribution (e.g. Burridge and White, 2000; Grant and Bowen, 1998; Grant et al., 2005; Halvorsen et al., 2012; Le Port et al., 2013; Liu et al., 2006) excluding vicariant explanations as well (e.g. Grant et al., 2005). Our outcomes do not support a vicariant explanation for *Sprattus* and *Clupea*, but we proposed that the equatorial warmer water area (i.e. temperature) has acted as a key barrier to dispersion across hemispheres, and that it has only been trespassed when the waters at the tropics lowered their temperatures during the Miocene.

Section 6: Tables

Table 3.1: Bayes factor comparison among different clock models incorporated in BEAST software.

SCM: strict clock model, ECM: exponential relaxed clock model, LNCM: lognormal relaxed clock model, LCM: random local clock model

	ln P(model | data)	S.E.	SCM	ECM	LNCM	LCM
SCM	-7590,619	+/- 0,174	-	0,023	0,03	0,036
ECM	-7586,832	+/- 0,224	**44,119**	**-**	**1,305**	**1,588**
LNCM	-7587,099	+/- 0,188	33,8	0,766	-	1,216
LCM	-7587,294	+/- 0,186	27,79	0,63	0,822	-

Section 7: Figures

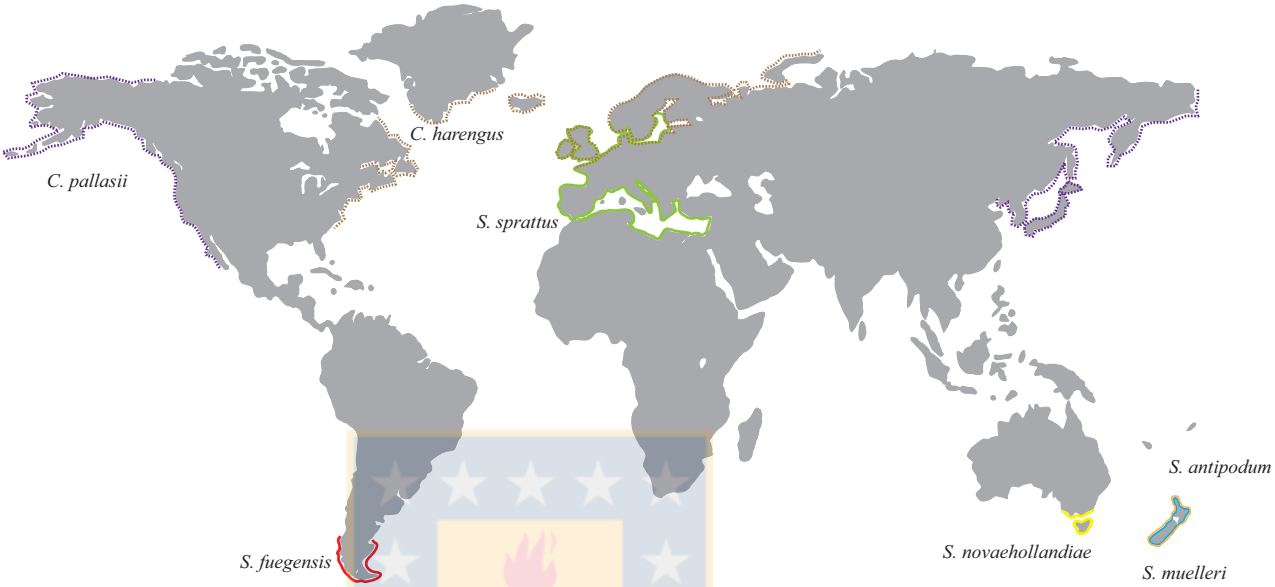


Figure 3.1: Distributional map of extant *Sprattus* and *Clupea* species.

Red solid line *S. fuegensis*, yellow solid line *S. novaehollandiae*, orange solid line *S. muelleri*, light blue solid line *S. antipodum*, green solid line *S. sprattus*, brown dashed line *C. harengus* and violet dashed line *C. pallasii*.

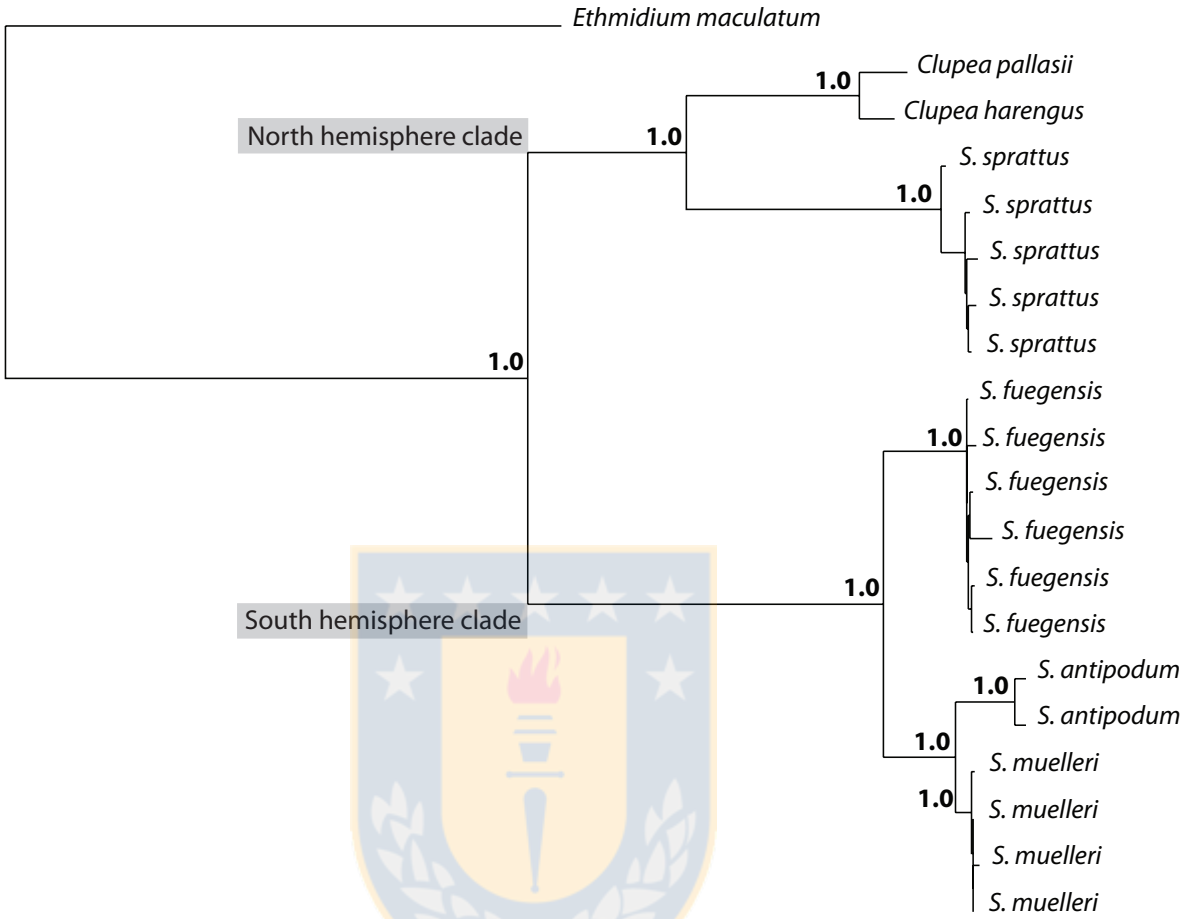


Figure 3.2: Bayesian consensus tree concatenating three-mitochondrial gene from 750 more likely trees.

Branch lengths are proportional to the number of substitutions per nucleotide position. Numbers at nodes are posterior probabilities from Bayesian analyses. Gray rectangles indicate current hemisphere distribution.

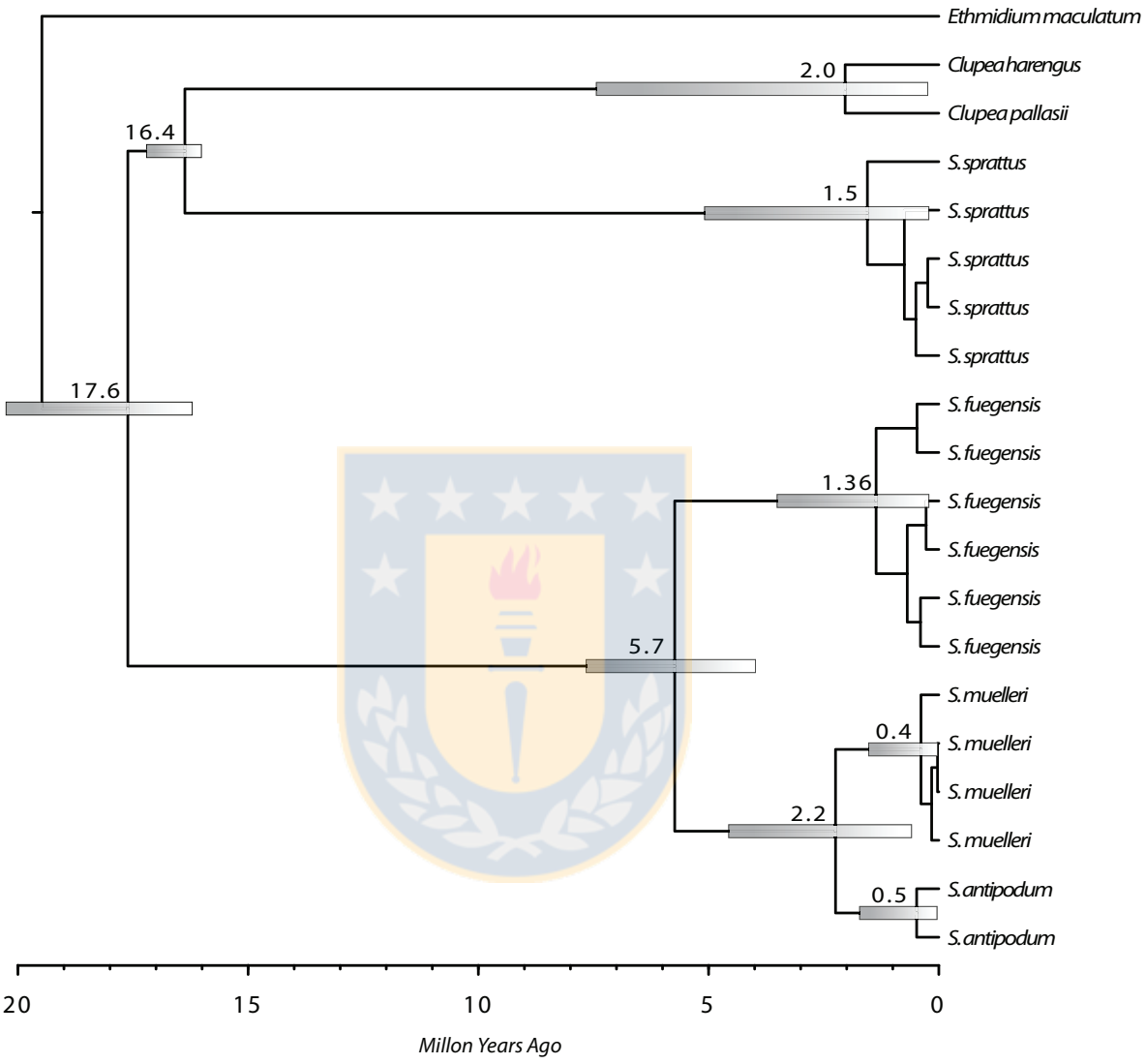


Figure 3.3: Time-calibrated phylogenetic tree based on a Bayesian relaxed clock analyses.

Numbers at nodes are divergence time since the root to each species. Horizontal light gray gradient bars indicate the 95% HPD of divergence times, and the scale axis shows divergence times as millions of years ago (MYA).

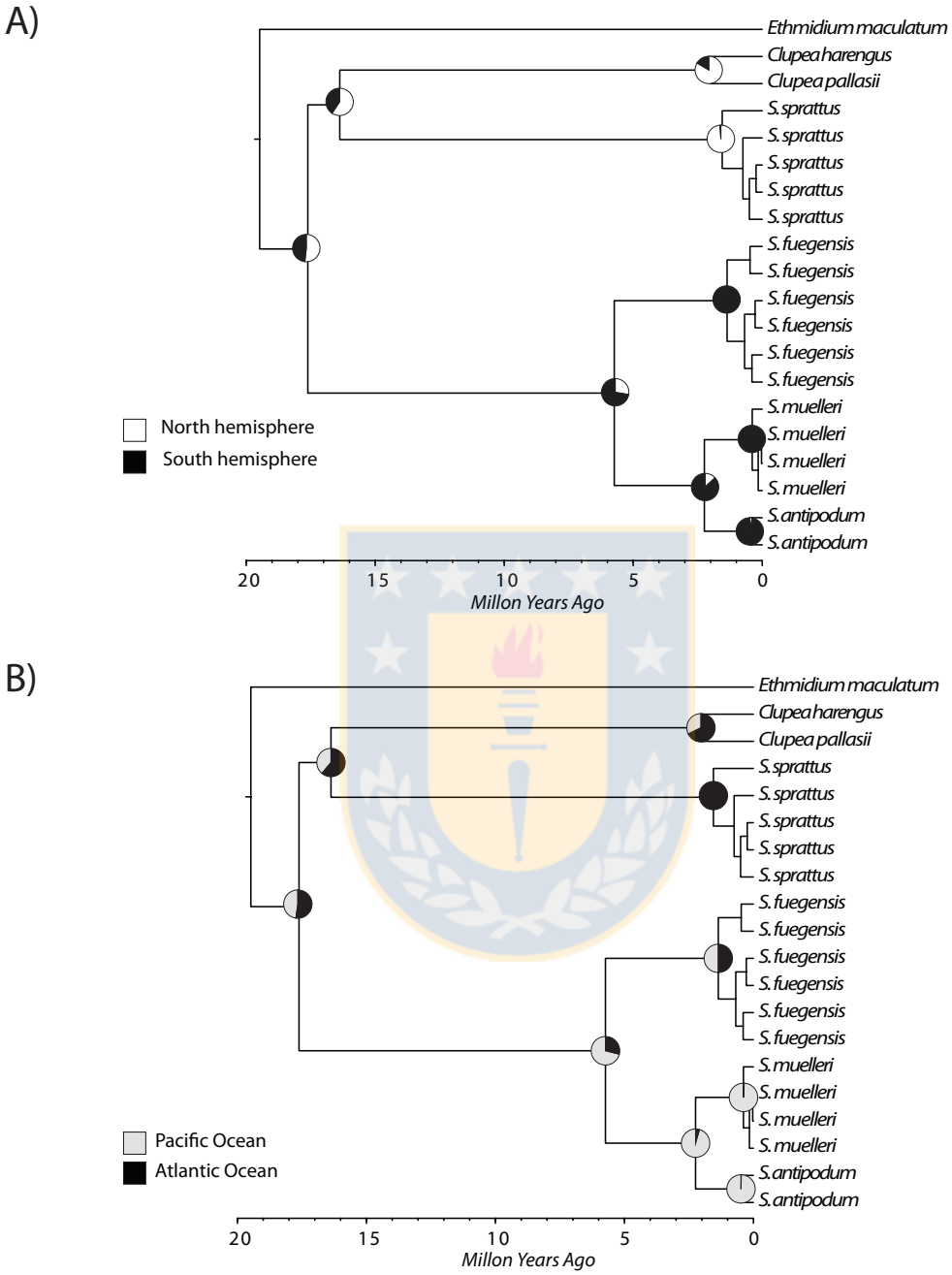


Figure 3.4: Biogeographical origin in *Sprattus* species.

A) Reconstruction of the ancestral state of the south hemisphere origin (in black) and north hemisphere origin (in white). B) Reconstruction of the ancestral state of the Atlantic Ocean origin (in black) and Pacific Ocean origin (in gray). Analyses based on both the topology and the branch lengths of the Bayesian phylogenetic trees.

Section 8: Supplementary figures

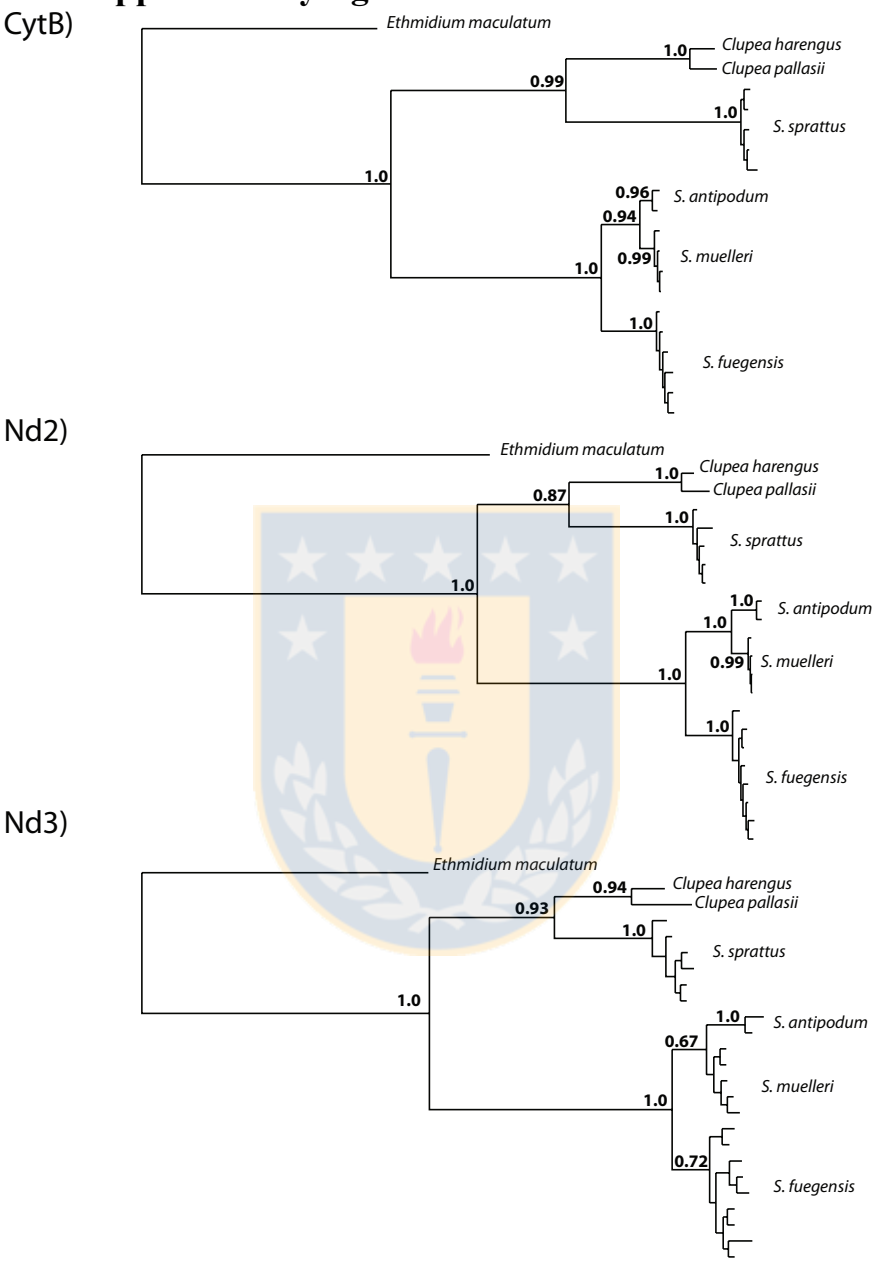


Figure 3.5: Bayesian consensus tree to each mitochondrial gene fragment

Gene fragments (i.e. CytB, Nd2 and Nd3). Branch lengths are proportional to the number of substitutions per nucleotide position. Numbers at nodes are posterior probabilities from Bayesian analyses.

Chapter 4 – Assessing the importance of oceanographic factors in diversification of Clupeiformes fishes: Court Jester Hypotheses

Section 1: Abstract

Disentangling the factors that determine why some clades are more diverse than others is a major focus of evolutionary biology. Different hypotheses have been proposed, where the importance of environmental scenario is contrasting. For example, the Red Queen Hypothesis, which empathize the importance of intrinsic biological factors versus the Court Jester Hypothesis that empathize the importance of external environmental factors. Here we evaluate these hypotheses assessing the relationship of oceanographic environmental factors on the diversification rates of a richness group of pelagic marine fishes, the Order Clupeiformes. We used a methodological framework of molecular phylogenetic, phylogenetic comparative methods, and birth-death models to test whether oceanographic environmental factors are associated to the processes of species diversification. We found that four of five the environmental factors tested showed a phylogenetic signal and that temperature, salinity, oxygen and chlorophyll were associated under a negative hump model (i.e. mean values showed a high diversification in extreme values) to the evolutionary diversification rates of the group. Finally, we could conclude that the diversification of Clupeiformes support the Court Jester Hypothesis, an it should promoted for colonization of extreme environments or that environment change through the time.

Section 2: Introduction

A major current focus of evolutionary biologist and paleontologist is to disentangling the factors that determine the structure of phylogenetic trees (i.e.

symmetrical or asymmetrical), in order to answer why some clades are more diverse than others, and to understand the underlying mechanism associated to biological diversity (Losos et al 1995). Richness diversity pattern for a lineage is the result of the diversification rate, that is the speciation rate (λ) minus extinction rate (μ) (Paradis 1997). This diversification rate could be the consequence of different specific attributes that could be influencing speciation or extinction rates (Barraclough et al. 1998, Cardillo et al. 1999). For instances, biological factors like multicellularity, body size, sexual selection, and breeding system have shown positive correlation with the net rate of cladogenesis (Barraclough et al. 1995, Avaria-Llautureo et al. 2012, Schirrmeister et al. 2013, Mooers & Mϕller 1996). Although, these and other studies evaluated how biological attributes are correlated to diversification rate, only few studies have paid attention to the effect that environmental variables have on diversification rate (e.g. Steeman et al. 2009, Suto et al. 2012, Williams et al. 2013). These two conceptual frameworks, speciation by biotic factors and by abiotic factors, invoke the Red Queen hypothesis (Van Valen 1973) and Court Jester Hypothesis (Barnoski 2001), respectively. The Red Queen hypothesis (RQH) describes that species are continuously evolving by biotic competition between species, which implies that they will not necessarily become better adapted to a given environment (Van Valen 1973, Vendetti et al. 2010); even those species that are currently well adapted to a given environment, are still undergoing continuous adaptation because the environment is never constant (Vendetti et al. 2010, Benton 2010). On the other hand, the Court Jester hypothesis (CJH) states that physical-environmental perturbations, such as climate change, are more important in biological diversification than biotic factors in opposition to what is predicted by the Red Queen Hypothesis (Barnoski 2001). For instance, Barnoski (2001) pointed out that an increase of speciation rate might be the result of "random perturbations to the physical environment such as climate change, tectonic events, or even bolide impacts that change the ground rules for the biota". Therefore, environmental variation plays a key role in intraspecific processes (Kozak *et al.* 2008, Canales-Aguirre et al. in prep) and has been defined like a driver of evolutionary change (Erwin 2009) in a macroevolutive scenario.

Currently, the increased in molecular data in worldwide repositories (i.e. GenBank), marine environmental information for global-scale (i.e. Bio-ORACLE, Tyberghein et al. (2012); MARSPECT, Sbrocco & Barber (2013)), as well as a high theoretical knowledge of diversification rates (Yule 1924, Nee et al. 1994, Paradis 2003, Rabosky et al. 2007, Morlon et al. 2011, Eastman et al. 2011, Rabosky 2013, Fitzjohn 2010, Silvestro et al. 2011, Maddison et al. 2007) plus the availability of analytical tools, could help elucidate the effects of abiotic factors that drive speciation. Putting together this diversification rate framework with approaches based on the comparative phylogenetic method (see Pagel 1997, 1999a, 2002; Blomberg, Garland & Ives 2003, Hernández et al. 2013), and specifically evaluating the phylogenetic signal of a trait, would allow us to study the effects of macroevolutionary processes on observed macroecological patterns (Hernández et al. 2013). Finally, merging these two analytical approaches could allow us to recognize causes that would explain the difference in species richness among clades of species.

In this study, we examine the relationship of oceanographic environmental factors on the diversification rates across the Order Clupeiformes, a group of pelagic marine fishes (Actinopterygii) which includes mostly species of economic and ecological importance. Clupeiformes is a rich and taxonomically diverse Order that includes herrings, sardines, pilchards, sprat, anchovies, and wolf herring. The order includes five families, 84 genera, and 364 species (Nelson 2006). Given that this group is composed by pelagic fishes, which are mostly associated to shallow superficial waters (e.g. *Clupea*, *Engraulis*, *Sardinops*, etc.), the oceanographic characteristics (i.e. superficial surface temperature) and chemistry (i.e. Salinity, Oxigen, Nutrients, etc) of these waters mass should be driving their evolutionary changes, which suggest that the CJH explain the distribution of extant taxa and Clupeiformes diversity. We evaluate the CJH using the methodological framework of molecular phylogenetic, phylogenetic comparative

methods and birth-death models to test whether oceanographic environmental factors are associated to the processes of species diversification in this fishes.

Section 3: Methods

Marine environmental and DNA database collection

We compiled a contemporary marine database of environmental factors that included temperature, salinity, oxygen, chlorophyll, and nitrate from MARSPEC database (Sbrocco & Barber 2013) and from Bio-Oracle databases (Tyberghein et al. 2012). From these databases we obtain available values of maximum, minimum, mean, range and variance of all environmental factors. We used sequences of the Cytochrome-b mitochondrial gene (hereafter Cyt b, respectively) from GenBank (GenBank 2014) to posterior phylogenetic analyses, because sequences of this marker are available for a large number of Clupeiformes species (Supplementary table 4.2). Two far relatives fishes were selected as outgroups (i.e. *Ictularus punctatus* and *Danio rerio*), based on the phylogeny Bloom & Jovejoy (2014). Finally, we obtain molecular data with the associated environmental data for approximately 30% of extant Clupeiformes species, representing 55% of the actual genus richness.

BMCMC molecular phylogeny and estimates of divergence times

Given that our hypothesis-testing framework was based on phylogenetic relationship over time, we reconstructed the molecular phylogenies of Order Clupeiformes. We simultaneously estimated phylogenetic relationships with aligned sequences, branch lengths, and divergence times for the group using BEAST 1.7.5 software (Drummond and Rambaut, 2007). This analysis was conducted using a BMCMC framework to estimate the posterior probability of phylogenetic trees. As prior information we used a GTR+Γ+I model of sequence evolution, the Birth-Dead process of speciation, and we used an uncorrelated exponential relaxed clock as was suggest by Lavoué et al. (2013). One independent MCMC chain was ran for 10,000,000 generations

(discarding the first phylogenetic trees, before the posterior probabilities distribution of the selected diversification model converged), with parameters sampled every 10,000 steps. Examination of MCMC sample was performed using TRACER v. 1.4 software (Rambaut and Drummond, 2007) to evaluate that the chain was adequately sampling the probability distribution; and that effective sample size for all parameters of interest were greater than 200. Finally, to obtain the consensus ultrametric tree from the Bayesian sample, we used the TreeAnnotator v1.7.5 using the maximum clade credibility tree, and we kept target heights to discard negative branch lengths. To obtain the posterior distribution of the estimated divergence time, three fossil calibration points were applied as prior to constrain the age in nodes of the genera *Alosa*, *Sardinella*, and *Sardinops* (i.e. www.paleobiology.org). Prior age distribution of this clade followed an Exponential distribution with a minimal age equal to the minimum age of the geological stage from which the fossil was excavated, and the maximum was equal to the maximum age of the stage (i.e. 95% credibility interval). The age of these fossils were: *Alosa sculptata* (28.4MYBP), *Sardinella* sp. (13.65 MYBP), and *Sardinops* sp. (35.7 MYBP), www.paleobiology.org). Finally we use a constraint of 180 MYBP to root the Order Clupeiformes based on Lavoué et al. (2013).

Inferring patterns of diversification rates over time and between clades.

With Bayesian statistical inference, we explored a variety of time-varying models that could account for diversification rate heterogeneity in time and among clades with BAMM software (Bayesian Analysis of Macroevolutionary Mixtures; Rabosky, 2014). BAMM uses rjMCMC to automatically explore a vast universe of candidate models of lineage splitting, finding locations for shifts to new time-varying or diversity-dependent evolutionary rate regimes that are maximally supported by the data (Rabosky 2014), with no a priori specification as to where these shifts in dynamics might have occurred (Rabosky 2014). These shifts can also be abrupt in time, and a mixture of models (i.e. many distinct combinations of evolutionary shift regimes) might have roughly equal posterior probabilities to explain the data (e.g. Rabosky et al. 2013, Rabosky et al. 2014).

Testing the phylogenetic signal in present-day environmental variables.

We evaluated the extent to which the phylogeny correctly predicts patterns of similarity in oceanographic environmental variables (i.e. phylogenetic signal, see Table 4.1) using the phylogeny scaling parameter, λ (Pagel 1999). In this approach λ reveals whether the phylogeny fits the patterns of covariance among species for a given trait (Pagel 1999). This parameter evaluates whether one of the key assumptions underlying the use of the comparative method (i.e. that species are not independent), fits the data for a given phylogeny and trait, assessing the strength of the phylogenetic signal (see Hernández et al. 2013). Values close to zero indicate that there is no concordance between phylogeny and the trait values of species (phylogenetic independence). If traits are evolving as expected, given the tree topology and branch lengths, λ takes the value of 1. Intermediate values of λ, between 0 and 1, indicate different degrees of a phylogenetic signal (Pagel 1999, Blomberg et al. 2003). We inferred the maximum likelihood values of lambda for each oceanographic variable based on the ultrametric tree (obtained by BEAST) in BayesTrait v2.0 software (Pagel, 1999, 1997). To test the significance of the phylogenetic signal we compared the estimated value with a model of lambda forced to 0 by BF. Only those variables that had significant phylogenetic signal were used to evaluate its past effect on diversification patterns.

Testing the effect of past environmental variables on diversification patterns

We tested the hypothesis that differences in lineage diversification rates were influenced by past climate using QuaSSE algorithm (FitzJohn, 2010), that takes a time-phylogeny and a set of continuous trait measurements for the tips. In order to fit a series of birth–death models in which the lineage diversification rates are independent of trait evolution, or vary along branches as a function of a continuous trait that evolves according to a diffusion process, with or without an evolutionary tendency (in this case the environmental variables with phylogenetic signal (Pagel 1999; Blomberg et al. 2003). In these models, rates varied as linear, sigmoidal, or hump-shaped function respect to past

variables, with and without a general tendency in the evolution of environmental variables that lineages occupy over time. Hence, we compared seven models of diversification rates: 1) a constant model of speciation (non environmental effect); 2) speciation that varies as a linear function of environmental variables, evolving by a diffusion process; 3) speciation that varies as a sigmoidal function of environmental variables, evolving by a diffusion process; 4) speciation that varies as a hump function of environmental variables, evolving by a diffusion process; 5) speciation that varies as a linear function of environmental variables, evolving by a diffusion process with a directional trend; 6) speciation that varies as a sigmoidal function of environmental variables, evolving by a diffusion process with a directional trend; and 7) speciation that varies as a hump function of environmental variables evolving by a diffusion process with a directional trend (see FitzJohn 2010; FitzJohn 2012). In a linear model, the diversification rates varies proportional to environmental variables, in a sigmoidal model species that inhabited low values of environmental variables had a low diversification rate compared to species of high values of environmental variables, and in a hump model, species that occupied middle values of environmental variables had the highest speciation rate. These models have the following parameters: the speciation and extinction rate parameters (λ, μ); the diffusion parameter (σ^2), which is the expected squared rate of change and captures the stochastic elements of trait evolution; and the directional trend ‘‘drift’’ parameter (θ), which captures the deterministic or directional component of trait evolution, this is the expected rate of change of the character over time and may be due to selection or any other within-lineage process that has a directional tendency (FitzJohn, 2010). These models were implemented in QuaSSE algoritm, and the analyses were done using the Diversitree package of R software (FitzJohn, 2012).

Section 4: Results

The rjMCMC algorithm implemented in BAMM does not showed non-shifts to new time varying or diversity-dependent evolutionary rate regimes changes. However,

indicated that the diversification of the Order Clupeiformes was best described by an increment of the general diversification rate since 140 to 115 mya ago, given for the speciation rate because the extinction rate was constant (Figure 4.1). In general, Clupeiformes showed a constant diversification rate through the time since 115 to the present, nevertheless this constant diversification rate toward more recent groups is result of the heterogeneity of diversification rate among clades. For instance, the clade of Pristigasteridae Family has species that show an increase and decrease diversification rate, as well as the Coilinae subfamily (Figure 4.2). This latter it is the result of diversification rate with intermediate values and constant. More recent clades show different patterns (increase or decrease diversification rate), which result in maintains constant diversification rates through the time (Figure 4.2).

All variables showed significant phylogenetic signal but whit intermediate values (0.3 – 0.8) (Table 4.1), indicating that the environment that currently inhabits species of Order Clupeiformes have been determined by the environment in which the ancestral species inhabited. The evolution of temperature, salinity, oxygen and chlorophyll were associated to a diversification patterns with a negative trend and a hump function (Table 4.1). Negative trend and hump function means that the diversification rate has accelerated thought of the history of Clupeiformes at which time the values of the environmental variables are extreme (e.g. high or low temperature, salinity, etc.).

Section 5: Discussion

Our results support the hypothesis that ancient marine environmental variables played a key role on the diversification rate of Clupeiformes species, whish support the Court Jester hypothesis. The general diversification rate increased from the 140 mya to 115 mya in the lower Cretaceous until reaching a constant rate (i.e. 115 mya to present; figure 1); however, there are some variations in recent clades associated to sub-families and families depending of the environmental habitats of those clades. For instance, the

clade of Pristigasteridae Family has species that show an increase and decrease diversification rate, as well as the Coilinae subfamily (Figure 2). The emergence of new species in the Cluperiformes was related to increases in speciation rates that were associated to extreme changes in the environment variables: so temperature, salinity, oxygen and chlorophyll promote diversification when they reach extreme values in a particular habitat (Barnosky 2001). To intermediate values of temperature, salinity oxygen and chlorophyll the speciation rate decreased, conversely the diversification rate increased in Clupeiformes when the value of the environmental variable as been extreme or it deviates from the mean. That means low or high temperatures, low or high salinity, and so on.

The ancestor of Clupeiformes originated in the lower Cretaceous, in tropical marine regions, and later dispersed and colonized new regions with different environmental conditions to which they adapt generating the present-day distribution of species richness (Lavoué et al. 2013). If the ancestor of Clupeiformes inhabited warmer waters, the ancestors might have had high temperature tolerance in order to colonize temperate regions with cooler waters. Our results support that the colonization of extreme environments promote diversification rate (Table 4.1). In the marine realm, temperature is a key environmental variable that have been shown to be directly related to the number of species in a given area, to act as a barrier to dispersal, to affect the developmental rates of eggs and larvae, and to influence recruitment, among others. For instance, several species of Clupeiformes have an antitropical distribution in temperate or cooler waters, where the current warm tropical water blocked free dispersal among hemispheres (e.g. Grant et al. 2005, Grant & Bowen 1998, Grant et al. 2010, Okazaki et al. 1996, Bowen & Grant 1997). During “El Niño” events, superficial warm water moves towards temperate areas, and species such as *Engraulis ringens* and *Sardinops sagax* change their bathymetrical distribution, seeking cooler waters or change their distributional range (Ñiquen and Bouchon 2004, Freon et al. 2005). In an ecophysiological context, temperature is a key factor for small pelagic fishes (Meskendal et al 2010, Peck et al.

2012, Pörtner & Peck 2010). On the other hand, in a reproductive context, several species have an optimum range of temperature in which they reproduce, or in which earlier stages have a higher survival rate (Peck et al. 2012a,b; Martinez-Porchas et al. 2009). For this latter, and based on our result we can suggest that this optimum temperature do not promote the diversification, because this optimum only promotes an ecological survivor to maintain the species along the time.

The most recent common ancestor of Clupeiformes was a marine species (Lavoué et al. 2013); however, several current species of this group are euryhalines or inhabit freshwater. In addition, in this group there are many species that shown diadromy, a life – history behavior where individuals migrate between ocean and freshwater (Bloom and Lovejoy, 2014). In Clupeiformes, salinity has been suggested as having its largest ecological impact during spawning (Petereit et al., 2009), on influencing changes in abundance (Ojaveer and Kalejs 2010), and also on acting as a key reproductive barrier (Limborg et al., 2009). Shultz & McCormick (2013) pointed out that ray-finned fishes can vary up to tenfolds in their range of tolerance to salinity levels, suggesting that euryaline species are a remarkable source of evolutionary diversity. In addition these authors suggested that posterior to the Cretaceous-Palaeogene extinction (around 65 My) there was an increment in the number of families that tended to be halohabitat-euryhaline.

All variables associated to chlorophyll showed phylogenetic signal and were associated to evolutionary diversification rates in Clupeiformes (Table 4.1). Chlorophyll is a proxy used to estimate primary productivity (Bot & Colijin 1996) or marine phytoplankton. Most of the Clupeiforme species eat close to the base of food web (i.e. plankton), thus a large portion of this species are associated to regions rich in nutrients where phytoplankton blooms seasonally or continuously through time (i.e. upwelling areas) (Whitehead 1985). Recently, Suto et al. (2012) showed that changes in the upwelling conditions drive the evolution of marine organisms by comparing marine phytoplankton species curves with chronological ranges of marine mammals. However,

although the authors did not incorporate marine fishes into their study, they suggested that diversification of phytoplankton (i.e. Chaetoceros) and its increase in abundances had a major impact in the evolution of marine mammals due to their effect through their food source, smaller herbivorous and carnivorous organisms (Suto et al. 2012). The latter and our results strongly suggest that primary productivity (i.e. as a proxy of chlorophyll concentration) plays a role in the diversification of Clupeiformes species.

Although, these environmental oceanographic variables do not act independently on the species and might have complex dynamics of interaction, our results show that together they played a major role in promoting speciation in the marine fishes Clupeiformes, which support the Court Jester Hypothesis. Thus, the use of the comparative phylogenetic method and the diversification rates analyses expanded our understanding of the effect of several environmental factors on the historical demography of marine organisms. As proposed by Benton (2009): “Comparative phylogenetic methods will illuminate questions about clade dynamics, species richness, and the origin of novelties”.

Section 6: Tables

Table 4.1: Environmental variables used to test phylogenetic signal (λ) and to test an association with diversification rate.

		(λ)	Diversification	Model	Tendency
Temperature	mean	0,821	-	-	-
	range	0,679	yes	Hump	Negative
	Var	0,619	-	-	-
Salinity	mean	0,304	-	-	-
	range	0,661	yes	Hump	Negative
	Var	0,593	yes	Hump	Negative
Oxygen	mean	0,721	yes	Hump	Negative
NO3	mean	0,387	-	-	-
Chlorophyll	mean	0,659	yes	Hump	Negative
	Max	0,572	yes	Hump	Negative
	Min	0,635	yes	Hump	Negative
	Ran	0,465	yes	Hump	Negative

Section 7: Figures

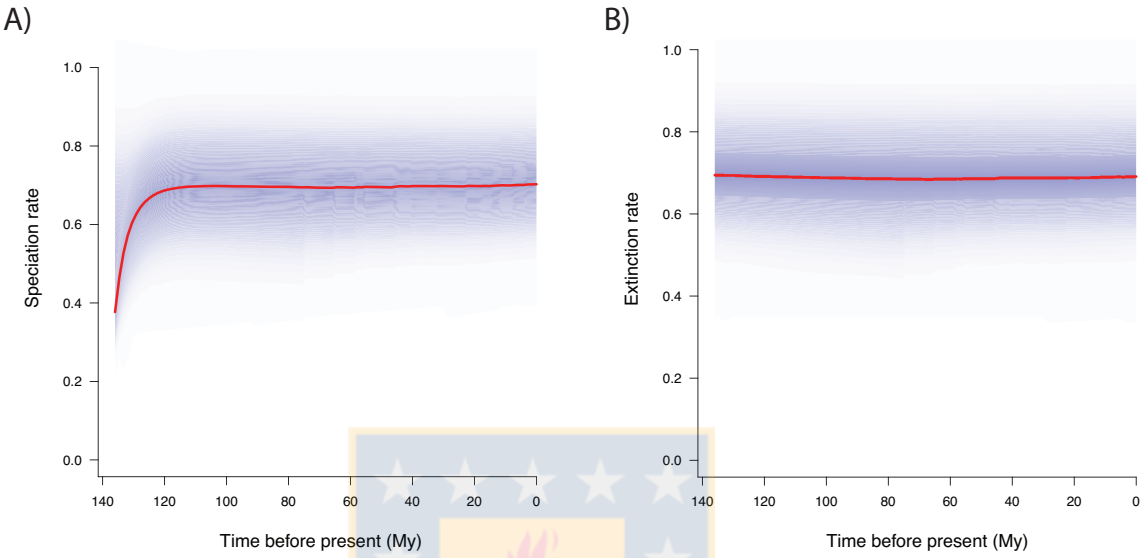


Figure 4.1: Speciation and extinction rate through the time.

A) Speciation rate and B) extinction rate of Clupeiformes fishes.

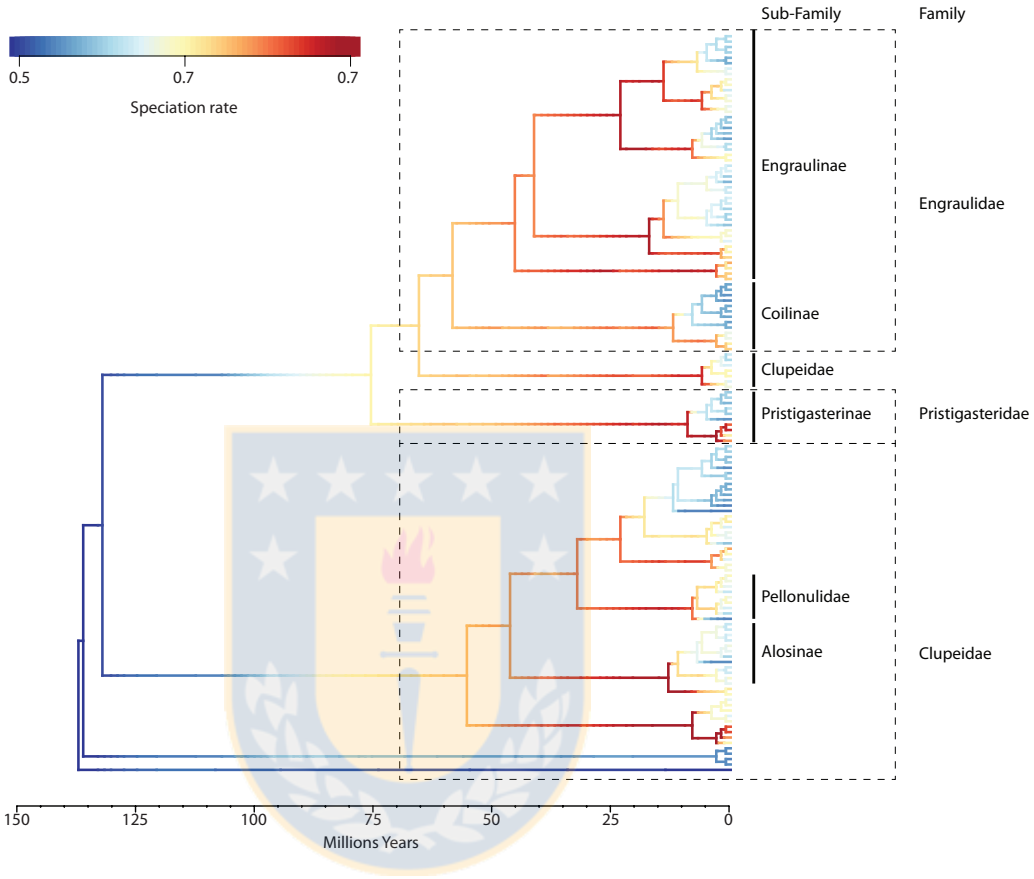


Figure 4.2: Time-calibrate phylogeny of Clupeiformes showing speciation rates.

Cool colors means slow speciation and, warm colors means fast speciation along each branch of the Clupeiformes phylogeny. Each unique color section of a branch represents the mean of the marginal posterior density of speciation rates on a localized segment of a phylogenetic tree. Horizontal black line showed clades that define each subfamily. Dashed box showed clades that define each Family found in Bayesian phylogenetic tree.

Section 8: Supplementary tables

Table 4.2: List of species used in this study and access number of Cyt B sequences data set obtained from GenBank

Family	Sub-Family	Genus	Species	GenBank acc.
Clupeidae	Alosinae	*Alosa*	*aestivalis*	EU552615
		Alosa	*alabamae*	KJ158091
		Alosa	*alosa*	NC_009575
		Alosa	*chrysochloris*	KJ158092
		Alosa	*fallax*	EU552574
		Alosa	*mediocris*	KJ158093
		Alosa	*pseudoharengus*	AP009132
		Alosa	*sapidissima*	EU552616
		Brevoortia	*patronus*	EU552618
		Brevoortia	*smithi*	KJ158094
		Brevoortia	*tyrannus*	EU552614
		Ethmalosa	*fimbriata*	NC_009582
		Ethmidium	*maculatum*	AP011602
		Hilsa	*kelee*	AP011613
		Tenualosa	*ilisha*	EU552622
	Dorosomatinae	*Anodontostoma*	*chacunda*	AP011614
		Dorosoma	*cepedianum*	EU552586
		Dorosoma	*petenense*	EU552581
		Konosirus	*punctatus*	AB548682
		Nematalosa	*erebi*	EU552592
		Nematalosa	*japonica*	NC_009586
	Dussumieriinae	*Etrumeus*	*teres*	EU552621
		Etrumeus	*whiteheadi*	EU552567
		Jenkinsia	*lamprotaenia*	EU552613
		Spratelloides	*delicatulus*	NC_009588
		Spratelloides	*gracilis*	NC_009589
	Pellonulinae	*Pellonula*	*leonensis*	EU552624
		Pellonula	*vorax*	EU552628
		Gilchristella	*aestuaria*	EU552578
		Sauvagella	*madagascariensis*	EU552610
		Potamalosa	*richmondia*	AP011594
		Hyperlophus	*vittatus*	EU552587
	Clupeinae	*Clupea*	*harengus*	EU552606
		Clupea	*pallasii*	EU552599
		Clupeonella	*cultriventris*	NC_015109
		Escualosa	*thoracata*	AP011601
		Harengula	*jaguana*	EU552617
		Lile	*stolifera*	KJ158080
		Opisthonema	*libertate*	KJ158081
		Opisthonema	*oglinum*	EU552620
		Rhinosardinia	*amazonica*	EU552550
		Sardina	*pilchardus*	AF472582
		Sardinella	*aurita*	EU552619

Family	Sub-Family	Genus	Species	GenBank acc.
Clupeidae	Clupeinae	*Sardinella*	*lemuru*	KJ158078
		Sardinella	*maderensis*	NC_009587
		Sardinops	*sagax*	EU552565
		Sprattus	*antipodum*	AP011608
		Sprattus	*muelleri*	AP011607
		Sprattus	*sprattus*	AP009234
Chirocentridae		*Chirocentrus*	*dorab*	NC_006913
Engraulidae	Coiliinae	*Coilia*	*lindmani*	NC_014271
		Coilia	*mystus*	EU694407
		Setipinna	*tenuifilis*	JQ012398
		Setipinna	*taty*	JQ012365
		Thryssa	*baelama*	NC_014264
		Thryssa	*dussumieri*	JQ012363
		Thryssa	*mystax*	JQ012366
	Engraulinae	*Anchoa*	*cayorum*	JQ012347
		Anchoa	*colonensis*	JQ012383
		Anchoa	*cubana*	JQ012342
		Anchoa	*lamprotaenia*	JQ012379
		Anchoa	*lyolepis*	JQ012344
		Anchoa	*mitchilli*	JQ012357
		Anchoa	*mundeoloides*	JQ012419
		Anchoa	*nasus*	JQ012373
		Anchoa	*panamensis*	JQ012392
		Anchoa	*parva*	JQ012377
		Anchoa	*schofieldi*	JQ012349
		Anchoa	*spinifer*	KJ158085
		Anchoa	*walkeri*	JQ012369
		Anchovia	*clupeoides*	KJ158087
		Anchovia	*macrolepidota*	JQ012394
		Anchovia	*surinamensis*	JQ012402
		Anchoviella	*balboae*	JQ012371
		Anchoviella	*brevirostris*	JQ012412
		Anchoviella	*guianensis*	JQ012324
		Anchoviella	*lepidentostole*	JQ012414
		Cetengraulis	*edentulus*	JQ012385
		Cetengraulis	*mysticetus*	JQ012390
		Encrasicholina	*devisi*	JQ012364
		Encrasicholina	*punctifer*	AP011561
		Engraulis	*anchoita*	JQ012416
		Engraulis	*encrasicolus*	JQ012359
		Engraulis	*eurystole*	JQ012427
		Engraulis	*japonicus*	NC_003097
		Engraulis	*mordax*	JQ012350
		Engraulis	*ringens*	JQ012426

Family	Sub-Family	Genus	Species	GenBank acc.
Engraulidae	Engraulinae	*Lycengraulis*	*batesii*	JQ012411
		Lycengraulis	*grossidens*	JQ012396
		Lycengraulis	*poeyi*	JQ012370
		Lycothrissa	*crocodilus*	JQ012420
		Pterengraulis	*atherinoides*	JQ012323
		Stolephorus	*waitei*	AP011567
Pristigasteridae	Pelloninae	*Ilisha*	*africana*	NC_009584
		Ilisha	*elongata*	AP009141
		Ilisha	*megaloptera*	KJ158079
		Pellona	*flavipinnis*	EU552551
		Pellona	*harroweri*	KJ158088
	Pristigasterinae	*Odontognathus*	*mucronatus*	KJ158082
Outgroups				
Cyprinidae		*Danio*	*rerio*	NC_002333
Ictuluridae		*Ictalurus*	*punctatus*	NC_003489

Conclusion (Spanish)

En esta tesis doctoral se evaluó como factores ambientales oceanográficos actuales y pasados están asociados con la diversidad genética, estructuración poblacional, demografía histórica y tasas de diversificación evolutiva a una escala microevolutiva y macroevolutiva. Por otro lado, se identificó un problema en la sistemática del genero *Sprattus* evaluando sus relaciones filogenéticas y biogeografía histórica. Se incorporó por primera vez en conjunto análisis que incorporan el método comparativo filogenético para evaluar la señal filogenética histórica de variables y su relación con las tasa de diversificación de la especie *S. fuegensis* y del Orden Clupeiformes. De esta manera se construyó una metodología robusta para responder a la pregunta si variables ambientales, tales como temperatura, salinidad, oxigeno, pH, nutrientes, entre otras, juegan un rol importante a diferentes niveles taxonómicos en peces marinos, tomando como modelos el Orden Clupeiformes y la especie *S. fuegensis*. Se observó que en diferentes sub-familias y familias las tasas de diversificación fueron variables, sugiriendo que no todos lo grupos taxonómicos han respondido de igual manera a los cambios u oscilaciones de factores ambientales oceanográficos. Finalmente y dado los resultados obtenidos en los diferentes capítulos de esta tesis, se demostró claramente que factores ambientales oceanográficos jugaron un rol importante en la historia evolutiva desde especies hasta niveles taxonómicos mayores.

En un contexto microevolutivo contemporáneo, capitulo 1, los factores ambientales que estarían relacionados con variabilidad en frecuencias alélicas y diferenciación poblacional de *S. fuegensis* fueron el oxigeno y el pH. Estos factores podrían explicar la heterogeneidad encontrada a través de su impacto en la disponibilidad de recursos en zonas altamente productivas (e.g. fitoplancton), como ha sido sugerido en trabajos de Iriarte *et al.* (2007, 2010), y Alvarez et al. (2010). Por otro lado, en un contexto microevolutivo histórico, capitulo 2, las paleovariables

oceanográficas ambientales como temperatura, salinidad y disminución en el nivel del mar estuvieron correlacionadas con un incremento del N_E a través del tiempo, principalmente previo al ultimo máximo glacial (UMG). Aunque estas paleovariables, producto de las oscilaciones ambientales ocurridas en el UMG, se han relacionado cualitativamente con cambios demográficos históricos en organismos marinos (e.g. González-Wevar et al., 2011; Fraser et al., 2010, 2009; Macaya y Zuccarello, 2010; Montecinos et al., 2012; Brante et al., 2012; Cárdenas et al., 2009a; Guzmán et al., 2011; Haye et al., 2014; Sánchez et al., 2011a; Weis y Melzer, 2012) estos no evalúan si hay una correlación entre ellas, como fue en este estudio. Para variables ambientales contemporáneas con señal filogenética histórica, la salinidad, el fosfato (PO_4) y principalmente el nitrato (NO_3) mostraron estar relacionados con una tasa de diversificación intraespecífica exponencial decreciente. En relación a la tolerancia a la salinidad, se ha descrito una innovación clave para colonizar nuevos ambientes o ampliar rangos de distribución promoviendo así aislamiento reproductivo y posterior diferencia genética (Schultz and McCormick, 2013). Especies cercanamente emparentas así como también *S. fuegensis* habitan fiordos, por lo tanto tienen una alta capacidad de tolerar bajas salinidades (Landaeta et al., 2012; Limborg et al., 2009; Petereit et al., 2009). Frank et al., (2005) propone que los nutrientes son uno de los factores limitantes en los ecosistemas marinos. Además, se ha descrito que están correlacionados con la tasa de especiación de géneros (Cardenas y Harries, 2010). Incluso, se ha descrito que la variación en nutrientes puede promover persistencia y diferenciación de poblaciones aisladas y finalmente eventos de especiación (Allmon, 2001).

En un contexto macroevolutivo a nivel de genero, capitulo 3, aunque no se evaluó explícitamente las variables asociadas a las relaciones filogenéticas del genero *Sprattus* y su patrón de distribución antitropical, se pudo inferir que producto de cambios en las características ambientales oceanográficas ocurridas en el Mioceno, el ancestro del genero pudo dispersarse desde el hemisferio norte en el Océano Atlántico al hemisferio sur cuando las condiciones de temperatura disminuyeron producto del enfriamiento de la

época. Esto basado en los tiempos de divergencia del genero. Adicionalmente se encontró que el genero *Sprattus* es parafilético con el genero *Clupea*, poniendo en duda la clasificación actual del grupo y en donde estudios posteriores deberán ser realizados para dilucidar esta problemática taxonómica.

Por último, en un contexto macroevolutivo a nivel de Orden (i.e. Clupeiformes), Capitulo 4, las variables temperatura, salinidad, oxigeno, y clorofila mostraron señal filogenética histórica y están asociadas a un patrón incremental en las tasas de diversificación de especies desde el pasado hacia el presente hasta alcanzar una tasa constante, sin embargo en clados recientes correspondientes a subfamilias y familias presentan variaciones en sus tasas de diversificación. Como se describió previamente en relación al efecto de estas variables ambientales oceanográficas en contextos microevolutivos y su efecto a niveles macroevolutivo (Schultz and McCormick, 2013; Suto et al. 2012; Limborg et al., 2009; Lavoué et al. 2013; Bloom and Lovejpy 2014; Cardenas y Harries, 2010; Allmon, 2001), nuestros resultados dan cuenta del papel que cumplen esos factores sobre la dinámica histórica a nivel de especies y familias y orden.

En resumen, nuestros resultados dan cuenta como estos factores ambientales promueven variabilidad genética, estructuración poblacional, cambios en la demografía histórica y tasas de diversificación, de acuerdo con la hipótesis propuesta por Barnoski (2001) que describe a factores ambientales como los responsables del origen de nuevas especies. Finalmente, esta tesis expande nuestro entendimiento de las variables ambientales a diferentes categorías taxonómicas.

List of References

Alvarez R, Farías A, Galvez-Larach M *et al.* (2010) Sintesis del estudio "Investigación para el desarrollo de Área Marina Costera Protegida Chiloé, Palena y Guaitecas." In: *Conservando el mar de Chiloé, Palena y Guaitecas* (eds Hucke-Gaete R, Lo Moro P, Ruiz J), p. 174. Valdivia, Chile.

Aranis A, Meléndez R, Germán P, Cerna T F (2007) *Sprattus fuegensis* en aguas interiores de Chiloé, Chile (Osteichthyes: Clupeiformes:Clupeidae). *Gayana*, **71**, 102–113.

Atarhouch T, Rüber L, Gonzalez EG *et al.* (2006) Signature of an early genetic bottleneck in a population of Moroccan sardines (*Sardina pilchardus*). *Molecular Phylogenetics and Evolution*, **39**, 373–383.

Avise JC (2000) *Phylogeography: The History and Formation of Species*. Harvard University Press, Cambridge.

Balbontín F (2006) Ictioplancton de los canales y fiordos australes. In: *Avances en el conocimiento oceanográfico de las aguas interiores chilenas, Puerto Montt a cabo de Hornos* (eds Silva N, Palma S), pp. 115–120. Comité Oceanográfico Nacional, Pontificia Universidad Católica de Valparaíso, Valparaíso, Chile.

Balbontín F, Bernal R (1997) Distribución y abundancia del ictioplancton en la zona austral de Chile. *Ciencia y Tecnología del Mar*, **20**, 155–163.

Barnosky A (2001) Distinguishing the effects of the Red Queen and Court Jester on Miocene mammal evolution in the northern Rocky Mountains. *Journal of Vertebrate Paleontology*, **21**, 172–185.

Basedow SL, Eiane K, Tverberg V, Spindler M (2004) Advection of zooplankton in an Arctic fjord (Kongsfjorden, Svalbard). *Estuarine, Coastal and Shelf Science*, **60**, 113–124.

Bekkevold D, André C, Dahlgren TG *et al.* (2005) Environmental correlates of population differentiation in Atlantic herring. *Evolution*, **59**, 2656–68.

Benton M (2009) The Red Queen and the Court Jester: species diversity and the role of biotic and abiotic factors through time. *Science*, **323**, 728–732.

Bernal R, Balbontín F (1999) Ictioplancton de los fiordos entre el golfo de Penas y estrecho de Magallanes y factores ambientales asociados. *Ciencia y Tecnología del Mar*, **22**, 143–154.

Bernatchez L, Guyomard R, Bonhomme F (1992) DNA sequence variation of the mitochondrial control region among geographically and morphologically remote European brown trout *Salmo trutta* populations. *Molecular Ecology*, **1**, 161–173.

Beuchel F, Gulliksen B, Carroll ML (2006) Long-term patterns of rocky bottom macrobenthic community structure in an Arctic fjord (Kongsfjorden, Svalbard) in relation to climate variability (1980–2003). *Journal of Marine Systems*, **63**, 35–48.

Bloom DD, Lovejoy NR (2014) The evolutionary origins of diadromy inferred from a time-calibrated phylogeny for Clupeiformes (herring and allies). *Proceedings. Biological sciences / The Royal Society*, **281**, 20132081.

Bowen BW, Grant WS (1997) Phylogeography of the Sardines (*Sardinops* SPP.): Assessing Biogeographic Models and Population Histories in Temperate Upwelling Zones. *Evolution*, **51**, 1601–1610.

Bradbury IR, Snelgrove PVR, Bentzen P *et al.* (2009) Structural and functional connectivity of marine fishes within a semi-enclosed Newfoundland fjord. *Journal of Fish Biology*, **75**, 1393–409.

Brante A, Fernández M, Viard F (2012) Phylogeography and biogeography concordance in the marine gastropod *Crepipatella dilatata* (Calyptraeidae) along the southeastern Pacific coast. *Journal of Heredity*, **103**, 630–637.

Burridge CP (2002) Antitropicality of Pacific Fishes: Molecular Insights. *Environmental Biology of Fishes*, **65**, 151–164.

Burridge CP, White RWG (2000) Molecular phylogeny of the antitropical subgenus *Goniistius* (Perciformes: Cheilodactylidae: Cheilodactylus): evidence for multiple transequatorial divergences and non-monophyly. *Biological Journal of the Linnean Society*, **70**, 435–458.

Bustos CA, Balbontín F, Landaeta MF (2007) Spawning of the southern hake *Merluccius australis* (Pisces: Merlucciidae) in Chilean fjords. *Fisheries Research*, **83**, 23–32.

Bustos CA, Landaeta MF, Balbontín F (2008) Efectos ambientales sobre la variabilidad espacial del ictioplancton de Chile austral durante noviembre de 2005. *Revista Chilena de Historia Natural*, **81**, 205–219.

Canales-Aguirre CB, Ferrada S, Hernández CE, Galleguillos R (2010) Population structure and demographic history of *Genypterus blacodes* using microsatellite loci. *Fisheries Research*, **106**, 102–106.

Canales-Aguirre CB, Ferrada-Fuentes S, Hernández CE, Galleguillos R Seascape genetic in Patagonian fjords: a small pelagic fish viewpoint.

Caniupán M, Lamy F, Lange CB *et al.* (2011) Millennial-scale sea surface temperature and Patagonian Ice Sheet changes off southernmost Chile (53°S) over the past ~60 kyr. *Paleoceanography*, **26**, PA3221.

Cárdenas G (2009) Análisis de la composición de tallas, indicadores nutricionales y parámetros de crecimiento *Strangomera bentincki* (sardina común) y *Sprattus fuegensis* (sardina austral) en el sur de Chile. Tesis para optar al grado de Biología Marina, Universidad Austral de Chile.

Cárdenas L, Castilla JC, Viard F (2009a) A phylogeographical analysis across three biogeographical provinces of the south-eastern Pacific: the case of the marine gastropod *Concholepas concholepas*. *Journal of Biogeography*, **36**, 969–981.

Cárdenas AL, Harries PJ (2010) Effect of nutrient availability on marine origination rates throughout the Phanerozoic eon. *Nature Geoscience*, **3**, 430–434.

Cárdenas L, Silva AX, Magoulas A *et al.* (2009b) Genetic population structure in the Chilean jack mackerel, *Trachurus murphyi* (Nichols) across the South-eastern Pacific Ocean. *Fisheries Research*, **100**, 109–115.

Carlsson J (2008) Effects of microsatellite null alleles on assignment testing. *Journal of Heredity*, **99**, 616–623.

Carrasco C, Silva N (2010) Comparación de las características oceanográficas físicas y químicas presentes en la zona de Puerto Montt a la Boca del Guafo entre el invierno y la primavera de 2004 y entre las primaveras de 1995 y 2004. *Ciencia y Tecnología del Mar*, **33**, 17–44.

Castellani M, Rosland R, Urtizberea A, Fiksen Ø (2013) A mass-balanced pelagic ecosystem model with size-structured behaviourally adaptive zooplankton and fish. *Ecological Modelling*, **251**, 54–63.

Castoe TA, Poole AW, de Koning APJ *et al.* (2012) Rapid microsatellite identification from Illumina paired-end genomic sequencing in two birds and a snake. *PloS one*, **7**, e30953.

Ceballos SG, Lessa EP, Victorio MF, Fernández DA (2011) Phylogeography of the sub-Antarctic notothenioid fish *Eleginops maclovinus*: evidence of population expansion. *Marine Biology*, **159**, 499–505.

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

Chapuis M-P, Lecoq M, Michalakis Y *et al.* (2008) Do outbreaks affect genetic population structure? A worldwide survey in *Locusta migratoria*, a pest plagued by microsatellite null alleles. *Molecular Ecology*, **17**, 3640–53.

Cheng QQ, Lu DR (2006) Phylogenetic analysis and relative-rate test of nine Clupeidae fishes Osteichthyes Clupeiformes inferred from cytochrome b gene sequence of mitochondrial DNA. *Marine Fisheries*, **28**, 5–12.

Clark PU, Dyke AS, Shakun JD *et al.* (2009) The Last Glacial Maximum. *Science*, **325**, 710–714.

Coleman MA, Feng M, Roughan M, Cetina-Heredia P, Connell SD (2013) Temperate shelf water dispersal by Australian boundary currents: Implications for population connectivity. *Limnology & Oceanography: Fluids & Environments*, **3**, 295–309.

Corander J, Waldmann P, Sillanpaa MJ (2003) Bayesian analysis of genetic differentiation between populations. *Genetics*, **163**, 367–374.

Cousseau MB (1982) Revisión taxonómica y análisis de los caracteres morfométricos y:merísticos de la sardina f ueguina, *Sprattus fuegensis* (Jenyns, 1842) (Pisces, Clupeidae). *Revista de Investigación y Desarrollo Pesquero*, **3**, 77–94.

Darriba D, Taboada GL, Doallo R, Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods*, **9**, 772.

DaSilva JL, Anderson JB, Stravers J (1997) Seismic facies changes along a nearly continuous 24° latitudinal transect: the fjords of Chile and the northern Antarctic Peninsula. *Marine Geology*, **143**, 103–123.

Debes P V., Zachos FE, 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.

DeWoody JA, Avise JC (2000) Microsatellite variation in marine, freshwater and anadromous fishes compared with other animals. *Journal of Fish Biology*, **56**, 461–473.

Drummond AJ, Ho SYW, Phillips MJ, Rambaut A (2006) Relaxed phylogenetics and dating with confidence. *PLoS biology*, **4**, e88.

Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, **7**, 214.

Drummond AJ, Suchard MA (2010) Bayesian random local clocks, or one rate to rule them all. *BMC biology*, **8**, 114.

Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian Phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, **29**, 1969–1973.

Earl DA, VonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, **4**, 359–361.

Erwin DH (2009) Climate as a driver of evolutionary change. *Current Biology*, **19**, R575–R583.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611–20.

Evans KL, Gaston KJ (2005) Can the evolutionary-rates hypothesis explain species-energy relationships? *Functional Ecology*, **19**, 899–915.

Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, **1**, 47–50.

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, **10**, 564–7.

Ezard THG, Aze T, Pearson PN, Purvis A (2011) Interplay between changing climate and species' ecology drives macroevolutionary dynamics. *Science*, **332**, 349–51.

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

Ferrada S, Hernández K, Montoya R, Galleguillos R (2002) Estudio poblacional del recurso anchoveta (*Engraulis ringens* Jenyns 1842) (Clupeiformes, Engraulidae), mediante análisis de ADN. *Gayana*, **66**, 243–248.

Ferrada-Fuentes S, Galleguillos R, Canales-Aguirre CB *et al.* (In press) Development and characterization of thirty-three microsatellite markers for the Patagonian sprat, *Sprattus fuegensis* (Jenyns, 1842), using paired-end Illumina shotgun sequencing. *Conservation Genetics Resources*.

Fevolden SE, Westgaard JI, Pedersen T, Præbel K (2012) Settling-depth vs. genotype and size vs. genotype correlations at the Pan I locus in 0-group Atlantic cod *Gadus morhua*. *Marine Ecology Progress Series*, **468**, 267–278.

FitzJohn RG (2010) Quantitative Traits and Diversification. *Systematic Biology*, **59**, 619–633.

FitzJohn RG (2012) Diversitree : comparative phylogenetic analyses of diversification in R. *Methods in Ecology and Evolution*, **3**, 1084–1092.

Foll M, Gaggiotti O (2006) Identifying the environmental factors that determine the genetic structure of populations. *Genetics*, **174**, 875–891.

Frank KT, Petrie B, Choi JS, Leggett WC (2005) Trophic cascades in a formerly cod-dominated ecosystem. *Science*, **308**, 1621–3.

Fraser CI, Nikula R, Spencer HG, Waters JM (2009) Kelp genes reveal effects of subantarctic sea ice during the Last Glacial Maximum. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 3249–3253.

Fraser CI, Thiel M, Spencer HG, Waters JM (2010) Contemporary habitat discontinuity and historic glacial ice drive genetic divergence in Chilean kelp. *BMC Evolutionary Biology*, **10**, 203.

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

Galindo HM, Olson DB, Palumbi SR (2006) Seascape genetics: a coupled oceanographic-genetic model predicts population structure of Caribbean corals. *Current Biology : CB*, **16**, 1622–6.

Galleguillos R, Canales-Aguirre CB, Ferrada S (2012a) Genetic variability in jack mackerel (*Trachurus murphyi*): New SSRs loci and application. *Gayana*, **76**, 58–62.

Galleguillos R, Ferrada-Fuentes S, Canales-Aguirre CB *et al.* (2012b) *Determinación de unidades poblacionales de sardina austral entre la X y XII regiones de Chile.* Final Report FIP N°2010-17.

Galleguillos R, Troncoso L, Monsalves J, Oyarzun C (1997) Diferenciación poblacional en la sardina chilena *Strangomera bentincki* (Pisces: Clupeidae) análisis genético de variabilidad proteínica. *Revista Chilena de Historia Natural*, **70**, 351–361.

Gelman A, Carlin JB, Stern HS, Rubin DB (1995) *Bayesian Data Analysis*. Chapman & Hall/CRC, London.

Girard P, Angers B (2008) Assessment of power and accuracy of methods for detection and frequency-estimation of null alleles. *Genetica*, **134**, 187–97.

Glover KA, Skaala Ø, Limborg M, Kvamme C, Torstensen E (2011) Microsatellite DNA reveals population genetic differentiation among sprat (*Sprattus sprattus*) sampled throughout the Northeast Atlantic, including Norwegian fjords. *ICES Journal of Marine Science*, **68**, 2145–2151.

González-Wevar CA, David B, Poulin E (2011) Phylogeography and demographic inference in *Nacella* (Patinigera) *concinna* (Strebel, 1908) in the western Antarctic Peninsula. *Deep Sea Research Part II: Topical Studies in Oceanography*, **58**, 220–229.

Goodman SN (1999) Toward evidence-based medical statistics. 2: The Bayes factor. *Annals of internal medicine*, **130**, 1005–1013.

Goodson MS, Giske J, Rosland R (1995) Growth and ovarian development of *Maurolicus muelleri* during spring. *Marine Biology*, **124**, 185–195.

Grant WS (1986) Biochemical Genetic Divergence between Atlantic, *Clupea harengus*, and Pacific, *C. pallasi*, Herring. *Copeia*, **1986**, 714–719.

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

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

Grant WS, Leslie RW (1996) Late Pleistocene dispersal of Indian-Pacific sardine populations in an ancient lineage of the genusSardinops. *Marine Biology*, **126**, 133–142.

Grant WS, Leslie RW, Bowen BW (2005) Molecular genetic assessment of bipolarity in the anchovy genus *Engraulis*. *Journal of Fish Biology*, **67**, 1242–1265.

Grant WS, 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.

Guillot G, Estoup A, Mortier F, Cosson JF (2005a) A spatial statistical model for landscape genetics. *Genetics*, **170**, 1261–1280.

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

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

Guzmán BE, Nuñez JJ, Vejar A, Barriga EH, Gallardo CS (2011) Genetic diversity and population structure of two South American marine gastropods, *Crepipatella dilatata* and *C . fecunda*

(Gastropoda: Calyptraeidae): Distinct patterns based on developmental mode. *Italian Journal of Zoology*, **78**, 444–454.

Guzmán D, Silva N (2002) Caracterización física y química y masa de agua en los canales australes de Chile entre Boca del Guafo y Golfo Elefantes (Crucero CIMAR-Fiordo 4). *Ciencia y Tecnología del Mar*, **25**, 45–76.

Halvorsen KAT, Árnason E, Smith PJ, Mork J (2012) Mitochondrial DNA differentiation between the antitropical blue whiting species *Micromesistius poutassou* and *Micromesistius australis*. *Journal of Fish Biology*, **81**, 253–269.

Hammer Ø, Harper DAT, Ryan PD (2001) PAST: Paleontological statistics software package for education and data analysis. *Palaeontologia Electronica*, **4**, 1–9.

Hansen JE (1999) *Estimación de parámetros poblacionales del efectivo de sardina fueguina (Sprattus fuegensis) de la costa continental Argentina*. INIDEP Informe Técnico 27, Mar del Plata, Argentina.

Haye PA, Segovia NI, Muñoz-Herrera NC *et al.* (2014) Phylogeographic structure in benthic marine invertebrates of the southeast Pacific coast of Chile with differing dispersal potential. (BR MacKenzie, Ed,). *PloS one*, **9**, e88613.

Hernández CE, Rodríguez-Serrano E, Avaria-Llautureo J *et al.* (2013) Using phylogenetic information and the comparative method to evaluate hypotheses in macroecology. *Methods in Ecology and Evolution*, **4**, 401–415.

Hipsley CA, Müller J (2014) Beyond fossil calibrations: realities of molecular clock practices in evolutionary biology. *Frontiers in genetics*, **5**, 138.

Ho SYW, Lanfear R, Bromham L *et al.* (2011) Time-dependent rates of molecular evolution. *Molecular ecology*, **20**, 3087–101.

Hollin JT, Schilling DH (1981) Late Wisconsin-Weichselian mountain glaciers and small ice caps. In: *The Last Great Ice Sheets* (eds Denton GH, Hughes TJ), pp. 179–220. Wiley, New York.

Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources*, **9**, 1322–32.

Hulton NRJ, Purves RS, McCulloch RD, Sugden DE, Bentley MJ (2002) The Last Glacial Maximum and deglaciation in southern South America. *Quaternary Science Reviews*, **21**, 233–241.

Huybrechts P (2002) Sea-level changes at the LGM from ice-dynamic reconstructions of the Greenland and Antarctic ice sheets during the glacial cycles. *Quaternary Science Reviews*, **21**, 203–231.

Imbrie JZ, Hays JD, Martinson DG *et al.* (1984) The orbital theory of Pleistocene climate: support from a revised chronology of the marine d18O record. In: *Milankovitch and Climate* (eds Berger A, Imbrie JZ, Hays JD, Kukla G, Saltzman B), pp. 269–305. Reidel, Dordrecht.

Iriarte JL, González HE, Liu KK, Rivas C, Valenzuela C (2007) Spatial and temporal variability of chlorophyll and primary productivity in surface waters of southern Chile (41.5–43° S). *Estuarine, Coastal and Shelf Science*, **74**, 471–480.

Iriarte JL, González HE, Nahuelhual L (2010) Patagonian fjord ecosystems in southern Chile as a highly vulnerable region: problems and needs. *AMBIO*, **39**, 463–466.

Jérôme M, Lemaire C, Bautista JM, Fleurence J, Etienne M (2003) Molecular phylogeny and species identification of sardines. *Journal of agricultural and food chemistry*, **51**, 43–50.

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

Jørgensen HBH, Pertoldi C, Hansen MM, Ruzzante DE, Loeschcke V (2008) Genetic and environmental correlates of morphological variation in a marine fish: the case of Baltic Sea herring (*Clupea harengus*). *Canadian Journal of Fisheries and Aquatic Sciences*, **65**, 389–400.

Kaartvedt S, Røstad A, Klevjer TA (2009) Sprat *Sprattus sprattus* can exploit low oxygen waters for overwintering. *Marine Ecology Progress Series*, **390**, 237–249.

Kalinowski ST (2004) Counting Alleles with Rarefaction: Private Alleles and Hierarchical Sampling Designs. *Conservation Genetics*, **5**, 539–543.

Kalinowski ST (2005) HP-RARE 1.0: a computer program for performing rarefaction on measures of allelic richness. *Molecular Ecology Notes*, **5**, 187–189.

Kozak KH, Graham CH, Wiens JJ (2008) Integrating GIS-based environmental data into evolutionary biology. *Trends in Ecology and Evolution*, **23**, 141–148.

Kristoffersen JB, Salvanes AG V. (1998) Life history of *Maurolicus muelleri* in fjordic and oceanic environments. *Journal of Fish Biology*, **53**, 1324–1341.

Kristoffersen JB, Salvanes AGV (2009) Distribution, growth, and population genetics of the glacier lanternfish (*Benthosema glaciale*) in Norwegian waters: Contrasting patterns in fjords and the ocean. *Marine Biology Research*, **5**, 596–604.

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

Lamy F, Kaiser J (2009) Glacial to Holocene paleoceanographic and continental paleoclimate reconstructions based on ODP Site 1233/GeoB 3313 off southern Chile. In: *Past Climate Variability in South America and Surrounding Regions* Developments in Paleoenvironmental Research. (eds Vimeux F, Sylvestre F, Khodri M), pp. 129–156. Springer Netherlands, Dordrecht.

Lamy F, Kaiser J, Ninnemann U *et al.* (2004) Antarctic timing of surface water changes off Chile and Patagonian ice sheet response. *Science*, **304**, 1959–1962.

Landaeta MF, Bustos CA, Palacios-Fuentes P, Rojas P, Balbontín F (2011) Distribucion del ictioplancton en la Patagonia austral de Chile: potenciales efectos del deshielo de Campos de Hielo Sur. *Latin American Journal of Aquatic Research*, **39**, 236–249.

Landaeta MF, López G, Suárez-Donoso N, Bustos CA, Balbontín F (2012) Larval fish distribution, growth and feeding in Patagonian fjords: potential effects of freshwater discharge. *Environmental Biology of Fishes*, **93**, 73–87.

Lavoué S, Miya M, Musikasinthorn P, Chen W-J, Nishida M (2013) Mitogenomic evidence for an Indo-West Pacific origin of the Clupeoidei (Teleostei: Clupeiformes). *PloS one*, **8**, e56485.

Lavoué S, Miya M, Saitoh K, Ishiguro NB, Nishida M (2007) Phylogenetic relationships among anchovies, sardines, herrings and their relatives (Clupeiformes), inferred from whole mitogenome sequences. *Molecular phylogenetics and evolution*, **43**, 1096–105.

Leal E, Canales TM, Aranis A, Gonzáles M (2011) Actividad reproductiva y longitud de madurez de sardina austral *Sprattus fuegensis* en el mar interior de Chiloé , Chile. *Revista de Biología Marina y Oceanografía*, **46**, 43–51.

Lee W-J, Conroy J, Howell WH, Kocher TD (1995) Structure and evolution of teleost mitochondrial control regions. *Journal of Molecular Evolution*, **41**, 54–66.

Lemey P, Rambaut A, Welch JJ, Suchard MA (2010) Phylogeography takes a relaxed random walk in continuous space and time. *Molecular Biology and Evolution*, **27**, 1877–1885.

Li C, Ortí G (2007) Molecular phylogeny of Clupeiformes (Actinopterygii) inferred from nuclear and mitochondrial DNA sequences. *Molecular phylogenetics and evolution*, **44**, 386–98.

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

Limborg MT, Pedersen JS, Hemmer-Hansen J, Tomkiewicz J, Bekkevold D (2009) Genetic population structure of European sprat *Sprattus sprattus*: differentiation across a steep environmental gradient in a small pelagic fish. *Marine Ecology Progress Series*, **379**, 213–224.

Lindberg DR (1991) Marine Biotic Interchange Between the Northern and Southern Hemispheres. *Paleobiology*, **17**, 308–324.

Liu J-X, Gao T-X, Zhuang Z-M *et al.* (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.

Liu J-X, Tatarenkov A, Beacham TD *et al.* (2011) Effects of Pleistocene climatic fluctuations on the phylogeographic and demographic histories of Pacific herring (*Clupea pallasii*). *Molecular Ecology*, **20**, 3879–3893.

Macaya EC, Zuccarello GC (2010) Genetic structure of the giant kelp *Macrocystis pyrifera* along the southeastern Pacific. *Marine Ecology Progress Series*, **420**, 103–112.

MacKenzie BR, Horbowy J, Köster FW (2008) Incorporating environmental variability in stock assessment: predicting recruitment, spawner biomass, and landings of sprat (*Sprattus sprattus*) in the Baltic Sea. *Canadian Journal of Fisheries and Aquatic Sciences*, **65**, 1334–1341.

Manel S, Schwartz MK, Luikart G, Taberlet P (2003) Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology & Evolution*, **18**, 189–197.

Margonski P, Hansson S, Tomczak MT, Grzebielec R (2010) Climate influence on Baltic cod, sprat, and herring stock–recruitment relationships. *Progress in Oceanography*, **87**, 277–288.

Mariani S, Hutchinson WF, Hatfield EMC *et al.* (2005) North Sea herring population structure revealed by microsatellite analysis. *Marine Ecology Progress Series*, **303**, 245–257.

Martin RE (1996) Secular increase in nutrient levels through the Phanerozoic: implications for productivity, biomass, and diversity of the marine biosphere. *Palaios*, **11**, 209–219.

Martin AP, Palumbi SR (1993) Body size, metabolic rate, generation time, and the molecular clock. *Proceedings of the National Academy of Sciences of the United States of America*, **90**, 4087–4091.

Martínez-Porchas M, Hernández-Rodríguez M, Bückle-Ramírez LF (2009) Thermal behavior of the Pacific sardine (*Sardinops sagax*) acclimated to different thermal cycles. *Journal of Thermal Biology*, **34**, 372–376.

Mathews D (1884) *Report on sprat fishing during the winter of 1883-84*.

McCulloch RD, Bentley MJ, Purves RS *et al.* (2000) Climatic inferences from glacial and palaeoecological evidence at the last glacial termination, southern South America. *Journal of Quaternary Science*, **15**, 409–417.

McMillan WO, Palumbi SR (1997) Rapid rate of control-region evolution in Pacific butterflyfishes (Chaetodontidae). *Journal of Molecular Evolution*, **45**, 473–484.

Meskendahl L, Herrmann J-P, Temming a. (2010) Effects of temperature and body mass on metabolic rates of sprat, *Sprattus sprattus* L. *Marine Biology*, **157**, 1917–1927.

Minin VN, Bloomquist EW, Suchard M a (2008) Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. *Molecular Biology and Evolution*, **25**, 1459–1471.

Montecinos A, Broitman BR, Faugeron S *et al.* (2012) Species replacement along a linear coastal habitat: phylogeography and speciation in the red alga *Mazzaella laminarioides* along the south east pacific. *BMC Evolutionary Biology*, **12**, 97.

Münkemüller T, Lavergne S, Bzeznik B *et al.* (2012) How to measure and test phylogenetic signal. *Methods in Ecology and Evolution*, **3**, 743–756.

Nelson GJ (1967) Gill Arches of Teleostean Fishes of the Family Clupeidae. *Copeia*, **1967**, 389–399.

Nelson JS (2006) *Fishes of the world*. John Wiley and Sons, New York.

Niklitschek E, Toledo P, Hernández E *et al.* (2009) *Evaluación hidroacústica de pequeños pelágicos en aguas interiores de la X y XI regiones, año 2007.* Final Report FIP N^{o} 2007-05.

Ñiquen M, Bouchon M (2004) Impact of El Niño events on pelagic fisheries in Peruvian waters. *Deep Sea Research Part II: Topical Studies in Oceanography*, **51**, 563–574.

Ojaveer E, Kalejs M (2010) Ecology and long-term forecasting of sprat (*Sprattus sprattus* balticus) stock in the Baltic Sea: a review. *Reviews in Fish Biology and Fisheries*, **20**, 203–217.

Okazaki T, Kobayashi T, Uozumi Y (1996) Genetic relationships of pilchards (genus: *Sardinops*) with anti-tropical distributions. *Marine Biology*, **126**, 585–590.

Oksanen J, Blanchet, F. Guillaume Kindt R, Legendre P *et al.* (2013) VEGAN: community ecology package. R package version 2.0-10. http://CRAN.R-project.or.

Olsen RB, Richardson K, Simonsen V (2002) Population differentiation of eelpout *Zoarces viviparus* in a Danish fjord. *Marine Ecology Progress Series*, **227**, 97–107.

Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, **4**, 535–538.

Pagel M (1997) Inferring evolutionary processes from phylogenies. *Zoologica Scripta*, **26**, 331–348.

Pagel M (1999) Inferring the historical patterns of biological evolution. *Nature*, **401**, 877–884.

Pagel M, Meade A (2004) A phylogenetic mixture model for detecting pattern-heterogeneity in gene sequence or character-state data. *Systematic Biology*, **53**, 571–581.

Pagel M, Meade A (2005) Mixture models in phylogenetic inference. In: *Mathematics of Evolution and Phylogeny* (ed Gascuel O), pp. 121–142. Oxford University Press, Oxford.

Pagel M, Meade A (2006) Bayesian analysis of correlated evolution of discrete characters by reversible-jump Markov chain Monte Carlo. *The American Naturalist*, **167**, 808–825.

Pagel M, Meade A (2008) Modelling heterotachy in phylogenetic inference by reversible-jump Markov chain Monte Carlo. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **363**, 3955–3964.

Pahnke K, Sachs JP (2006) Sea surface temperatures of southern midlatitudes 0–160 kyr B.P. *Paleoceanography*, **21**, PA2003.

Palumbi SR (1996a) PCR and molecular systematics. In: *Molecular Systematics* (eds Hillis D, Moritz C, Mable B). Sinauer Press.

Palumbi SR (1996b) What can molecular genetics contribute to marine biogeography? An urchin's tale. *Journal of Experimental Marine Biology and Ecology*, **203**, 75–92.

Pantoja S, Luis Iriarte J, Daneri G (2011) Oceanography of the Chilean Patagonia. *Continental Shelf Research*, **31**, 149–153.

Parker J, Rambaut A, Pybus OG (2008) Correlating viral phenotypes with phylogeny: accounting for phylogenetic uncertainty. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases*, **8**, 239–46.

Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, **6**, 288–295.

Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics*, **28**, 2537–9.

Peck MA, Baumann H, Bernreuther M *et al.* (2012a) The ecophysiology of *Sprattus sprattus* in the Baltic and North Seas. *Progress in Oceanography*, **103**, 42–57.

Peck MA, Kanstinger P, Holste L, Martin M (2012b) Thermal windows supporting survival of the earliest life stages of Baltic herring (*Clupea harengus*). *ICES Journal of Marine Science*, **69**, 529–536.

Peck MA, Reglero P, Takahashi M, Catalán IA (2013) Life cycle ecophysiology of small pelagic fish and climate-driven changes in populations. *Progress in Oceanography*, **116**, 220–245.

Penzo E, Gandolfi G, Bargelloni L, Colombo L, Patarnello T (1998) Messinian salinity crisis and the origin of freshwater lifestyle in western Mediterranean gobies. *Molecular Biology and Evolution*, **15**, 1472–1480.

Pequeño G (1989) Peces de Chile. Lista sistemática revisada y comentada. *Revista de Biología Marina, Valparaíso*, **24**, 1–132.

Pespeni MH, Chan F, Menge BA, Palumbi SR (2013) Signs of adaptation to local pH conditions across an environmental mosaic in the California Current Ecosystem. *Integrative and Comparative Biology*, **53**, 857–70.

Petereit C, Haslob H, Kraus G, Clemmesen C (2008) The influence of temperature on the development of Baltic Sea sprat (*Sprattus sprattus*) eggs and yolk sac larvae. *Marine Biology*, **154**, 295–306.

Petereit C, Hinrichsen H-H, Voss R *et al.* (2009) The influence of different salinity conditions on egg buoyancy and development and yolk sac larval survival and morphometric traits of Baltic Sea sprat (*Sprattus sprattus* balticus Schneider). *Scientia Marina*, **73**, 59–72.

Phillips MJ (2009) Branch-length estimation bias misleads molecular dating for a vertebrate mitochondrial phylogeny. *Gene*, **441**, 132–40.

Porobić J, Canales-Aguirre CB, Ernst B, Galleguillos R, Hernández CE (2013) Biogeography and historical demography of the Juan Fernández Rock Lobster, *Jasus frontalis* (Milne Edwards, 1837). *The Journal of Heredity*, **104**, 223–233.

Le Port A, Pawley M, Lavery S (2013) Speciation of two stingrays with antitropical distributions: low levels of divergence in mitochondrial DNA and morphological characters suggest recent evolution. *Aquatic Biology*, **19**, 153–165.

Prahl FG, Muehlhausen LA, Zahnle DL (1988) Further evaluation of long-chain alkenones as indicators of paleoceanographic conditions. *Geochimica et Cosmochimica Acta*, **52**, 2303–2310.

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

Rabosky DL (2014) Automatic Detection of Key Innovations, Rate Shifts, and Diversity-Dependence on Phylogenetic Trees. *PLoS ONE*, **9**, e89543.

Rambaut A, Drummond AJ (2007) Tracer version 1.4. *Computer program and documentation distributed by the author, website http://beast. bio. ed. ac. uk/Tracer [accessed August 2009]*.

Rand DM (1994) Thermal habit, metabolic rate and the evolution of mitochondrial DNA. *Trends in Ecology & Evolution*, **9**, 125–131.

Raymond M, Rousset F (1995) GENEPOP (Version 1.2): Population Genetics Software for Exact Tests and Ecumenicism. *Journal of Heredity*, **86**, 248–249.

Reusch TBH (2014) Climate change in the oceans: evolutionary versus phenotypically plastic responses of marine animals and plants. *Evolutionary Applications*, **7**, 104–22.

Rice WR (1989) Analyzing tables of statistical test. *Evolution*, **43**, 223–225.

Riginos C, Liggins L (2013) Seascape Genetics: Populations, Individuals, and Genes Marooned and Adrift. *Geography Compass*, **7**, 197–216.

Rijnsdorp AD, Peck MA, Engelhard GH, Mollmann C, Pinnegar JK (2009) Resolving the effect of climate change on fish populations. *ICES Journal of Marine Science*, **66**, 1570–1583.

Rodríguez F, Oliver JL, Marín A, Medina JR (1990) The general stochastic model of nucleotide substitution. *Journal of Theoretical Biology*, **142**, 485–501.

Rousset F (2008) Genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources*, **8**, 103–106.

Sánchez RP, Remeslo A, Madirolas A, de Ciechomski JD (1995) Distribution and abundance of post-larvae and juveniles of the Patagonian sprat, *Sprattus fuegensis*, and related hydrographic conditions. *Fisheries Research*, **23**, 47–81.

Sánchez R, Sepúlveda R, Brante A, Cárdenas L (2011) Spatial pattern of genetic and morphological diversity in the direct developer *Acanthina monodon* (Gastropoda: Mollusca). *Marine Ecology Progress Series*, **434**, 121–131.

Schultz ET, McCormick SD (2013) Euryhalinity in an Evolutionary Context. *EEB Articles*, **Paper29**, 1–100.

Selkoe KA, Henzler CM, Gaines SD (2008) Seascape genetics and the spatial ecology of marine populations. *Fish and Fisheries*, **9**, 363–377.

Semenova A V., Stroganov AN, Smirnov AA, Afanas'ev KI, Rubtsova GA (2014) Genetic variations in Clupea pallasii herring from Sea of Okhotsk based on microsatellite markers. *Russian Journal of Genetics*, **50**, 175–179.

Sernapesca (2012) Anuario estadístico de pesca.

Shaw PW, Turan C, Wright JM, O'Connell M, Carvalho GR (1999) Microsatellite DNA analysis of population structure in Atlantic herring (*Clupea harengus*), with direct comparison to allozyme and mtDNA RFLP analyses. *Heredity*, **83**, 490–499.

Shirakova EN (1978a) Some biological features of Tierra del Fuego sprat. *Soviet Journal of Marine Biology*, **4**, 697–702.

Shirakova EN (1978b) Contribution to the biology of Tierra del Fuego sprat *Sprattus fuegensis* (Jenyns, 1842). *Biology of the Sea*, **3**, 78–84.

Sievers HA (2006) Temperatura y salinidad en canales y fiordos australes. In: *Avances en el conocimiento oceanográfico de las aguas interiores chilenas, Puerto Montt a Cabo de Hornos* (eds Silva N, Palm S), pp. 31–36. Valparaíso, Chile.

Sievers HA, Calvete C, Silva N (2002) Distribución de características físicas, masas de agua y circulación general para algunos canales australes entre el Golfo de Penas y el Estrecho de Magallanes (Crucero CIMAR-FIORDO 2), Chile. *Ciencia y Tecnología del Mar*, **25**, 17–43.

Sievers HA, Silva N (2006) Masas de agua y circulación en los canales y fiordos australes. In: *Avances en el conocimiento oceanográfico de las aguas interiores chilenas, Puerto Montt a Cabo de Hornos* (eds

Silva N, Palma S), pp. 53–58. Comité Oceanográfico Nacional, Pontificia Universidad Católica de Valparaíso, Valparaíso, Chile.

Silva N, Calvete C (2002) Características oceanográficas físicas y químicas de canales australes chilenos entre el Golfo de Penas y el Estrecho de Magallanes. *Ciencia y Tecnología del Mar*, **25**, 23–88.

Silva N, Calvete C, Sievers HA (1997) Características oceanográficas físicas y químicas de canales australes chilenos entre Puerto Montt y Laguna San Rafael (Crucero Cimar-Fiordo 1). *Ciencia y Tecnología del Mar*, **20**, 23–106.

Silva N, Calvete C, Sievers HA (1998) Masas de agua y circulación general para algunos canales australes entre Puerto Montt y Laguna San Rafael, Chile (Crucero Cimar-Fiordo 1). *Ciencia y Tecnología del Mar*, **21**, 17–48.

Silva N, Guzmán D (2006) Condiciones oceanográficas físicas y químicas, entre la Boca del Guafo y Fiordo Aysén (Crucero CIMAR 7 Fiordos). *Ciencia y Tecnología del Mar*, **29**, 25–44.

Silva N, Valdenegro A (2008) Caracterización oceanográfica de canales australes chilenos entre la Boca del Guafo y los canales Pulluche - Chacabuco (CIMAR 8 fiordos)*. *Ciencia y Tecnología del Mar*, **31**, 5–44.

Spalding MD, Fox HE, Allen GR *et al.* (2007) Marine ecoregions of the world: a bioregionalization of coastal and shelf areas. *BioScience*, **57**, 573–583.

Stepien CA, Rosenblatt RH (1996) Genetic Divergence in Antitropical Pelagic Marine Fishes (*Trachurus*, *Merluccius*, and *Scomber*) between North and South America. *Copeia*, **1996**, 586–598.

Stepputtis D, Hinrichsen H-H, Böttcher U, Götze E, Mohrholz V (2011) An example of meso-scale hydrographic features in the central Baltic Sea and their influence on the distribution and vertical migration of sprat, *Sprattus sprattus* balticus (Schn.). *Fisheries Oceanography*, **20**, 82–88.

Storfer A, Murphy MA, Evans JS *et al.* (2007) Putting the "landscape" in landscape genetics. *Heredity*, **98**, 128–42.

Storfer A, Murphy MA, Spear SF, Holderegger R, Waits LP (2010) Landscape genetics: where are we now? *Molecular Ecology*, **19**, 3496–514.

Sugaya T, Sato M, Yokoyama E *et al.* (2008) Population genetic structure and variability of Pacific herring *Clupea pallasii* in the stocking area along the Pacific coast of northern Japan. *Fisheries Science*, **74**, 579–588.

Sugden DE, Bentley MJ, Fogwill CJ *et al.* (2005) Late-Glacial glacier events in southernmost South America: a blend of "northern" and "southern" hemispheric climatic signals? *Geografiska Annaler*, **87A**, 273–288.

Sverdrup HU, Johnson MW, Fleming RH (1942) *The Oceans*. Prentice-Hall, Englewood Cliffs.

Teacher AG, André C, Jonsson PR, Merilä J (2013) Oceanographic connectivity and environmental correlates of genetic structuring in Atlantic herring in the Baltic Sea. *Evolutionary Applications*, **6**, 549–67.

Valdenegro A, Silva N (2003) Caracterización oceanográfica física y química de la zona de canales y fiordos australes de Chile entre el Estrecho de Magallanes y Cabo de Hornos (CIMAR 3 FIORDOS). *Ciencia y Tecnología del Mar*, **26**, 19–60.

VanValen L (1973) A new evolutionary law. *Evolutionary theory*, **1**, 1–30.

Vimeux F, Sylvestre F, Khodri M (2009) *Past Climate Variability in South America and Surrounding Regions* (F Vimeux, F Sylvestre, M Khodri, Eds,). Springer Netherlands.

Vodolazhskii DI, Kornienko I V., Voinova N V. (2008) Hypervariability of the D-loop region in mitochondrial DNA of Russian sturgeon *Acipenser gueldenstaedtii* (Acipenseriformes, Acipenseridae). *Journal of Ichthyology*, **48**, 188–197.

Wang TH, Donaldson YK, Brettle RP, Bell JE, Simmonds P (2001) Identification of shared populations of human immunodeficiency virus type 1 infecting microglia and tissue macrophages outside the central nervous system. *Journal of virology*, **75**, 11686–99.

Weis A, Melzer RR (2012) How did sea spiders recolonize the Chilean fjords after glaciation ? DNA barcoding of Pycnogonida , with remarks on phylogeography of *Achelia assimilis* (Haswell, 1885). *Systematics and Biodiversity*, **10**, 361–374.

Whitehead PJP (1964) A new genus and subgenus of clupeid fishes and notes on the genera Clupea, Sprattus and Clupeonella. *Journal of Natural History Series 13*, **7**, 321–330.

Whitehead PJP (1985) FAO species catalogue. Vol.7. Clupeoid fishes of the world. An annotated and illustrated catalogue of the herrings, sardines, pilchards, sprats, anchovies and wolf- herrings. Part 1 - Chirocentridae, Clupeidae and Pristigasteridae. *FAO Fisheries Synopsis*, **125**, 1–303.

Whitehead PJP, Smith PJ, Robertson DA (1985) The two species of sprat in New Zealand waters (*Sprattus antipodum* and *S. muelleri*). *New Zealand Journal of Marine and Freshwater Research*, **19**, 261–271.

Wildes SL, Vollenweider, Johanna J. Nguyen HT, Guyon JR (2010) Genetic variation between outer-coastal and fjord populations of Pacific herring (*Clupea pallasii*) in the eastern Gulf of Alaska. *Fishery Bulletin*, **109**, 382–393.

Willis K, Cottier F, Kwasniewski S, Wold A, Falk-Petersen S (2006) The influence of advection on zooplankton community composition in an Arctic fjord (Kongsfjorden, Svalbard). *Journal of Marine Systems*, **61**, 39–54.

Wilson AB, Teugels GG, Meyer A (2008) Marine incursion: the freshwater herring of Lake Tanganyika are the product of a marine invasion into west Africa. (C Moritz, Ed,). *PloS one*, **3**, e1979.

Xia X (2013) DAMBE5: a comprehensive software package for data analysis in molecular biology and evolution. *Molecular biology and evolution*, **30**, 1720–8.

Xia X, Xie Z, Salemi M, Chen L, Wang Y (2003) An index of substitution saturation and its application. *Molecular Phylogenetics and Evolution*, **26**, 1–7.

Appendix : Manuscript accepted

Development and characterization of thirty-three microsatellite markers for the Patagonian sprat, *Sprattus fuegensis* (Jenyns, 1842), using paired-end Illumina shotgun sequencing

Sandra Ferrada-Fuentes[1], Ricardo Galleguillos[1], **Cristian B. Canales-Aguirre**[1], Cara N. Love[2], Kenneth L. Jones[3], Stacey L. Lance[2]

1Laboratorio de Genética y Acuicultura, Departamento de Oceanografía, Facultad de Ciencias Naturales y Oceanográficas, Universidad de Concepción, Casilla 160-C, Concepción, Chile.
2Savannah River Ecology Laboratory, University of Georgia, Aiken, SC 29802
3Department of Biochemistry and Molecular Genetics, University of Colorado School of Medicine, Aurora, CO 80045

Manuscript accepted in Conservation Genetics Resources

We isolated and characterized a total of 33 microsatellite loci from the Patagonian sprat *Sprattus fuegensis*, a recent exploited marine resource with a conservation status unknowing. Loci were screened in 24 individuals from the inshore waters of the Aysén Fjord, Chile. The number of alleles per locus ranged from 7 to 24, observed heterozygosity ranged from 0.217 to 0.875, and the probability of identity values ranged from 0.006 to 0.133. These new loci will provide tools for examining population genetic structure, estimating effective population size and provide information to fisheries management and conservation.

Microsatellite loci are widely used in marine resources providing useful information to fishery management and promote future long-term strategies (e.g. Canales-Aguirre et al. 2010). *Sprattus fuegensis* is a small pelagic fish that inhabit in shallow waters, around 50 m deep, and near the coastal line distributed in fjords and channels of the Pacific Southeast Ocean and Atlantic Southwest Ocean (Aranis et al.

2007). Although *S. sprattus* has been exploited since 2006, little is known about its genetic diversity and population structure.

Total DNA was extracted from one individual of *Sprattus fuegensis,* using Nucleospin Tissue Kit (Machery & Nagel), following the manufacturer's protocol. An Illumina paired-end shotgun library was prepared by shearing 1 μg of DNA using a Covaris S220 and following the standard protocol of the Illumina TruSeq DNA Library Kit and using a multiplex identifier adaptor index. Illumina sequencing was conducted on the HiSeq with 100 bp paired-end reads. Five million of the resulting reads were analyzed with the program *PAL_FINDER_v0.02.03* (Castoe et al. 2012) to extract those reads that contained di-, tri-, tetra-, penta-, and hexanucleotide microsatellites. Once positive reads were identified in *PAL_FINDER_v0.02.03* they were batched to a local installation of the program Primer3 (version 2.0.0) for primer design. Loci for which the primer sequences only occurred one time in the 5 million reads were selected. Forty-eight loci of the 2446 that met this criterion were chosen. One primer from each pair was modified on the 5' end with an engineered sequence (CAG tag 5'-CAGTCGGGCGTCATCA-3') to enable use of a third primer in the PCR (identical to the CAG tag) that was fluorescently labeled. The sequence GTTT was added to primers without the universal CAG tag addition.

Forty-eight primer pairs were tested for amplification and polymorphism using DNA obtained from four individuals. PCR amplifications were performed in a 12.5 μL volume (10 mM Tris pH 8.4, 50 mM KCl, 25.0 μg/ml BSA, 0.4 μM unlabeled primer, 0.04μM tag labeled primer, 0.36μM universal dye-labeled primer, 3.0 mM $MgCl_2$, 0.8 mM dNTPs, 0.5 units AmpliTaq Gold® Polymerase (Applied Biosystems), and 20 ng DNA template) using an Applied Biosystems GeneAmp 9700. A touchdown thermal cycling program encompassing a 10°C span of annealing temperatures ranging between 65-55°C (TD65) was used for all loci. PCR products were run on an ABI-3130xl

sequencer and sized with Naurox size standard, except that unlabeled primers started with GTTT. Results were analyzed using GeneMapper version 3.7 (Applied Biosystems). Unambiguous scoring was possible for 33 polymorphic loci. We assessed the variability of the 33 polymorphic loci in 24 specimens collected from the inshore waters of the Aysén Fjord, southern Chile.

Conditions and characteristics of the loci are provided in Table 1. We estimated the number of alleles per locus (*k*), observed and expected heterozygosity (H_o and H_e), and probability of identity (PI) using GenAlEx v6 (Peakall and Smouse 2006). Tests for deviations from Hardy-Weinberg equilibrium (HWE) and for linkage disequilibrium were conducted using GENEPOP v4.0 (Rousset 2008).

Thirty-three of the tested primer pairs amplified high quality PCR product that exhibited polymorphism. The number of alleles per locus ranged from 7 to 24, observed heterozygosities ranged from 0.217 to 0.875, and the probability of identity values from 0.006 to 0.133 (Table 1). Seven loci showed moderate polymorphism (7-9 alleles) and 26 loci were highly polymorphic (11-24 alleles). After Bonferroni correction for multiple comparisons 12 loci showed significant deviations from expectations under HWE. Linkage disequilibrium was detected in 3 of 528-paired loci comparisons; no evidence exists of physical linkage between microsatellites developed.

These new loci will provide tools for examining genetic diversity and aid in better understanding the, ecology, and conservation of *S. fuegensis*. This information will help determine appropriate fisheries management action along its distribution and monitory changes on its genetic diversity in the new marine protected area Bahia Tictoc in Corcovado gulf, Chile.

References

Aranis A, Meléndez R, Germán P, Cerna T F (2007) *Sprattus fuegensis* en aguas interiores de Chiloé, Chile (Osteichthyes: Clupeiformes:Clupeidae). Gayana 71:102–113.

Canales-Aguirre CB, Ferrada S, Hernández CE, Galleguillos R (2010) Population structure and demographic history of *Genypterus blacodes* using microsatellite loci. Fisheries Research 106:102–106. doi: 10.1016/j.fishres.2010.06.010

Castoe TA, Poole AW, de Koning APJ, et al. (2012) Rapid microsatellite identification from Illumina paired-end genomic sequencing in two birds and a snake. PloS one 7:e30953. doi: 10.1371/journal.pone.0030953

Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes 6:288–295. doi: 10.1111/j.1471-8286.2005.01155.x

Rousset F (2008) Genepop'007: a complete re-implementation of the genepop software for Windows and Linux. Molecular Ecology Resources 8:103–106. doi: 10.1111/j.1471-8286.2007.01931.x

Table 1: Details for 33 polymorphic microsatellite loci developed for *Sprattus fuegensis*. The size indicates the range of observed alleles in base pairs and includes the length of the CAG tag; number of individuals genotyped is *N*; *k* is number of alleles observed; H_o and H_e are observed and expected heterozygosity, respectively; and PI is the probability of identity for each locus.

Locus	Primer Sequence 5' --> 3'	Repeat motif	Size (bp)	N	*K*	H_o	H_e	PI
Spfu2†	F: *CCTGCTGCTCTCATGTCTCTAGC	TCTG	202-318	22	20	0.409	0.942	0.006
	R: AACCACAGTCACCCATTCTCC							
Spfu3	F: *CGTCGCTGGAGACACAAAGG	AAAG	138-170	24	9	0.75	0.817	0.058
	R: AAGCACTTGATGGTGTCAGGC							
Spfu4†	F: *ATGGGTCCTTCCACAGTCC	TCTG	134-214	23	11	0.391	0.868	0.032
	R: AAGTTTCCCTGCCTGTCCG							
Spfu5†	F: *CCTCTCTCCGTCCATCTATTCC	TCTG	263-377	23	13	0.217	0.889	0.022
	R: AATTGATACAGAAGGTGTAGGTAGGC							
Spfu6	F: *CTTCTCCCGTGACGTGTAGC	TCTG	230-280	22	10	0.87	0.847	0.039
	R:ACACACACCTTGCAGAAGGC							
Spfu7	F: *CACTACAGTGCGACTCAGCC	ACTG	139-183	24	9	0.583	0.811	0.06
	R:ACATTGAACCAGGCACACC							
Spfu9	F: *CAGCACAGCATCTCAAAGCC	TCTG	200-252	24	14	0.667	0.908	0.016
	R: ACTGACCTGACAAGGACAAAGC							
Spfu10	F: *AGTTTGCACCGAGTCTGGC	TCTG	200-272	23	13	0.609	0.853	0.036
	R: ACTGTATCCTCTTGATAACCGGC							
Spfu11	F: *CAGAACTTAAACGGCCCAGC	AATC	223-285	24	16	0.875	0.878	0.025
	R: AGAAGGTGTACCAACTGAGGGC							
Spfu12†	F: *GGGTGTCTTTGACAGGTCG	AATC	150-186	21	9	0.333	0.83	0.049
	R: AGCAGTGATCACATCCAGACC							
Spfu13†	F: *TGACGTGTTGGTTTCCTTGG	AAAG	333-397	21	18	0.333	0.92	0.011
	R: AGCTGAGCACCATCTCCTCC							
Spfu14†	F: *TCCCTGTCTCATAGATAGAGATGCC	TCTG	128-222	23	17	0.217	0.926	0.01
	R: ATTGGCTGCATTGAGAGAGC							
Spfu17	F: *ACTCATGACTTGCTGCCTCC	AAAC	201-227	24	7	0.375	0.706	0.133
	R: CAGACTTGCCTTTCCTTCCC							
Spfu23	F: *GATAAGGCCCAGACTGAAGC	TCTG	124-192	24	17	0.708	0.894	0.02
	R: CCCAGTGTCACGAGCACC							
Spfu24†	F: *CCACCTGCTTGCTTGC	AAAG	142-182	24	11	0.458	0.875	0.028
	R: CTGGCACGTATAGCATCAGCC							
Spfu25	F: *CCAACTACCCAGTGTTGAGTGC	TCTG	158-286	24	17	0.625	0.924	0.011
	R: CTTCTTGGCCTGCTCGTCC							
Spfu26	F: *CAGTCAAGGAGATAGAGAGAGGC	TCTG	140-232	22	18	0.727	0.916	0.013
	R: CTTTATTGTTCTCTGTCCCGC							
Spfu27†	F: *TCAACTCCGCTGTGATGAGG	ATCT	255-379	20	21	0.45	0.94	0.007
	R: GAGTCTGTGTGCTAACAGAAGTGC							
Spfu28†	F: *CATTTAGTATCTCTCAACCTCAGTTGC	TCTG	191-239	20	12	0.25	0.895	0.02
	R: GCAATGTCGGTTTCAAACTCC							
Spfu29	F: *TGGGAGAAGGAATTTAGAGAGGG	AATC	238-306	22	11	0.682	0.857	0.036
	R: GCACAGAAATTCAGCACTGGG							

Spfu30	F: *CAACCTCTAACTCTACGGCAGG	AATC	111-191	24	16	0.833	0.92	0.012
	R: GCTGATTGGTTGAAGTCAGGC							
Spfu31†	F: *CAATCTTCTCTCTCAATCCAGCC	ATCT	175-299	21	15	0.333	0.789	0.055
	R: GGAGGAGAGGACAAATAAATAGGG							
Spfu32†	F: *TGTCTGATTTGGACTGATTGGG	AATC	160-188	21	9	0.238	0.833	0.049
	R: GGGAACGCATTATGTTACCTCC							
Spfu38	F: *CACTCATTTCAATTCAACTCC	ACTG	259-303	21	12	0.429	0.788	0.067
	R: GTGTTAAGGAGAGAACCGGC							
Spfu39†	F: *TCCCTCACGAAGTGATGAGC	ATCT	147-227	21	13	0.238	0.898	0.019
	R: TCACAATACTAACACCTGAAATTCCC							
Spfu41	F: *GGAACAGCTCAGGTGCAGG	TCTG	298-398	22	14	0.818	0.904	0.017
	R: TGCATGTGTCCGTGATAAAGC							
Spfu42	F: *TCCACAAGTTTCCACTGCCC	AATG	165-233	24	14	0.875	0.885	0.023
	R: TGCCTGGTGAGGTTGTACTCC							
Spfu43	F: *CAAAGGATTGGTGGCGTAGC	TCTG	229-293	23	16	0.696	0.863	0.029
	R: TGCTTCAGCTGTGCATAGGG							
Spfu44	F: *GCTGATGCTGGAGAGCTGC	TCTG	184-302	23	24	0.696	0.942	0.006
	R: TGTCAGACATATTGTTTGAAGTTAGACC							
Spfu45	F: *TGTACCTGTCCCTCAGTGCC	TCTG	133-193	23	14	0.826	0.895	0.02
	R: TGTGTCTGCCACACTGAACG							
Spfu46	F: *GGACAATGCTTTCCCTGACC	AATC	124-152	24	8	0.833	0.863	0.034
	R: TTCAGGCCTAACTCATCAGAGG							
Spfu47	F: *CCACCTCCTGCTTCTGTGC	AAAC	147-183	24	9	0.542	0.803	0.066
	R: TTTAATTAGGTGGGTGTATCGCC							
Spfu48	F: *GGCAGAGCACTCAATTCACC	AAAC	138-172	24	13	0.792	0.885	0.024
	R: TTTGACACCTGGCTGAGTGG							

* indicates CAG tag (5'- CAGTCGGGCGTCATCA-3') label; † indicates significant deviations from Hardy-Weinberg expectations after Bonferroni corrections.