

**PATRONES DE DIVERGENCIA GENÓMICA EN
DIFERENTES ETAPAS DEL CONTINUO DE
ESPECIACIÓN EN EL GÉNERO *ORESTIAS*
(TELEOSTEI; CYPRINODONTIDAE)**

Tesis entregada a la Universidad de Chile
en cumplimiento parcial de los requisitos para optar al grado de

Doctora en Ciencias con Mención en Ecología y Biología Evolutiva

Facultad De Ciencias

Por

PAMELA MARITZA MORALES HENRIQUEZ

Mayo, 2018

Director de Tesis:

Dr. Marco A. Méndez T.

FACULTAD DE CIENCIAS
UNIVERSIDAD DE CHILE
INFORME DE APROBACION
TESIS DE DOCTORADO

Se informa a la Escuela de Postgrado de la Facultad de Ciencias que la Tesis de
Doctorado presentada por la candidata.

PAMELA MARITZA MORALES HENRIQUEZ

ha sido aprobada por la comisión de Evaluación de la tesis como requisito para optar al
grado de Doctora en Ciencias con mención en Ecología y Biología Evolutiva, en el
examen de Defensa Privada de Tesis rendido el día 30 de Abril de 2018.

Director de Tesis:

Dr. Marco A. Méndez T.

Comisión de Evaluación de la Tesis:

Dr. David Véliz

Dr. Elie Poulin

Dra. Leyla Cárdenas

Dra. Alejandra González

A la Aba Luz y a la Aba Mafa.

A mis papás.

Y a Isidora y María Jesús.



Nací en Santiago de Chile, donde he vivido toda mi vida. Entré a estudiar Licenciatura en Biología en la Pontificia Universidad Católica de Chile. Terminando la carrera ya me había inclinado hacia el área ecológica: realicé un seminario de investigación con el Dr. Patricio Ojeda, y al terminar me propuso ir al laboratorio del Dr. Elie Poulin en la Universidad de Chile. Ahí conocí lo que realmente me gusta: la biología evolutiva. Con Elie realicé mi seminario de investigación final para obtener la Licenciatura, y al año siguiente comencé el Magíster en Ciencias en la Universidad de Chile, bajo su tutela. Fue una época increíble de aprendizaje. Mientras cursaba el Magíster, el Dr. David Véliz me propuso hacer una pasantía con la Dr. Julie Turgeon en la Université Laval (Quebec, Canadá). Fue mi primera gran experiencia trabajando en ciencia en un país al otro lado del mundo, sin ningún familiar o amigo cerca y sin hablar el idioma. Al volver, comencé a trabajar como ayudante de proyecto con el Dr. Marco Méndez. Después de tres años decidí entrar al Doctorado en Ciencias con mención en Ecología y Biología Evolutiva de la Universidad de Chile, bajo la tutela de Marco. Intenté hacer lo mejor y lo más que pude: fui a todos los cursos sobre genómica que me aceptaron, intenté desarrollar análisis que no se habían hecho hasta el momento en Chile, hice una pasantía con un rock-star de la biología evolutiva como es el Dr. Ole Seehausen (una pasantía que iba a ser de 6 meses, pero terminó siendo de 11). Estoy orgullosa con la mayoría de lo que hice. Después de todos estos años de estudio, dedicación, aprendizaje y fascinación con la ciencia, me siento preparada para desenvolverme bien en cualquier desafío que me proponga. Y tengo muchas ideas que espero poder desarrollar, en Chile o donde me quieran recibir.

AGRADECIMIENTOS

Quisiera agradecer profundamente a todos aquellos que me ayudaron directa o indirectamente a desarrollar esta tesis:

A mi familia, a mis papás Alfredo y Maritza, a mis hermanos, Daniel y David, a mi tía Nena, a cuñadas/hermanas-putativas, Marisol y Eileen. No tengo palabras suficientes para agradecerles todo su amor incondicional.

A mi tutor, Dr. Marco A. Méndez, por su constante apoyo y confianza en mí, por las discusiones/peleas que todo estudiante de doctorado debe tener con su tutor. Espero que podamos continuar trabajando juntos investigando este interesante grupo de peces.

A mis amigos del laboratorio de Genética y Evolución, GEVOL, a los integrantes actuales y antiguos: Moisés Valladares, Franco Cruz-Jofré, Camilo Valdivieso, Paola Sáez, Pablo Fibla, Álvaro Zúñiga, Carolina Gallardo, Claudio Correa, Dayana Vásquez, Gonzalo Collado, Luis Pastenes, Hugo Salinas, Carolina Araya, Carla Rojas, Alejandra Fabres, José Serrano. Muchas gracias por todas las discusiones científicas y acerca de la vida y por todos los buenos momentos.

A quienes me ayudaron en terreno, a Hugo Salinas, Álvaro Zúñiga, Pablo Fibla, Franco Cruz-Jofré, y al profesor Michel Sallaberry. Muchísimas gracias, porque los terrenos fueron particularmente extenuantes.

Al Dr. Ole Seehausen, a FishEc y al instituto EAWAG (Suiza), por el inmenso apoyo que me entregaron, tanto en la discusión del tema general como con los análisis. A los integrantes de FishEc, a quienes agradezco especialmente: David Marques, Joana Meier, Ayana De Brito, Moritz Muschick, Philine Feulner, Mathew McGee, Jaime Anaya, Oliver Selz, Jacco Van Rijssel, Cas Retel, Florian Moser, Carmela Doenz, Miguel Leal, y por supuesto, a la maravillosa Salome Mwaiko. También quisiera agradecer la asistencia del Genetic Diversity Center en ETH Zurich. De EAWAG Kastanienbaum también

agradezco infinitamente a quienes me apoyaron moralmente: Typhaine Guillemot, Adrien Gaudard, Jonas Walker, Bernard Plachy, Kathrin Bross, Rohini Athavale, Pamela “Paloma” Gumpinger, Sergio “Checho” Di Michelangeli, Raquel Arias, Andrea del Pilar, Carolina Madeira, Stephanie Koch, Stephen Thomas, Nadine Czekalski y Kirsten Oswald.

A Franco Cruz-Jofré, por todas las discusiones acerca de *Orestias* que tuvimos. Quiero agradecer también al grupo mejicano-francés con el que trabajamos en el Muséum National d'Histoire Naturelle de París, Francia: Yareli Esquer-Garrigos, Philippe Gaubert y Bernard Hugueny y Thierry Oberdoff con quienes llevamos a cabo el proyecto de colaboración. ECOS-CONICYT C10802. Fue maravilloso encontrar más personas tan dedicadas a la comprensión de la evolución de *Orestias*.

A mis compañeros de generación del Doctorado, Nicolás Segovia y Esteban Botero, porque logramos formar un pequeño, pero muy buen grupo de estudio (y de apoyo moral).

A mis grandes amigas, Bárbara Núñez y Lorena Pizarro, porque sus consejos siempre me han ayudado.

A Paula Díaz, por salvarme de tantas situaciones difíciles.

A quienes me ayudaron en algún aspecto de esta tesis, ya sea por preguntas específicas o por la revisión de todos mis resultados: Natalia Lam, Cristian Canales, Daniel Gómez-Uchida, Robin Cristofari, Pablo Orozco-terWengel, Nicolas Puillandre y José Rogan.

Al comité de revisión de esta tesis, por sus valiosos comentarios que sin dudas mejoraron el producto final: Dr. David Véliz, Dr. Elie Poulin, Dra. Leyla Cárdenas y Dra. Alejandra González.

A los proyectos científicos que financiaron esta tesis, a los proyectos FONDECYT 1140540 y FONDECYT 1140543 que permitieron financiar los trabajos de terreno, y parte de los datos de RAD-Seq. Al Dr. Miguel Allende y al Centro de Regulación del Genoma FONDAP- CRG-1509007 que permitieron obtener el genoma de referencia de *O. ascotanensis*, el primer genoma completo de un pez endémico de Chile, y la otra parte de los datos de RAD-Seq. Y a la Beca de Doctorado Nacional otorgada por CONICYT, CONICYT-PCHA/doctorado Nacional/2012-21120972.

INDICE DE CONTENIDOS

RESUMEN.....	16
ABSTRACT.....	18
CAPÍTULO I. INTRODUCCIÓN GENERAL.....	20
1. GENÓMICA DE LA ESPECIACIÓN	21
1.1. El proceso continuo de la especiación.....	21
1.2 Genómica de la especiación	22
1.3 Peces como modelo de estudio de los procesos de especiación.	25
2. EL GÉNERO <i>ORESTIAS</i> (TELEOSTEI; CYPRINODONTIDAE).....	26
2.1 Estudio sistemático de las especies chilenas de <i>Orestias</i>	34
2.1.1 Morfología.	34
2.1.2 Cariotipos.....	37
2.1.3 Análisis filogenéticos.....	40
2.1.4 Dietas	42
2.1.5 Fauna bentónica y zooplancton.....	42
2.2 Perspectiva general del problema	43
2.3. Historia geológica de los hábitats de las especies chilenas de <i>Orestias</i>	46
2.4. Pares de especies a comparar en este estudio.	48
HIPÓTESIS.....	49
OBJETIVOS.	50

RESULTADOS.....	51
CAPITULO II. ESTRUCTURA GENÉTICA EN CUATRO ESPECIES DEL GÉNERO <i>ORESTIAS</i>	52
1. Abstract	53
2. Introduction	54
3. Methods.....	55
3.1 Sampling, DNA extraction, sequencing of mtDNA control region and RAD-Seq.....	55
3.2 Phylogeny and phylogeography of mtDNA control region sequences.	59
3.3 Data preparation and variant, genotype and SNP calling.....	60
3.4 RAxML.....	61
3.5 Principal Component Analysis (PCA) between species.....	62
3.6 Hierarchical Structure Analysis.	62
3.7 Pairwise F_{ST} between species.....	63
4. Results	63
4.1 Phylogeny and phylogeography of Chilean <i>Orestias</i>	63
4.2 Genomic differentiation in pairs of <i>Orestias</i> species.....	66
5. Discussion.	70
6. Supporting Information.....	73
CAPITULO III. ANÁLISIS COMPARATIVO DE LOS PATRONES DE DIFERENCIACIÓN GENÓMICA EN CUATRO ESPECIES DE <i>ORESTIAS</i>	76
1. Resumen.....	77
2. Introducción	78
3. Metodología	79
3.1 Estimación de los valores de F_{ST} global y F_{ST} por cada sitio.	80
3.2 Estimación de SNPs y categorías de Gene Ontology comunes y únicas por especie.....	80
3.3 Detección de sitios más diferenciados entre especies.	81

4.	Resultados	82
4.1	Valores de F_{ST} a distintos niveles de diferenciación.	82
4.2	Número de SNPs y categorías GO comunes y únicos por especie.....	87
4.3	Identificación de los SNPs de mayor contribución a la diferenciación entre especies	91
5.	Discusión.....	93
6.	Material Suplementario	97
DISCUSIÓN GENERAL.....		101
1.	Aportes a la comprensión del proceso de especiación en <i>Orestias</i>	101
2.	Crítica a la utilización de RAD-Seq en estudios de especiación.	103
BIBLIOGRAFIA.....		106

LISTA DE TABLAS

Tabla 1. Especies de <i>Orestias</i> del complejo <i>agassii</i> (<i>sensu</i> Parenti 1984a) y su ubicación geográfica. ID corresponde al número identificador que aparece en la Fig. 1	29
Tabla 2. Análisis discriminante realizado a partir de los resultados del análisis de componentes principales de las mediciones de morfología lineal de las especies chilenas de <i>Orestias</i> . Se indican el número de individuos asignados a cada especie y los porcentajes de clasificación correcta. Modificado de Vila et al. (2010).	35
Tabla 3. Fórmula cromosómica de las especies chilenas de <i>Orestias</i> (faltan los datos de <i>O. cf. agassii</i> del río Collacagua, cuyo fórmula aún no ha sido descrita). 2n: número cromosómico diploide; NF: número fundamental de brazos; M: cromosomas metacéntricos; SM: cromosomas submetacéntricos; ST: cromosomas subteloicéntricos; T: cromosomas telocéntricos; Mc: microcromosomas. *: descrito en Vila <i>et al.</i> (2011); **: mencionado en Araya (2009).	38
Tabla 4. Linajes cromosómicos en las especies chilenas de <i>Orestias</i> . Se indican los cromosomas con bandeo C, que indica la heterocromatina constitutiva, la ubicación del cluster ribosomal mayor (28S) y del cluster ribosomal menor (5S). Se especifica la morfología del par cromosómico 1 (metacéntrico) para todas las especies, y luego los pares cromosómicos que contienen los cluster ribosomales, diferenciándolos en submetacéntricos y telocéntricos. Para las especies <i>O. parinacotensis</i> , <i>O. piacotensis</i> y <i>O. gloriae</i> sólo se han obtenido datos de la heterocromatina constitutiva. Modificado de Araya (2009). ..	39
Tabla 5. Resumen de los caracteres analizados en las especies chilenas de <i>Orestias</i> . *: estudios realizados sólo para las especies del PNL.	45

Tabla 6. Sampled sites in the Chilean distribution of <i>Orestias</i> that were included in the phylogenetic analysis, including the seven type localities and the sixteen recently discovered sites. a: Type locality; b: recently discovered site. * indicates the sample size for mtDNA control region sequences and then for RAD-Seq, respectively.....	58
Tabla 7. Valores de F_{ST} globales y resumen de los resultados de los valores de F_{ST} locus por locus, en cada par especies de <i>Orestias</i> . N: tamaño muestral. ^a : p-value < 0.05. ^b : entre paréntesis se indica el porcentaje de loci con respecto al total de loci. ^c : se aplicó la corrección FDR para estimar el número de loci con F_{ST} significativo.....	84
Tabla 8. Detalle del número de SNPs por especies, número de SNPs que más contribuyen a la diferenciación de las especies, y el número de SNPs que más contribuyen a la diferenciación y que son únicos y comunes entre set de datos. Se indican los SNPs que se encuentran en genes anotados (y los que no) en el genoma de referencia, número de los genes anotados que tienen un ID conocido (y los que no) en la base de datos de PANTHER y el número de categorías GO enriquecidas que se obtienen a partir de éstos.....	90
Tabla 9. Número de categorías GO enriquecidas comunes entre especies de Procesos Biológicos (bajo la diagonal principal) y Función Molecular (sobre la diagonal principal).	91
Tabla 10. Resultados y conclusiones obtenidas en esta tesis luego de poner a prueba cada hipótesis planteada.	105

LISTA DE FIGURAS

- Figura 1.** Proceso continuo de especiación. Se indican tres estados o etapas de divergencia diferentes de este proceso, determinados por diferentes medidas que cuantifican la divergencia: aislamiento reproductivo, agrupamiento genotípico o distribución de las frecuencias génicas de cada población y segregación de linajes. Al comienzo de este proceso no existe aislamiento reproductivo o es muy escaso, la distribución de las frecuencias génicas en los individuos muestreados en ambas poblaciones o agrupamiento genotípico es unimodal y no existen linajes diferenciados. En una etapa intermedia, se puede observar cierto grado de diferenciación entre grupos y comienzan a distinguirse grupos genotípicos diferentes. Al final del proceso de especiación, se establece el aislamiento reproductivo entre grupos divergentes, los cuales están caracterizados por frecuencias alélicas diferentes, que se observan en una distribución bimodal de éstas y hay una marcada segregación de los linajes. Modificado de Nosil *et al.* (2009b). 22
- Figura 2.** Distribución de las 27 especies de *Orestias* del complejo *agassii*, además de las poblaciones que han sido asignadas a *O. agassii* (*O. cf. agassii*) y las poblaciones de *Orestias* sin estatus taxonómico establecido (*Orestias* sp.). En la esquina superior derecha se muestra la región de Titicaca ampliada. El detalle de las especies y poblaciones comprendidas en cada punto se presentan en la Tabla 1. 28
- Figura 3.** Localidades donde han sido descritas las especies de *Orestias* en Chile. En el rectángulo azul se indican los sitios que se encuentran dentro del Parque Nacional Lauca. 33
- Figura 4.** Análisis de variables canónicas de datos de morfometría geométrica basado en 15 landmarks para la comparación por especie. Se grafican las dos primeras variables canónicas (CV1 y CV2), en las

que se incluyen los diagramas de cambio para los valores -4 y 4 en ambos ejes: en gris se indica la forma de consenso para todos los individuos y en azul, la forma ponderada. Se indican los individuos de las especies que se distinguen en este análisis: *O. gloriae* (Carcote, en naranja), *O. ascotanensis* (Ascotán, en amarillo) y *O. cf. agassii* de Isluga (en rojo). (F. Cruz-Jofré, comunicación personal). 36

Figura 5. Reconstrucción filogenética (máxima verosimilitud e inferencia bayesiana) de las especies chilenas de *Orestias* realizada con tres marcadores mitocondriales (región control, cytb y ND2; 3035 pb en total). Cada especie descrita forma un grupo monofilético, excepto las del PNL que quedan reunidas dentro de un mismo grupo monofilético (indicado como “linaje PNL” con la barra gris), dentro del cual no se recupera la monofilia de ninguna de estas especies. En amarillo se indica el grupo monofilético de *O. ascotanensis* (salar Ascotán). En naranja, se observa *O. gloriae* (salar Carcote). En azul y verde se indican las secuencias de *O. chungarensis* (lago Chungará) y *O. parinacotensis* (bofedal Parinacota), respectivamente. Los valores en los nodos corresponden a los valores de bootstrap >70% (indicado con un ●) y la probabilidad a posterior >0,95 (indicado con un ◇). Modificado de Morales *et al.* (2013)..... 41

Figura 6. Reconstrucción del nivel más alto que habría alcanzado el paleolago Tauca (3.770 m sobre el nivel del mar). Se indica la ubicación del salar Carcote, que habría formado parte de este lago, mientras que el salar Ascotán (al sur del salar Carcote) no habría sido inundado por este paleolago. Modificado de Blard *et al.* (2011). 47

Figura 7. a) Sampled localities in the Lauca National Park (LNP) and nearby sites. b) Ascotán salt pan and the twelve springs with *O. ascotanensis* populations. c) Carcote salt pan and the three springs with *O. gloriae* populations. The red circles indicate the populations used for genomic analysis. 57

Figura 8. Maximum likelihood phylogenetic reconstruction for the mtDNA control region sequences of Chilean populations of *Orestias*. Bootstrap values >70% (ML) and posterior probabilities (BI) >0.95 of each node is indicated (* : no support). In a) it is shown the ingroup, presenting the three main collapsed clades: b) the CHAL clade, c) the *O. gloriae* clade, and d) the LNP clade. b), c), and d) are schematic representations of each of these clades, showing the close relationship between the RAD individuals and

each of their own species (for simplicity, there are not indicated any other species or population within each of these clades; see the Figure S1 for the complete uncollapsed clades). 65

Figura 9. Haplotype networks of the control region mtDNA sequences of a) *O. ascotanensis*, b) *O. gloriae* and c) *O. chungarensis* and *O. laucaensis* along all the LNP populations. 67

Figura 10. RAxML phylogenetic reconstruction of the RAD individuals, comprising *O. ascotanensis* (blue), *O. gloriae* (green), *O. chungarensis* (yellow) and *O. laucaensis* (red). 68

Figura 11. Principal Component Analysis (PCA) for the RAD individuals of the four species of *Orestias*. Variance explained by each principal component (PC) is indicated in parenthesis. a) PC1 versus PC2, and b) PC2 versus PC3. 69

Figura 12. Hierarchical *structure* analysis, showing the number of genetic cluster at each step of the analysis. Each individual is represented as a bar and the different color represents the assignment probability to different clusters. First, the complete dataset was analyzed, founding 3 clusters. Then, the genetic cluster comprising the LPN species was analyzed, obtaining a best grouping of 2, where each cluster was formed by each species. 70

Figura 13. Distribuciones de los valores de F_{ST} entre especies de *Orestias* a) entre las especies del PNL, que recientemente divergieron, *O. laucaensis* y *O. chungarensis*, y b) entre las especies de los salares, que presentan un largo tiempo de divergencia, *O. ascotanensis* y *O. gloriae*. También se indican las distribuciones de los valores de F_{ST} significativos (después de aplicar la corrección de FDR) c) entre especies del PNL y d) entre especies de los salares. Las líneas punteadas (rojas y negras) indican la mediana de cada distribución: en a) es 0,00095, en b) es 0,07, en c) es 0,26y en d) es 0,52. 85

Figura 14. Niveles de diferenciación para cada locus en cada par de especies de *Orestias*. En a) se indican los valores de F_{ST} de cada locus entre las especies del PNL, *O. chungarensis* y *O. laucaensis*, que han divergido recientemente, y en b) entre las especies de los salares, *O. ascotanensis* y *O. gloriae*, con mayor tiempo de divergencia. En el eje x se indican las posiciones de cada locus dentro del scaffold al que pertenecen (indicados en gris claro y gris oscuro para diferenciar scaffolds adyacentes), y en el eje y se

indican los valores de F_{ST} para cada locus. Los puntos negros indican los loci con valores de F_{ST} con un valor de $p > 0,05$ y los puntos rojos indican los loci con valores de F_{ST} significativos al 5%, después de la corrección de FDR. 86

Figura 15. a) Número de sitios (SNPs) totales por cada especie. b) y c) Diagramas de Venn relacionando el número de SNPs de b) las especies de los salares y c) las especies del PNL. En la intersección de cada diagrama se indica el número de sitios que son comunes entre especies y el porcentaje que representa, considerando el total de sitios como la suma de SNPs de ambas especies..... 87

Figura 16. Diagrama de Venn que relaciona el número de SNPs que más contribuyen a la diferenciación entre especies recientes y divergentes. 92

MATERIAL SUPLEMENTARIO

Figure S1. a) The *O. gloriæ* clade. The RAD individuals of *O. gloriæ* are shown in green. b) LNP clade comprising the populations from the Lauca National Park; the RAD individuals of *O. chungarensis* are shown in yellow, while the RAD individuals of *O. laucaensis* are shown in red. c) CHAL clade including the populations from Collacagua river, Huasco salt pan, Ascotán salt pan and Lirima wetland; the RAD individuals of *O. ascotanensis* are shown in blue. 73

Tabla MS 1. Categorías GO enriquecidas de Proceso Biológico por especie. Se indican las categorías compartidas por pares de especies: ^a: *O. ascotanensis* y *O. gloriæ*; ^b: *O. ascotanensis* y *O. chungarensis*; ^c: *O. ascotanensis* y *O. laucaensis*; ^d: *O. gloriæ* y *O. laucaensis*. *O. chungarensis* no comparte categorías enriquecidas con *O. laucaensis* ni con *O. gloriæ*. 97

Tabla MS 2. Categorías GO enriquecidas de Función Molecular por cada una de las especies. Se indican las categorías compartidas por pares de especies: ^a: *O. ascotanensis* y *O. gloriæ*; ^b: *O. ascotanensis* y *O. laucaensis*; ^c: *O. gloriæ* y *O. laucaensis*. *O. chungarensis* no comparte categorías enriquecidas con ninguna de las otras tres especies. 99

Tabla MS 3. Categorías GO enriquecidas detectadas a partir del 10% de loci que más contribuyen a la diferenciación de las especies en tres set de datos: todas las cuatro especies, las especies antiguas (de los salares) y especies recientes (del PNL). 100

Tabla MS 4. Categorías GO enriquecidas de Procesos Biológicos detectadas a partir de los 769 SNPs que sólo diferencian a las especies recientes. Ninguna otra categoría presentó subcategorías enriquecidas. Ninguno de los otros dos set de datos presentó categorías enriquecidas. 100

RESUMEN

Durante el proceso continuo de la especiación se genera la divergencia genética y el establecimiento del aislamiento reproductivo. La descripción de los patrones genéticos de diferenciación entre pares de taxa cercanamente relacionados en diferentes etapas de este continuo podría ayudar a determinar la proporción del genoma que contribuye a la divergencia y la naturaleza de los genes involucrados. En el contexto de especiación alopátrica, se espera que la magnitud del primer aspecto sea proporcional al tiempo de divergencia, mientras que la deriva génica debería hacer aparecer mutaciones al azar en el genoma, afectando a diferentes regiones génicas e intergénicas en diferentes etapas del continuo de especiación.

En esta tesis se describen los patrones de divergencia genómica entre dos pares de especies chilenas del género *Orestias*, pupfishes que habitan el Altiplano de Chile, Perú y Bolivia, que se encuentran en etapas diferentes del continuo de especiación. En una etapa inicial de este proceso se encuentran *O. chungarensis* y *O. laucaensis*, ambas presentes en ambientes aislados (Lago Chungará y Río Lauca, respectivamente). Por otra parte, en una etapa tardía se encuentran *O. ascotanensis* y *O. gloriae*, quienes habitan en vertientes de salares cercanos, pero desconectados (salar Ascotán y salar Carcote, respectivamente). Por una parte, debiera existir mayor diferenciación genómica entre las especies de la etapa más avanzada que entre las especies de la etapa más reciente. Por otra parte, y dado que estas especies se originaron en un contexto de especiación alopátrica que se ha mantenido hasta el presente, los patrones de divergencia genómica en cada par de especies debieran haber seguido rutas independientes.

Se aplicó la técnica RAD-Seq, un tipo de secuenciación genómica de representación reducida, a los individuos muestreados de las cuatro especies. Los análisis de estructuración genética detectaron una

fuerte divergencia entre las especies de los salares y entre éstas y *O. chungarensis* y *O. laucaensis*, y una divergencia mucho menor entre éstas últimas. Los niveles de diferenciación global, medidos con el índice F_{ST} , indicaron que las especies recientes se han diferenciado tres veces menos que las especies más divergentes. Además se observó que ~20% de los loci totales se diferencia entre las especies recientes, mientras que esa cantidad aumenta a ~50% entre especies divergentes. Estos loci no estarían concentrados en ninguna región en particular, sino que se encontrarían distribuidos a lo largo de todo el genoma. Los análisis del número total de SNPs y de SNPs que más diferencian a las especies indicaron que estos polimorfismos son particulares de cada especie al igual que las funciones biológicas en las que están involucrados. Estos resultados permitieron observar empíricamente cómo el grado de divergencia a nivel genómico aumenta a medida que se avanza en el continuo de especiación, tanto a nivel de diferenciación global, como de la diferenciación de cada locus, y que el proceso de diferenciación ha seguido un camino independiente en cada una de las especies y pares de especies, lo cual es concordantes con un modelo de especiación alopátrica.

ABSTRACT

During the continuum process of speciation the genetic diversity is generated and the reproductive isolation is established. The description of genomic patterns of differentiation from pairs of closely related taxa at different stages of this continuum would help identify the proportion of the genome that contributes to the divergence and the nature of the genes involved. In an allopatric speciation context, it is expected that the magnitude of the first aspect is proportional to divergence time, while the genetic drift would give rise mutations randomly in the genome affecting therefore different genic and intergenic regions at different stages of the speciation continuum.

This study described the genomic patterns of divergence between two pairs of Chilean species of the genus *Orestias* at different stages of the speciation continuum. An initial stage involves *O. chungarensis* and *O. laucaensis*, both inhabiting isolated environments (Lake Chungara and Lauca River, respectively). On the other hand, *O. ascotanensis* and *O. glorieae* represent a late stage of this continuum. They both inhabit close, unconnected salt pans (Ascotan and Carcote salt pan, respectively). On one hand, there should be a higher genomic differentiation between species of the late stage than species of the recent stage. On the other hand, and given these species were originated in an allopatric speciation context that persist until today, then the patterns of genomic divergence of each species pair should have follow different and independent paths.

We obtained RAD-Seq data, a reduced representation sequencing technique, from individuals of each of these species. Genetic structure analyses found a deep divergence between salt pans samples and between these and *O. chungarensis* and *O. laucaensis* samples, and much less divergence between these last two.

Overall F_{ST} values, as a measure of genetic differentiation, are three times higher between the distant species than the close related pair of species. Moreover, ~20% of the loci are differentiated between *O. chungarensis* and *O. laucaensis*, while ~50% of the total loci are differentiated between the distant species, and these loci are not concentrated in any specific region, but distributed along the whole genome. Analyses of the total number of SNPs and the SNPs that more differentiate the species indicate that the polymorphisms are particular of each species, as well as the biological functions they are associated with. These results allowed to empirically observing how the genomic divergence increase as the speciation continuum advance, at both overall differentiation and differentiation of locus-by-locus, and that the differentiation process has followed an independent path in each of species and species pairs, in concordance with an allopatric speciation model.

CAPÍTULO I.

INTRODUCCIÓN GENERAL

1. GENÓMICA DE LA ESPECIACIÓN

1.1. El proceso continuo de la especiación

La especiación es un proceso que se ha entendido como un continuo durante el cual se establece la divergencia genética y el aislamiento reproductivo (Nosil & Feder 2012a). La comprensión de este proceso ha sido uno de los problemas más fascinantes de la biología evolutiva y que más controversia ha generado (Hey 2001; Barton 2001). En la actualidad se han comenzado a indagar preguntas tales como cuál es la importancia relativa de distintas fuerzas evolutivas como selección natural, mutación, flujo génico, entre otros, en la generación de divergencia y en el proceso de especiación (Butlin *et al.* 2012; Nosil & Feder 2012b).

Una dificultad obvia del estudio de la divergencia y especiación en la mayoría de las poblaciones naturales de organismos multicelulares de reproducción sexual es que este proceso ocurre a una escala de tiempo muy larga para la observación directa (Seehausen *et al.* 2014). Como estrategia de investigación para el estudio de este tipo de organismos se ha propuesto analizar pares de especies cercanamente relacionadas que presenten diferentes grados de divergencia a lo largo de este continuo (Fig. 1) (Nadeau *et al.* 2012; Nosil & Feder 2013). De esta manera, el estudio de los estados intermedios de divergencia permitiría determinar la importancia relativa de los diferentes factores que pueden estar provocando el aislamiento reproductivo en esas etapas, ya sea mediante el análisis de un locus específico o realizando un análisis al genoma completo (Chapman *et al.* 2013). Esto permitiría inferir, de manera indirecta, el proceso de divergencia genómica y de especiación que ocurre en los taxa a lo largo del tiempo (Nosil & Feder 2013).

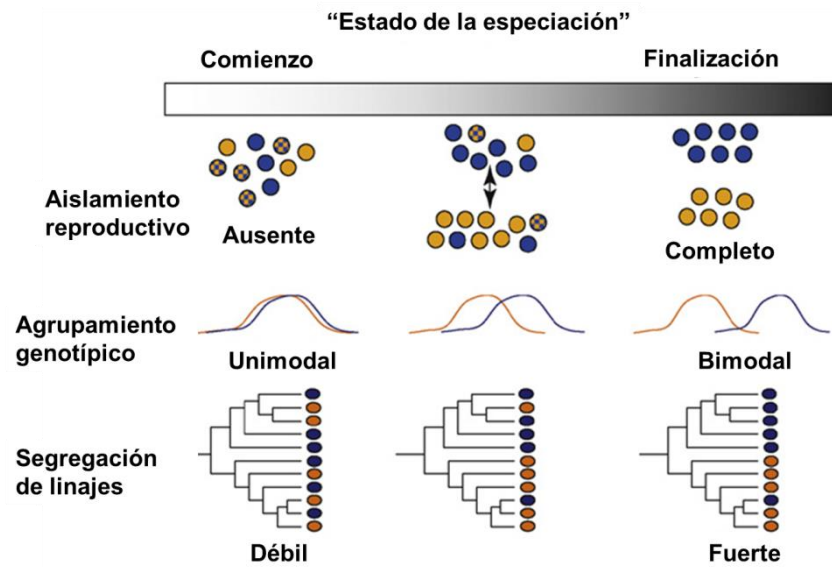


Figura 1. Proceso continuo de especiación. Se indican tres estados o etapas de divergencia diferentes de este proceso, determinados por diferentes medidas que cuantifican la divergencia: aislamiento reproductivo, agrupamiento genotípico o distribución de las frecuencias génicas de cada población y segregación de linajes. Al comienzo de este proceso no existe aislamiento reproductivo o es muy escaso, la distribución de las frecuencias génicas en los individuos muestreados en ambas poblaciones o agrupamiento genotípico es unimodal y no existen linajes diferenciados. En una etapa intermedia, se puede observar cierto grado de diferenciación entre grupos y comienzan a distinguirse grupos genotípicos diferentes. Al final del proceso de especiación, se establece el aislamiento reproductivo entre grupos divergentes, los cuales están caracterizados por frecuencias alélicas diferentes, que se observan en una distribución bimodal de éstas y hay una marcada segregación de los linajes. Modificado de Nosil *et al.* (2009b).

1.2 Genómica de la especiación

Los trabajos que han abordado el proceso continuo de la especiación desde un punto de vista genético han estado limitados al uso de las herramientas moleculares disponibles que han permitido el análisis de un

número restringido de marcadores moleculares (Mallet *et al.* 2007; Mallet 2008; Hendry 2009). En la actualidad, el desarrollo de técnicas de secuenciación masiva o Next-Generation Sequencing (NGS) (Ellegren 2008; Metzker 2010), han permitido una exploración mucho más detallada de la divergencia, haciendo posible observarla a lo largo de todo el genoma y no sólo en algunos genes (Nadeau *et al.* 2012; Jones *et al.* 2012; ROESTI *et al.* 2012; Andrew & Rieseberg 2013; Gagnaire *et al.* 2013; Powell *et al.* 2013; Martin *et al.* 2013; Seehausen *et al.* 2014). Estas técnicas han sido aplicadas al estudio de la especiación, desarrollándose una nueva área de investigación: la genómica de la especiación (Seehausen *et al.* 2014). La genómica de la especiación intenta comprender la influencia de las propiedades genómicas en la evolución de las barreras reproductivas, además de la relación entre distintos procesos evolutivos y los patrones de divergencia genómica, indagando en las causas ecológicas y no ecológicas de la especiación (Seehausen *et al.* 2014). Las investigaciones realizadas han estado dirigidas a responder preguntas tales como ¿dónde se genera la divergencia, en zonas localizadas o está diseminada a lo largo del genoma?, ¿la divergencia se genera por procesos selectivos o variaciones azarosas?, ¿la divergencia genómica sigue una misma trayectoria a medida que avanza en el proceso de especiación?, ¿cuál es el efecto de distintos grados de aislamiento geográfico sobre la generación de la divergencia genómica? (Seehausen *et al.* 2014).

Los principales actores involucrados en la generación de divergencia genómica son flujo génico, selección divergente o diversificadora y recombinación. Los dos primeros tendrían un efecto opuesto, ya que la divergencia se vería favorecida por la selección, mientras que el flujo génico tendría un efecto homogeneizador de los acervos genéticos en diferentes poblaciones (Nosil *et al.* 2009a). La recombinación, por otra parte, causaría que distintas partes del genoma muestren diferentes historias (Sousa & Hey 2013). Por lo tanto, alelos de loci neutrales podrían encontrarse a lo largo de todo el rango de distribución de la especie a una misma frecuencia alélica por efecto del flujo génico y la recombinación que permiten el traspaso del material genético entre poblaciones (Wu 2001; Sousa & Hey 2013). Por otra parte, aquellos alelos que confieren mayor fitness en un cierto tipo de ambiente, aumentarían en frecuencia en la población que habita ese ambiente, y presentará altos niveles de divergencia con otra población donde ese mismo alelo no esté siendo seleccionado a favor. Las regiones que flanquean el loci bajo

selección, y que están ligadas a él, también presentarán altos niveles de divergencia, aunque sean loci neutrales, debido al efecto de hitchhiking. La única manera de revertir este efecto es mediante recombinación en esa región, rompiendo la asociación entre el alelo bajo selección y el alelo neutral. Por lo tanto, en dos poblaciones habitando ambientes diferentes y en presencia de flujo génico, pueden encontrarse, coexistiendo en un mismo genoma, loci neutrales que mostrarán frecuencias alélicas similares entre poblaciones, loci bajo selección divergente y loci neutrales con altos niveles de divergencia porque están ligados a los anteriores (Barton 2000; Sousa & Hey 2013).

Hasta el momento, la mayoría de los estudios que han investigado los aportes de estos tres procesos a la divergencia genómica se han enfocado en el estudio de la especiación ecológica de especies incipientes que habitan en simpatría (Rice & Hostert 1993; Smadja & Butlin 2011). Los escaneos genómicos realizados han mostrado que la divergencia es altamente heterogénea. Por una parte, la selección produciría “islas genómicas de divergencia” entre poblaciones (Wu 2001; Turner *et al.* 2005; Harr 2006; Nosil *et al.* 2009a; Feder *et al.* 2012). Una isla genómica de divergencia se define como una región del genoma, de cualquier tamaño (puede ser un nucleótido o un cromosoma completo), cuya divergencia excede lo esperado por neutralidad (Nosil & Feder 2012a). Estas zonas divergentes pueden contener genes que contribuyen al aislamiento reproductivo (Nosil & Feder 2012a): se ha observado que en estas zonas residen los genes que causan disfunciones en híbridos de *Drosophila* (McGaugh & Noor 2012), genes que impiden la formación de híbridos entre dos subespecies de *Mus musculus* (Harr 2006), y que contienen loci involucrados en adaptación divergente, que podrían ser la causa de la selección en contra de migrantes e híbridos de mariposas del género *Heliconius* (Nadeau *et al.* 2012; Renaut *et al.* 2012). Por otra parte, en los escaneos genómicos también se han detectado zonas no diferenciadas debido al flujo génico presente y la introgresión de alelos neutrales. Si se llegase a establecer un aislamiento reproductivo suficientemente fuerte para permitir la persistencia de especies incipientes en simpatría, entonces el resto del genoma comenzaría a diferenciarse, a pesar de la existencia de flujo génico (Lawniczak *et al.* 2010; Jones *et al.* 2012; Gagnaire *et al.* 2013; Seehausen *et al.* 2014).

1.3 Peces como modelo de estudio de los procesos de especiación.

Se ha estimado que existen más de 28.000 especies de peces, las cuales habitan en prácticamente todo tipo de ambiente acuático alrededor del mundo, y exhiben una enorme diversidad en morfología y biología (Nelson 2006). Estas características han convertido a los peces en el grupo con mayor número de especies y el más diversificado dentro de los vertebrados (Nelson 2006), por lo que es posible encontrar variados escenarios de diversificación. Por estas razones, los peces son los organismos acuáticos más utilizados para el estudio de los procesos de especiación (Bernardi 2013). Un modelo de estudio clásico en esta materia han sido los peces cíclidos que habitan los grandes lagos africanos. Por ejemplo, en el lago Victoria, éstos habrían pasado por un proceso de radiación intenso hace tan sólo 17-12 kya (Johnson *et al.* 1996), el cual habría generado la enorme riqueza de especies que es posible observar en la actualidad (Turner *et al.* 2001). Para explicar cómo se ha generado toda esta variabilidad se han propuesto diversos mecanismos (Kornfield & Smith 2000; Salzburger & Meyer 2004), como especiación alopátrica relacionada con variaciones en los niveles lacustres (Sturmbauer & Meyer 1992; Sturmbauer *et al.* 2001; Verheyen *et al.* 2003; Joyce *et al.* 2005), especiación ecológica relacionada con especializaciones a un tipo de hábitat particular (Salzburger *et al.* 2002), y divergencia por selección sexual relacionada con sesgo sensorial en las hembras y preferencia por una coloración corporal particular de los machos (Seehausen *et al.* 2008; Miyagi *et al.* 2012).

Estos taxa también están siendo usados en el estudio de la genómica de la especiación. Las especies más utilizadas hasta el momento han sido *Coregonus clupeaformis* (whitefish) (St-Cyr *et al.* 2008; Renaut *et al.* 2010, 2012; Bernatchez *et al.* 2010; Gagnaire *et al.* 2013) y *Gasterosteus aculeatus* (threespine stickleback) (Hohenlohe *et al.* 2010; Jones *et al.* 2012; Roesti *et al.* 2012; Hendry *et al.* 2013; Feulner *et al.* 2013; Roesti *et al.* 2013), además de especies de *Xiphophorus* (Schartl *et al.* 2013; Jones *et al.* 2013; Cui *et al.* 2013), cíclidos africanos (Baldo *et al.* 2011; Keller *et al.* 2013) y los cíclidos del complejo Midas de Nicaragua (Elmer *et al.* 2010; Henning *et al.* 2013; Recknagel *et al.* 2013; Bradic *et al.* 2013; Franchini *et al.* 2014).

En esta tesis se utilizó a los peces del género *Orestias* como modelo de estudio en genómica de la especiación, porque entre las especies de este género existen distintos grados de diferenciación lo que representa distintas etapas dentro del continuo de especiación, probablemente debido a distintos escenarios de diversificación. A continuación se presenta este género y se detallan estas etapas de diferenciación.

2. EL GÉNERO ORESTIAS (TELEOSTEI; CYPRINODONTIDAE)

El género *Orestias* es un grupo de peces pequeños de aguas continentales del Altiplano sudamericano. Forma parte de la familia Cyprinodontidae conocidos comúnmente como *pupfish*, la que se encuentra dentro del orden Cyprinodontiformes, también conocidos como *killifish*.

La historia del género *Orestias* Valenciennes 1846 se remonta a 80 a 100 millones de años atrás, cuando el ancestro de la familia Cyprinodontidae, a la que pertenece *Orestias*, habría colonizado Sudamérica. Esto habría ocurrido durante el período en que el nivel del mar alcanzó su máximo nivel durante el Cretácico, lo que permitió la formación de mares epicontinentales que cubrieron extensas zonas de África y Norteamérica y gran parte del norte de Sudamérica, hasta el lado este de la Cordillera de los Andes (Willis 1910; Haq *et al.* 1987; Barron & Peterson 1989). Posteriormente, los cambios asociados a la orogénesis de Los Andes habrían provocado modificaciones en las cuencas, generando que pequeños ríos y cuerpos de agua hayan quedado capturados, hasta finalmente quedar aislados hace 15 millones de años atrás (Vandervoort *et al.* 1995). De esta manera, se habría producido la dispersión de estos peces hacia el sur de lo que es hoy el lago Titicaca, y el género *Orestias* habría quedado restringido al Altiplano de Sudamérica (Bolivia, Perú y Chile).

Actualmente hay 46 especies descritas en este género: Parenti (1984a) describió 43 especies en la revisión más acabada que se ha hecho del género (para comparaciones de revisiones previas, ver Esquer Garrigos *et al.* 2013 y Takahashi & Moreno 2015; ver también Loubens (1989) quien plantea una posible

sinonimización de dos especies, *O. farfani* y *O. rotundipinnis*, con *O. lutea*¹). Posteriormente, Vila et al. (1986, 2006 & 2011) contribuyeron con la descripción de tres especies chilenas más: *O. chungarensis* (Vila & Pinto 1986), *O. piacotensis* (Vila 2006) y *O. gloriae* (Vila et al. 2011).

Las especies de *Orestias* han sido clasificadas en cuatro complejos de acuerdo a sus similitudes morfológicas (Parenti 1984a): (i) *mulleri* (5 especies), (ii) *gilsoni* (10 especies), (iii) *cuvieri* (4 especies), y (iv) *agassii* (25 especies). Los primeros tres complejos son endémicos del Lago Titicaca y forman un grupo polifilético. Los complejos *gilsoni* y *mulleri* forman un grupo monofilético y representa un clásico ejemplo de *species flock* (Parenti 1984b). El complejo *agassii* es el único complejo que contiene especies que habitan fuera del Lago Titicaca (Figura 2), distribuidas desde el Lago Llacsha, Perú (*O. hardini*; Parenti 1984a) hasta el bodefal Chalviri en Bolivia (*O. cf. agassii*; Esquer Garrigos 2013). 14 de las 25 especies están exclusivamente fuera del Lago Titicaca, y sólo dos, *O. agassii* y *O. lutea* están presentes tanto en el lago como fuera de éste. *O. agassii* es la especie que presenta la mayor distribución del género, desde las cuencas de Urubamba y del Titicaca en Perú hasta el río Isluga en Chile (Parenti 1984a; Cruz-Jofré et al. 2014). Una de las características más llamativas de esta especie es la alta diversidad morfológica que es posible encontrar en ella. Lauzanne (1982, 1991) recomienda que no se realicen distinciones de subespecies o variedades dentro de esta especie basándose sólo en caracteres morfológicos, porque se podría llegar a un extremo de “crear tantas especies como especímenes existentes” (Lauzanne 1982).

¹ En esta tesis se considerarán los nombres de las especies corregidos según el trabajo de Cruz-Jofré et al. (2013)

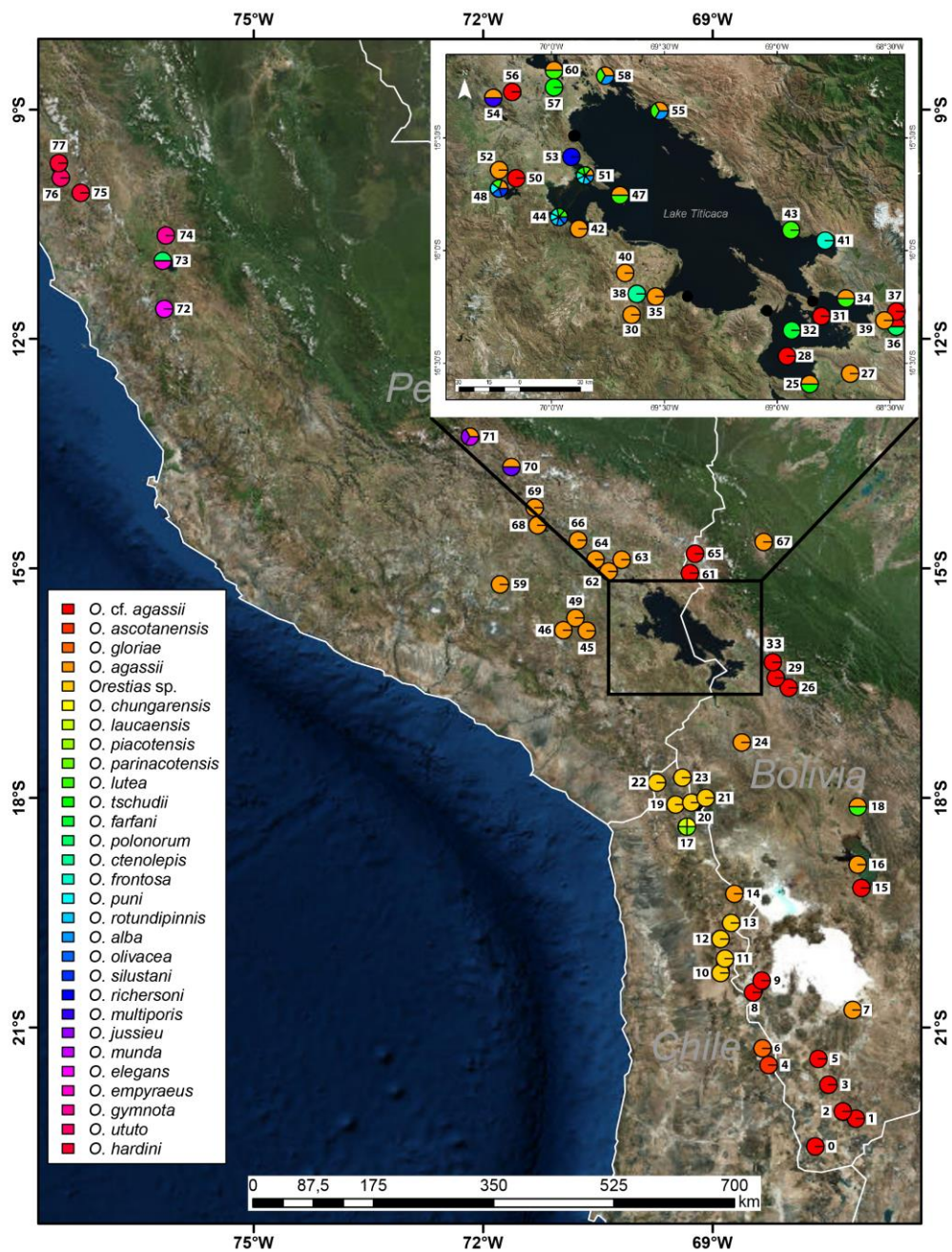


Figura 2. Distribución de las 27 especies de *Orestias* del complejo *agassii*, además de las poblaciones que han sido asignadas a *O. agassii* (*O. cf. agassii*) y las poblaciones de *Orestias* sin estatus taxonómico establecido (*Orestias sp.*). En la esquina superior derecha se muestra la región de Titicaca ampliada. El detalle de las especies y poblaciones comprendidas en cada punto se presentan en la Tabla 1.

Tabla 1. Especies de *Orestias* del complejo *agassii* (*sensu* Parenti 1984a) y su ubicación geográfica. ID corresponde al número identificador que aparece en la Fig. 1

ID	Especies	Localidad	País	Coordenadas geográficas
0	<i>O. cf. agassii</i>	Bofedal Chalviri	Bolivia	22° 33' 55"S, 67° 39' 24"W
1	<i>O. cf. agassii</i>	Lago Chipapa, bofedales Celeste y Celeste Este	Bolivia	22° 11' 36"S, 67° 06' 40"W
2	<i>O. cf. agassii</i>	Río Sol de Mañana	Bolivia	22° 07' 40"S, 67° 16' 56"W
3	<i>O. cf. agassii</i>	Río Villa Mar - Sur Lipez	Bolivia	21° 45' 11"S, 67° 29' 34"W
4	<i>O. ascotanensis</i>	Salar Ascotán	Chile	21° 29' 21"S, 68° 15' 25"W
5	<i>O. cf. agassii</i>	Bofedal Villa Alota	Bolivia	21° 24' 01"S, 67° 37' 21"W
6	<i>O. gloriae</i>	Salar Carcote	Chile	21° 16' 46"S, 68° 19' 21"W
7	<i>O. agassii</i>	Río Grande de Lipez	Bolivia	20° 46' 33"S, 67° 09' 59"W
8	<i>O. cf. agassii</i>	Río Empexa	Bolivia	20° 31' 01"S, 68° 26' 03"W
9	<i>O. cf. agassii</i>	Río Barras	Bolivia	20° 25' 13"S, 68° 21' 00"W
10	<i>Orestias sp.</i>	Salar Huasco	Chile	20° 15' 46"S, 68° 52' 32"W
11	<i>Orestias sp.</i>	Río Collacagua	Chile	20° 08' 48"S, 68° 50' 37"W
12	<i>Orestias sp.</i>	Bofedal Lirima	Chile	19° 51' 08"S, 68° 54' 38"W
13	<i>Orestias sp.</i>	Bofedal Cotaculco	Chile	19° 38' 28"S, 68° 45' 48"W
14	<i>O. agassii</i>	Río Isluga	Chile	19° 15' 23"S, 68° 42' 23"W
15	<i>O. cf. agassii</i>	Río Pampa Aullagas	Bolivia	19° 11' 04"S, 67° 02' 52"W
16	<i>O. agassii</i>	Lago Poopó	Bolivia	18° 52' 34"S, 67° 05' 33"W
17	<i>O. chungarensis</i>	Parque Nacional Lauca	Chile	18° 22' 50"S, 69° 20' 58"W
	<i>O. laucaensis</i>			
	<i>O. piacotensis</i>			
	<i>O. parinacotensis</i>			
18	<i>O. agassii</i> <i>O. lutea</i>	Lago Uru Uru	Bolivia	18° 07' 24"S, 67° 05' 03"W
19	<i>Orestias sp.</i>	Bofedal Chañopalca	Chile	18° 04' 32"S, 69° 16' 22"W
20	<i>Orestias sp.</i>	Bofedal Caquena	Chile	18° 03' 55"S, 69° 12' 26"W
21	<i>Orestias sp.</i>	Bofedal Colpa	Chile	18° 03' 34"S, 69° 13' 59"W
22	<i>Orestias sp.</i>	Bofedal Surapalca	Chile	17° 48' 26"S, 69° 42' 53"W
23	<i>Orestias sp.</i>	Bofedal Humaquilca	Chile	17° 44' 22"S, 69° 23' 21"W
24	<i>O. agassii</i>	Río Desaguadero	Bolivia	17° 16' 45"S, 68° 36' 52"W
25	<i>O. agassii</i> <i>O. tschudii</i>	Guaqui	Bolivia	16° 35' 20"S, 68° 51' 29"W
26	<i>O. cf. agassii</i>	Lago Huni	Bolivia	16° 32' 53"S, 68° 00' 45"W

Tabla 1. (continuación)

ID	Especies	Localidad	País	Coordenadas geográficas
27	<i>O. agassii</i>	Río Tiahuanaco	Bolivia	16° 32' 49"S, 68° 40' 37"W
28	<i>O. cf. agassii</i>	Lagos San José y Santa Rosa	Bolivia	16° 28' 03"S, 68° 57' 26"W
29	<i>O. cf. agassii</i>	Siete Lagunas	Bolivia	16° 25' 58"S, 68° 08' 55"W
30	<i>O. agassii</i>	Río Caminaque	Perú	16° 18' 52"S, 69° 38' 02"W
31	<i>O. cf. agassii</i>	Isla Taquiri	Bolivia	16° 17' 24"S, 68° 48' 07"W
32	<i>O. farfani</i>	Lago Pequeño	Perú	16° 16' 01"S, 69° 02' 37"W
33	<i>O. cf. agassii</i>	Lago Tuni	Bolivia	16° 14' 47"S, 68° 14' 31"W
34	<i>O. agassii</i> <i>O. lutea</i>	Huatajata	Bolivia	16° 12' 49"S, 68° 41' 24"W
35	<i>O. agassii</i>	Río Molino	Perú	16° 12' 27"S, 69° 28' 53"W
36	<i>O. polonorum</i> <i>O. cf. agassii</i>	Lago Hichu Khota	Bolivia	16° 10' 27"S, 68° 22' 48"W
37	<i>O. cf. agassii</i>	Lago Khara Khota	Bolivia	16° 09' 45"S, 68° 22' 21"W
38	<i>O. ctenolepis</i>	Río Zapatilla	Perú	16° 09' 08"S, 69° 37' 40"W
39	<i>O. agassii</i>	Lago Khotia	Bolivia	16° 07' 23"S, 68° 21' 05"W
40	<i>O. agassii</i>	Río Ilave	Perú	16° 06' 08"S, 69° 38' 33"W
41	<i>O. frontosa</i>	Sotalaya, costa este	Bolivia	15° 57' 13"S, 68° 47' 17"W
42	<i>O. agassii</i>	Chucuito	Perú	15° 54' 10"S, 69° 52' 51"W
43	<i>O. lutea</i>	Sotalaya	Bolivia	15° 54' 08"S, 68° 56' 33"W
	<i>O. frontosa</i>			
	<i>O. puni</i>			
	<i>O. tschudii</i>			
44	<i>O. lutea</i> <i>O. rotundipinnis</i> <i>O. alba</i> <i>O. olivacea</i>	Bahía Puno	Perú	15° 50' 56"S, 69° 57' 51"W
45	<i>O. agassii</i>	Lago Saracocha	Perú	15° 48' 45"S, 70° 37' 41"W
46	<i>O. agassii</i>	Crucero Alto	Perú	15° 46' 30"S, 70° 55' 07"W
47	<i>O. agassii</i> <i>O. lutea</i>	Isla Taquile	Perú	15° 45' 08"S, 69° 41' 29"W
	<i>O. agassii</i>			
	<i>O. lutea</i>			
48	<i>O. olivacea</i> <i>O. puni</i> <i>O. silustani</i>	Lago Umayo	Perú	15° 44' 21"S, 70° 10' 15"W

Tabla 1. (continuación)

ID	Especies	Localidad	País	Coordenadas geográficas
49	<i>O. agassii</i>	Lagunillas	Perú	15° 42' 03"S, 70° 48' 36"W
50	<i>O. cf. agassii</i>	Río Ilpa	Bolivia	15° 41' 43"S, 70° 04' 06"W
	<i>O. agassii</i>			
	<i>O. lutea</i>			
	<i>O. rotundipinnis</i>			
51	<i>O. frontosa</i>	Península Capachica	Perú	15° 39' 55"S, 69° 51' 02"W
	<i>O. puni</i>			
	<i>O. tschudii</i>			
	<i>O. alba</i>			
52	<i>O. agassii</i>	Río Ilpa	Perú	15° 38' 24"S, 70° 10' 14"W
53	<i>O. richersoni</i>	Río Coata	Perú	15° 34' 42"S, 69° 54' 35"W
54	<i>O. agassii</i>	Río Lampa	Perú	15° 23' 09"S, 70° 21' 26"W
	<i>O. multiporis</i>			
55	<i>O. agassii</i>	Bahía Moho	Perú	15° 22' 30"S, 69° 31' 20"W
	<i>O. lutea</i>			
	<i>O. alba</i>			
56	<i>O. cf. agassii</i>	Lago Juliaca	Bolivia	15° 22' 28"S, 70° 12' 39"W
57	<i>O. tschudii</i>	Taraco	Perú	15° 17' 27"S, 69° 59' 03"W
	<i>O. agassii</i>			
58	<i>O. lutea</i>	Huancané	Perú	15° 12' 55"S, 69° 46' 15"W
	<i>O. alba</i>			
59	<i>O. agassii</i>	Cailloma	Perú	15° 12' 10"S, 71° 46' 16"W
60	<i>O. agassii</i>	Laguna Arapa y Ríos Ramis y Chupa	Perú	15° 11' 48"S, 69° 59' 13"W
	<i>O. lutea</i>	Laguna Arapa		
61	<i>O. cf. agassii</i>	Río Suches, Lagos Chojña Khota y Cañuhma Janka Khota y bofedal Ulla Ulla	Bolivia	15° 05' 11"S, 69° 17' 11"W
62	<i>O. agassii</i>	Ríos Pucara y Porque y Tirapata	Perú	15° 02' 24"S, 70° 21' 17"W
63	<i>O. agassii</i>	Río Azangaro	Perú	14° 54' 20"S, 70° 11' 07"W
64	<i>O. agassii</i>	Río Ayaviri	Perú	14° 53' 34"S, 70° 31' 29"W
65	<i>O. cf. agassii</i>	Bofedales Catantica y Catantica Alto	Bolivia	14° 48' 26"S, 69° 13' 24"W
66	<i>O. agassii</i>	Río Santa Rosa	Perú	14° 38' 42"S, 70° 45' 20"W
67	<i>O. agassii</i>	Río Colorado	Bolivia	14° 38' 42"S, 68° 19' 33"W

Tabla 1. (continuación)

ID	Especies	Localidad	País	Coordenadas geográficas
68	<i>O. agassii</i>	Río Langui, Laguna Langui y Río Layo	Perú	14° 26' 01"S, 71° 17' 07"W
69	<i>O. agassii</i>	San Pablo	Perú	14° 12' 28"S, 71° 19' 08"W
70	<i>O. agassii</i> <i>O. jussieu</i>	Lago Urcos	Perú	13° 40' 53"S, 71° 37' 27"W
71	<i>O. agassii</i> <i>O. jussieu</i> <i>O. munda</i>	Pueblo Urubamba, Hacienda Urco, Lago Huaypo Lago Chinchero Cuzco	Perú	13° 16' 42"S, 72° 10' 01"W
72	<i>O. elegans</i>	Río Santa Eulalia	Perú	11° 35' 47"S, 76° 10' 08"W
73	<i>O. empyraeus</i> <i>O. polonorum</i>	Lago Junín y Río Mantaro Lago Junín	Perú	10° 58' 56"S, 76° 11' 27"W
74	<i>O. gymnota</i>	Provincia Pasco	Perú	10° 39' 11"S, 76° 08' 52"W
75	<i>O. hardini</i>	Lago Yanacocha	Perú	10° 05' 39"S, 77° 15' 39"W
76	<i>O. ututo</i>	Lago Ututo	Perú	9° 50' 49"S, 77° 31' 13"W
77	<i>O. hardini</i>	Lago Llacsha	Perú	9° 44' 30"S, 77° 32' 56"W

En Chile, se han descrito siete de estas especies (Fig. 3):

- 1) *Orestias parinacotensis* Arratia 1982; habita en el bofedal de Parinacota.
- 2) *Orestias laucaensis* Arratia 1982, que habita en el río Lauca y en las lagunas de Cotacotani.
- 3) *Orestias chungarensis* Vila & Pinto 1986, que habita en el lago Chungará.
- 4) *Orestias piacotensis* Vila 2006, que habita en la laguna Piacota
- 5) *Orestias* cf. *agassii* Valenciennes (1846); se encuentra asignada a las poblaciones del río Isluga, del salar Huasco y del río Collacagua.
- 6) *Orestias gloriae* Vila, Scott, Méndez, Valenzuela, Iturra & Poulin 2012; habita en tres vertientes del salar Carcote.
- 7) *Orestias ascotanensis* Parenti 1984, que habita en 12 vertientes del salar Ascotán.

Las cuatro primeras especies están circunscritas al Parque Nacional Lauca (PNL de aquí en adelante) ubicado en el extremo norte de Chile (XV Región de Arica y Parinacota).

Mann (1954) propuso que *O. agassii* se encuentra presente en las regiones de Tarapacá, Antofagasta y Atacama, por lo que los estudios posteriores que han incluido a las poblaciones de *Orestias* del río Isluga, el salar Huasco y el río Collacagua las han asignado a esta especie. Sin embargo, Cruz-Jofré *et al.* (2014) sugieren que sólo la población de Isluga estaría más relacionadas con *O. agassii* de Bolivia, mientras que la población de Huasco correspondería a una unidad taxonómica diferente. Este resultado es concordante con el análisis filogenético de Vila *et al.* (2013), quienes recuperan un grupo monofilético por cada una de estas tres poblaciones (ver detalles más adelante).



Figura 3. Localidades donde han sido descritas las especies de *Orestias* en Chile. En el rectángulo azul se indican los sitios que se encuentran dentro del Parque Nacional Lauca.

2.1 Estudio sistemático de las especies chilenas de *Orestias*

El estudio de estas especies ha estado enfocado en análisis morfológicos (Arratia 1982; Parenti 1984a; Scott 2010; Vila *et al.* 2010), cariotípicos (Araya 2009; Vila *et al.* 2010; Araya-Jaime *et al.* 2017), filogenéticos (Scott 2010; Esquer Garrigos 2013; Vila *et al.* 2013), filogeográficos (Morales *et al.* 2011; Esquer Garrigos 2013; Guerrero 2013; Cruz-Jofré *et al.* 2016) y de dietas (Guerrero 2013).

2.1.1 Morfología.

Los estudios que han incluido el análisis de la morfología de *Orestias* han estado centrados en el contexto de la sistemática tradicional. Los primeros trabajos que llevaron a cabo análisis de rasgos morfológicos corresponden a los realizados en las descripciones sistemáticas de las especies. Estos trabajos incluyeron caracteres morfológicos utilizando relaciones corporales (por ej., tamaño de la cabeza en relación a la longitud estándar), patrones de escamación corporal, coloración y descripción del aparato branquial, entre otros (Tchernavin 1944; Parenti 1984a). Vila *et al.* (2010) realizaron una comparación de la morfología entre las especies chilenas de *Orestias*, considerando mediciones morfológicas lineales y merísticas (además de caracteres cariotípicos que se detallan más adelante). Entre los resultados destaca el análisis discriminante (Tabla 2), que indicó altas tasas de clasificación correcta para *O. ascotanensis*, *O. cf. agassii* de Isluga, *O. laucaensis*, y *O. chungarensis*, mientras que *O. parinacotensis* y *O. piacotensis*, ambas especies del PNL, no fueron bien clasificadas. Esto evidencia que la morfología de estas dos últimas especies no se diferenció ni de las otras especies del PNL, ni de las que están fuera de esa cuenca. Resulta interesante observar que ocurre lo contrario con *O. chungarensis*, otra especie del PNL, ya que todos sus individuos fueron clasificados correctamente. Por lo tanto, el análisis discriminante de los datos morfológicos lineales diferencian a la mayoría de las especies, excepto a dos especies del PNL (*O. parinacotensis* y *O. piacotensis*).

Por otra parte, los datos merísticos del trabajo de Vila *et al.* (2010) presentaron una alta sobreposición interespecífica que no permite la diferenciación de las especies. En *Orestias*, estos rasgos serían inadecuados para diferenciar especies dado que no consideran la severa asimetría bilateral individual en el

número y disposición de las escamas, ni las diferencias entre sexos o entre juveniles y adultos (Arratia 1982). En consecuencia, se hace necesario realizar futuros análisis que consideren esta variación, para intentar responder con mayor precisión si la morfología de estas especies es diferente o no.

Finalmente, es importante tener en cuenta que resultados descritos por Vila *et al.* (2010) fueron realizados sin considerar el dimorfismo sexual en tamaño corporal descrito para el género. Esto implica que las mediciones realizadas sobre las hembras de una especie no son comparables a las de los machos de otra especie. Para solucionar este problema, se deben realizar análisis por separado para cada sexo.

Tabla 2. Análisis discriminante realizado a partir de los resultados del análisis de componentes principales de las mediciones de morfología lineal de las especies chilenas de *Orestias*. Se indican el número de individuos asignados a cada especie y los porcentajes de clasificación correcta. Modificado de Vila et al. (2010).

	1	2	3	4	5	6	N	% clasificación correcta
1. <i>O. ascotanensis</i>	15	0	0	0	2	3	20	75
2. <i>O. agassii</i>	1	20	0	0	1	0	22	91
3. <i>O. laucaensis</i>	0	0	19	1	0	4	24	79
4. <i>O. parinacotensis</i>	3	4	0	7	0	6	20	35
5. <i>O. chungarensis</i>	0	0	0	0	20	0	20	100
6. <i>O. piacotensis</i>	2	1	7	5	0	10	25	40
Total	21	25	26	13	23	23	131	69

Recientemente, se realizó un análisis de morfometría geométrica para analizar la forma corporal de las especies chilenas de *Orestias* y de otras poblaciones chilenas de *Orestias* recientemente descubiertas, cuyo estatus taxonómico aún no ha sido determinado (F. Cruz-Jofré, comunicación personal). En este estudio se analizaron sólo hembras adultas, ya que se contaba con un tamaño muestral muy bajo de machos. Los análisis de estos datos permitieron distinguir a las especies *O. ascotanensis* (salar Ascotán), *O. gloriae* (salar Carcote) y *O.cf. agassii* de Isluga, mientras que las del PNL (*O. laucaensis*, *O. chungarensis*, *O. parinacotensis* y *O. piacotensis*) y *O. cf. agassii* del salar Huasco no se diferenciaron entre ellas, ni con las poblaciones recientemente descubiertas (Fig. 4).

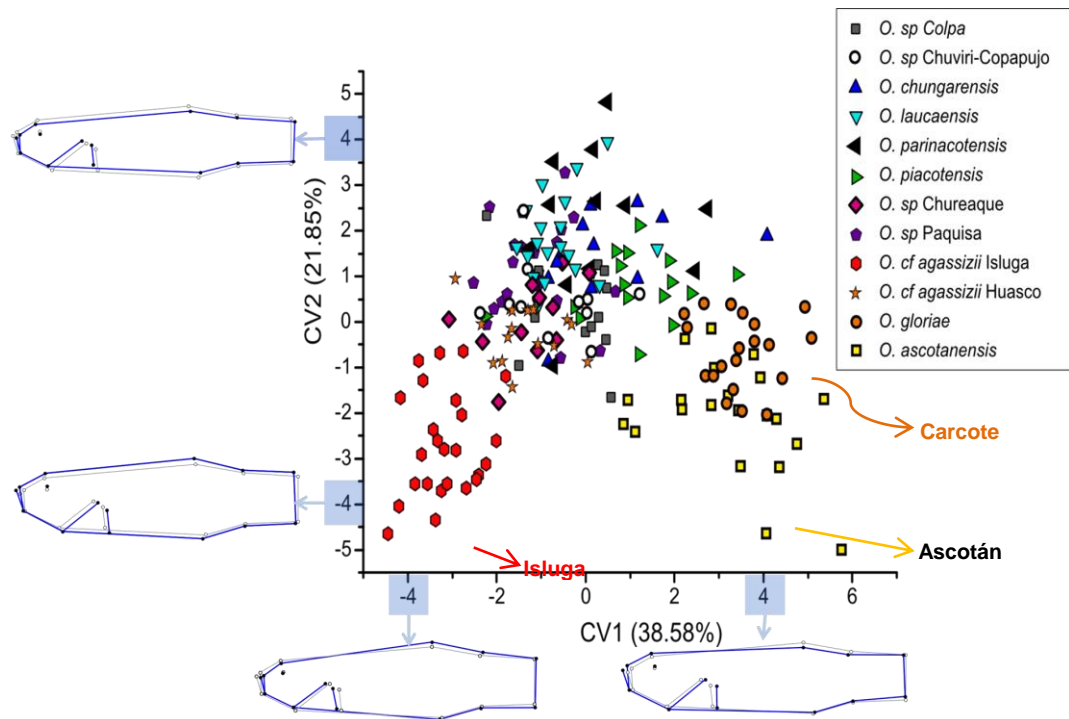


Figura 4. Análisis de variables canónicas de datos de morfometría geométrica basado en 15 landmarks para la comparación por especie. Se grafican las dos primeras variables canónicas (CV1 y CV2), en las que se incluyen los diagramas de cambio para los valores -4 y 4 en ambos ejes: en gris se indica la forma de consenso para todos los individuos y en azul, la forma ponderada. Se indican los individuos de las especies que se distinguen en este análisis: *O. gloriae* (Carcote, en naranja), *O. ascotanensis* (Ascotán, en amarillo) y *O. cf. agassii* de Isluga (en rojo). (F. Cruz-Jofré, comunicación personal).

2.1.2 Cariotipos.

Los análisis de caracteres cariotípicos se han basado en el número y la morfología de los cromosomas (Araya 2009; Vila *et al.* 2010; Araya-Jaime *et al.* 2017). Se ha observado que cada una de las especies chilenas de *Orestias* posee una fórmula cromosómica distinta, donde las diferencias radican principalmente en el número de cromosomas telocéntricos, subtelocéntricos y microcromosomas (Tabla 3). En la Tabla 4 se indican las variaciones en la distribución de la heterocromatina constitutiva y mapeo de genes ribosomales (cluster ribosomal mayor 28S y cluster ribosomal menor 5S) en las especies que se ha realizado estos análisis (para las especies *O. gloriae*, *O. laucaensis* y *O. piacotensis* sólo se ha determinado el presencia de heterocromatina constitutiva; P. Iturra, comunicación personal). Dadas las diferencias existentes en estos caracteres, se ha propuesto que en las especies chilenas de *Orestias* existirían al menos tres linajes cromosómicos (Araya 2009). Uno de estos estaría conformado por las especies del PNL, caracterizadas por el par cromosómico 1 que presenta abundante heterocromatina constitutiva y contiene el cluster ribosomal mayor 28S (esto último sólo ha sido confirmado para *O. chungarensis*). Un segundo linaje estaría constituido por las especies *O. ascotanensis* y *O. gloriae* (además de *O. cf. agassii* del salar Huasco). Este linaje difiere del anterior en la morfología del par cromosómico 1 metacéntrico que no presenta bloques conspicuos de heterocromatina, su composición de bases nitrogenadas es homogénea y no presenta genes ribosomales 28S (Tabla 4). Un tercer linaje estaría formado por *O. cf. agassii* del río Isluga, que se distingue por la distribución de la heterocromatina y la ubicación del cluster ribosomal mayor 28S en los pares de cromosomas 3, 15 y 18.


















En resumen, si bien cada especie presenta un número y una fórmula cromosómicos particulares que las diferencian de todas las otras especies (Tabla 3), la morfología de los cromosomas permiten agruparlas en linajes. En particular, la morfología del cromosoma 1 de las especies del PNL es similar entre ellas y característico sólo de estas especies (Tabla 4). El resto de las especies, por el contrario, presentan diferencias morfológicas de los cromosomas que permiten agruparlas en dos linajes diferentes (Tabla 4), a pesar de tener un número cromosómico igual (Tabla 3).




Finalmente, a pesar de las diferencias y similitudes encontradas en los cariotipos de estas especies, no se han realizado estudios que pongan a prueba la existencia de barreras reproductivas entre ellas.

Tabla 3. Fórmula cromosómica de las especies chilenas de *Orestias* (faltan los datos de *O. cf. agassii* del río Collacagua, cuyo fórmula aún no ha sido descrita). 2n: número cromosómico diploide; NF: número fundamental de brazos; M: cromosomas metacéntricos; SM: cromosomas submetacéntricos; ST: cromosomas subtelocéntricos; T: cromosomas telocéntricos; Mc: microcromosomas. *: descrito en Vila *et al.* (2011); **: mencionado en Araya (2009).

Especie	2n	NF	Fórmula cromosómica					Referencia
			M	SM	ST	T	Mc	
<i>Orestias parinacotensis</i>	48	54	2	4	10	32	-	Vila <i>et al.</i> 2010
<i>Orestias laucaensis</i>	51♀ 52♂	54*	2	2	4	40	2 o 3**	Arratia 1982
<i>Orestias chungarensis</i>	55	54	2	2	20	26	5	Vila <i>et al.</i> 2010 Araya 2009
<i>Orestias piacotensis</i>	52	54	2	2	12	34	2	Vila <i>et al.</i> 2010
<i>Orestias gloriae</i>	48	56	2	6	20	20	-	Vila <i>et al.</i> 2011
<i>Orestias ascotanensis</i>	48	54	2	4	4	38	-	Vila <i>et al.</i> 2010 Araya 2009
<i>Orestias cf. agassii</i> Huasco	48	54	2	4	14	28	-	Araya 2009
<i>Orestias cf. agassii</i> Isluga	48	54	2	4	24	18	1	Araya 2009

Tabla 4. Linajes cromosómicos en las especies chilenas de *Orestias*. Se indican los cromosomas con bandeo C, que indica la heterocromatina constitutiva, la ubicación del cluster ribosomal mayor (28S) y del cluster ribosomal menor (5S). Se especifica la morfología del par cromosómico 1 (metacéntrico) para todas las especies, y luego los pares cromosómicos que contienen los cluster ribosomales, diferenciándolos en submetacéntricos y telocéntricos. Para las especies *O. parinacotensis*, *O. piacotensis* y *O. gloriae* sólo se han obtenido datos de la heterocromatina constitutiva. Modificado de Araya (2009).

Linaje	Especies	2n	Característica cromosómica		
			M	SM	T
1	<i>O. chungarensis</i>	55	 Par 1	 Par 2	 Par 23
	<i>O. parinacotensis</i>	48	 Par 1		
	<i>O. piacotensis</i>	52	 Par 1		
2	<i>O. ascotanensis</i>	48	 Par 1	 Par 2	 Par 3
	<i>O. cf. agassii</i> Huasco	48	 Par 1	 Par 2	 Par 12
	<i>O. gloriae</i>	48	 Par 1		
3	<i>O. cf. agassii</i> Isluga	48	 Par 1	 Par 2	 Par 3
				 Par 15	 Par 18

 Bando C
  Cluster ribosomal mayor 28S
  Cluster ribosomal menor 5S

2.1.3 Análisis filogenéticos

Se han realizado estudios filogenéticos de las especies chilenas de *Orestias* (Scott 2010; Vila *et al.* 2013) y filogeográficos para determinar la diversidad y estructura genética de algunas de sus poblaciones (Morales *et al.* 2011; Cruz-Jofré *et al.* 2016; Guerrero-Jiménez *et al.* 2017). Para dilucidar las relaciones filogenéticas se han utilizado tres marcadores moleculares mitocondriales: la región control, el gen del citocromo b (cytb) y el gen la subunidad 2 de la enzima NADH deshidrogenasa (ND2) (Figura 5). La reconstrucción filogenética recuperó las especies *O. ascotanensis*, *O. gloriae*, cada una formando un grupo monofilético. Por otra parte, *O. cf. agassii* correspondería un grupo polifilético, pues cada localidad asignada a esta especie en Chile forma un grupo monofilético diferente. Finalmente, las especies presentes en el PNL también forman un grupo monofilético con buen soporte estadístico, pero dentro de este linaje no se recuperaron las especies (Vila *et al.* 2013). Esquer Garrigos (2013) realizó una filogenia molecular (utilizando la región control del ADN mitocondrial y cytb) del complejo *agassii*, en la cual incluyó especies bolivianas, peruanas y las siguientes especies chilenas: *O. ascotanensis*, *O. gloriae*, *O. laucaensis* y *O. cf. agassii* del salar Huasco y del río Isluga. Si bien, en este análisis la matriz de datos fue diferente al estudio anterior, nuevamente, se recuperó la monofilia de todas las especies chilenas, excepto de *O. laucaensis* (no es posible hacer mayores inferencias acerca de las relaciones filogenéticas de esta especie con otros taxa al no estar incluidas el resto de las especies del PNL).

Por otra parte, Guerrero-Jiménez *et al.* (2017) realizó un estudio centrado en las poblaciones y especies del PNL, analizando su diversidad y estructura genética, utilizando la región control del ADN mitocondrial y ocho marcadores microsátélites. Uno de sus principales resultados es que éstas corresponden a grupos genéticos diferenciados (sus frecuencias alélicas son diferentes), especialmente *O. chungarensis* y *O. piacotensis*, pero aún no son grupos divergentes (aún comparten alelos), y representarían entonces un caso de especiación incipiente. La formación de la cuenca del PNL hace menos de 15 mil años podría explicar la falta de divergencia de estos grupos (ver detalles más adelante).

De esta manera, ambos trabajos obtienen resultados similares a los ya mencionados: las especies chilenas de *Orestias* se diferencian genéticamente, excepto las del PNL, que aún no han divergido.

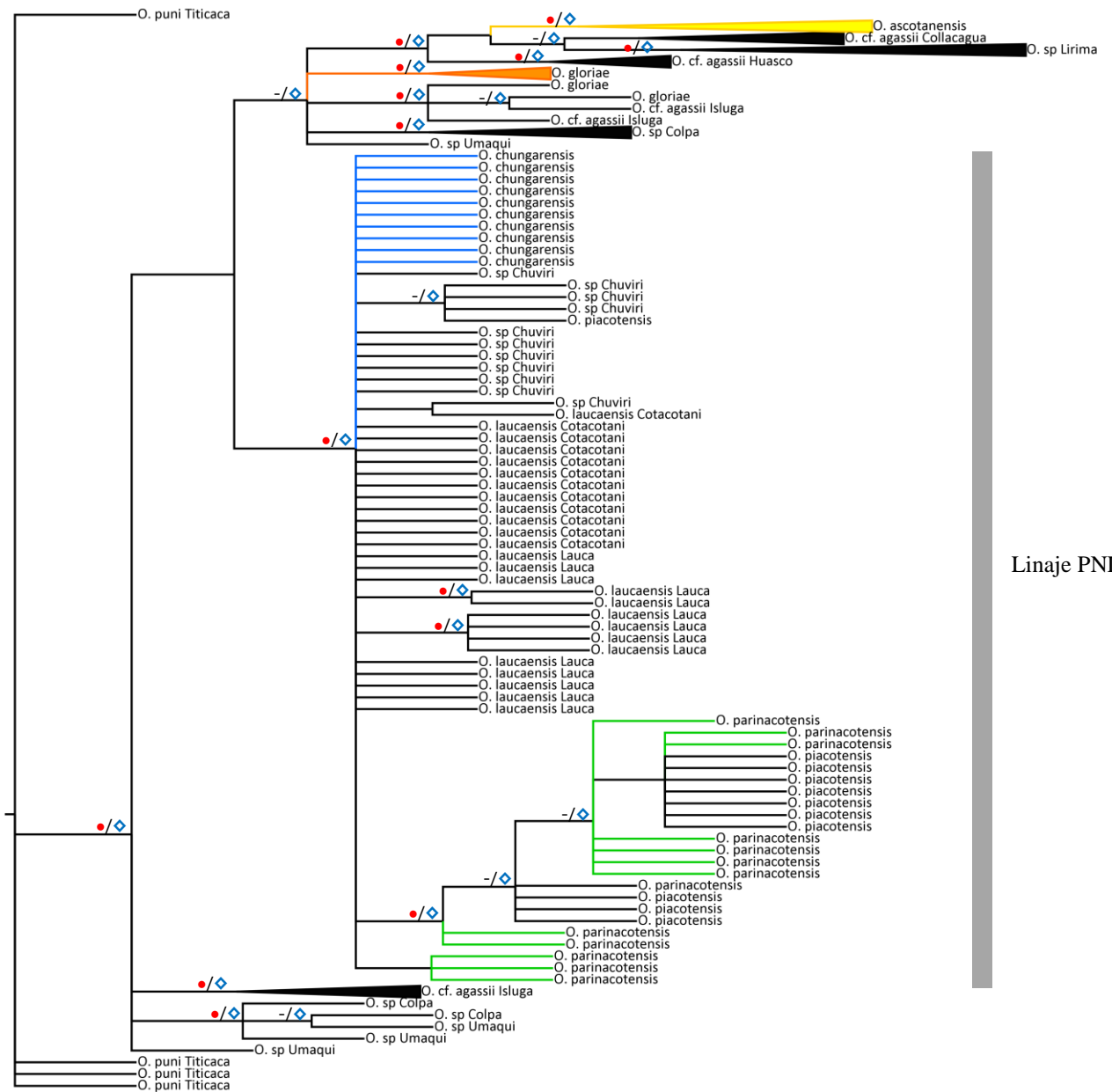


Figura 5. Reconstrucción filogenética (máxima verosimilitud e inferencia bayesiana) de las especies chilenas de *Orestias* realizada con tres marcadores mitocondriales (región control, cytb y ND2; 3035 pb en total). Cada especie descrita forma un grupo monofilético, excepto las del PNL que quedan reunidas dentro de un mismo grupo monofilético (indicado como “linaje PNL” con la barra gris), dentro del cual no se recupera la monofilia de ninguna de estas especies. En amarillo se indica el grupo monofilético de *O. ascotanensis* (salar Ascotán). En naranja, se observa *O. gloriae* (salar Carcote). En azul y verde se indican

las secuencias de *O. chungarensis* (lago Chungará) y *O. parinacotensis* (bofedal Parinacota), respectivamente. Los valores en los nodos corresponden a los valores de bootstrap >70% (indicado con un ●) y la probabilidad a posterior >0,95 (indicado con un ◇). Modificado de Morales *et al.* (2013)

2.1.4 Dietas

Guerrero (2013) realizó un estudio de las dietas de las especies del PNL. Los resultados mostraron que las principales presas de estas especies son zooplancton y macroinvertebrados. *O. chungarensis* presentaría la mayor riqueza de especies en su dieta, basada mayoritariamente en *Cyprinotus* y *Nauplius*. Por otra parte, *Biomphalaria* es una de las presas más consumida por todas las especies, pero en mayor proporción por *O. parinacotensis* y *O. laucaensis*. En general, las distintas especies no presentan grandes diferencias entre sus dietas. A pesar de la similitud en las dietas, no se observa una alta sobreposición de nicho, lo que podría deberse a que la proporción de presas consumidas difiere entre especies. Es decir, las especies del PNL presentarían dietas similares, pero las presas serían consumidas en distinta proporción. Con esta evidencia, no es posible hacer inferencias acerca de si son especies generalistas o especialistas, ya que en este trabajo no se estimó la oferta ambiental de presas.

2.1.5 Fauna bentónica y zooplancton

Márquez-García *et al.* (2009) realizaron un estudio acerca de la composición y estructura de la biota acuática de los sistemas donde habitan las especies de *Orestias* chilenas. Este estudio mostró que las vertientes de los salares Ascotán y Carcote presentan una fauna bentónica y zooplancton similar que se mantienen constante a lo largo del año. Por otra parte, las localidades del PNL, donde cada cual es un ambiente diferente (lago, río, bofedal, lagunas), presentan diferencias entre ellas, ya sea en la fauna bentónica durante el invierno (no hay datos acerca de la composición durante el verano), o en la composición de zooplancton durante el invierno, cuando el lago Chungará se diferencia de todos los sitios (los del PNL y los que están fuera de esta cuenca). El lago Chungará, entonces, es la localidad del PNL

que presenta mayores variaciones en su biodiversidad acuática en comparación al resto de las localidades del PNL, en cuanto a su composición y comportamiento estacional (excepto para el zooplancton durante la temporada primavera-verano). Finalmente, las localidades donde se ha asumido que hay *O. agassii*, Isluga y Huasco, se observan muy diferentes en cuanto a la composición de fauna bentónica, y Huasco presenta diferencias estacionales, al contrario de Isluga donde la composición permanece igual durante el año. En cuanto a la composición de zooplancton, Isluga permanece similar a lo largo del año, la cual es levemente diferente de la de Huasco en primavera-verano (no hay datos acerca de la composición durante el invierno).

2.2 Perspectiva general del problema

En la Tabla 5 se resumen los distintos caracteres que han sido analizados en las especies chilenas de *Orestias* y que fueron detallados en la sección anterior, de los que se desprende que:

- (a) *O. ascotanensis* y *O. gloriae* son especies que se diferencian entre ellas y del resto, ya sea analizando rasgos fenotípicos, morfológicos, o genéticos, como cariotipos o marcadores moleculares. Por otra parte, el hábitat de estas especies es muy similar en dos aspectos: (i) ambos sistemas corresponden a afloramientos de aguas subterráneas que fluyen sobre la superficie evaporítica de los salares, y (ii) la composición de la fauna bentónica y de zooplancton es similar en ambos ambientes (Keller & Soto 1998; Risacher *et al.* 2003; Márquez-García *et al.* 2009).
- (b) En relación a las especies del PNL, distintos análisis han mostrado resultados contradictorios en cuanto a la diferenciación de estas especies. La morfología lineal y cariotipos han permitido distinguir a estas especies como entidades separadas, sin embargo, las mediciones merísticas, la morfometría geométrica, la filogenia molecular y la filogeografía han indicado que estas especies son grupos no diferenciados (morfometría geométrica y filogenia molecular), o por lo menos, son grupos con bajo nivel de diferenciación (filogeografía). Por otra parte, el desacople entre la morfología (diferenciación de las especies) y la genética (no diferenciación de las especies) se ha interpretado como resultado de separación incompleta de linajes (o *incomplete lineage sorting* en

inglés; Esquer Garrigos 2013; Vila *et al.* 2013). Esto quiere decir que no habría transcurrido tiempo suficiente para que se diferencien las frecuencias génicas de las distintas especies. Una segunda interpretación es que la falta de diferenciación genética se deba a la existencia de flujo génico entre las poblaciones, que mezclaría los pools génicos evitando la divergencia. Por el momento, la primera posibilidad parece más plausible dados los datos de la formación de la cuenca (Clavero *et al.* 2004; Sáez *et al.* 2007; ver más adelante) y los tiempos de divergencia de las especies (Scott 2010). Por otra parte, Guerrero-Jiménez *et al.* (2017) sugirieron un posible flujo asimétrico hacia el bofedal Parinacota que podría ocurrir durante eventos de aumento del nivel de agua en la zona, dados los resultados de análisis de microsatélites y la menor elevación en la que se ubica este sitio con respecto a los otros del PNL. Pero hasta la fecha no hay evidencias de una conexión física que facilite el flujo génico entre los distintos sistemas del PNL.

Tabla 5. Resumen de los caracteres analizados en las especies chilenas de *Orestias*. *: estudios realizados sólo para las especies del PNL.

Carácter	Resultado
Diagnosis y descripción	Rangos de mediciones merísticas se sobreponen entre especies. Ni asimetría bilateral en caracteres de escamación ni dimorfismo sexual fueron considerados en claves de identificación, por lo que el número y distribución de escamas y relaciones corporales no serían buenos caracteres diagnósticos.
Morfología lineal	Diferencias significativas entre especies, a pesar de que no se consideró el dimorfismo sexual, excepto <i>O. piacotensis</i> .
Morfometría geométrica	Reconoce sólo a <i>O. ascotanensis</i> , <i>O. gloriae</i> y <i>O. cf. agassii</i> del río Isluga como grupos con forma corporal diferente.
Cariotipos	2n = 48, excepto para <i>O. laucaensis</i> (51♀ y 52♂), <i>O. chungarensis</i> (55) y <i>O. piacotensis</i> (52). NF = 54, excepto para <i>O. gloriae</i> (56). Fórmulas cromosómicas diferentes para cada especie, debido principalmente al número de cromosomas telocéntricos y subtelocéntricos. Presencia de microcromosomas en <i>O. laucaensis</i> (2 ó 3), <i>O. chungarensis</i> (5), <i>O. piacotensis</i> (2) y <i>O. cf. agassii</i> de Isluga (1). Linaje cromosómico 1 compuesto por las especies del PNL. Linaje cromosómico 2 compuesto por <i>O. ascotanensis</i> , <i>O. gloriae</i> y <i>O. cf. agassii</i> del salar Huasco. Linaje cromosómico 3 compuesto por <i>O. cf. agassii</i> del río Isluga.
Filogenia molecular	Cada una de las siguientes especies forman un grupo monofilético por sí mismas: <i>O. ascotanensis</i> , <i>O. gloriae</i> , <i>O. cf. agassii</i> de Huasco, <i>O. cf. agassii</i> de Isluga y <i>O. cf. agassii</i> de Collacagua. Especies del PNL contenidas dentro de un grupo monofilético, pero no se reconocen las especies dentro de este clado.
Filogeografía*	Las especies nominales del PNL corresponden a grupos diferenciados, pero no divergentes.
Dietas*	Las especies del PNL presentarían dietas similares, pero las presas serían consumidas en distinta proporción.
Fauna bentónica	Ascotán y Carcote son similares entre sí. Río Lauca, bofedal de Parinacota y lago Chungará son diferentes entre sí y con el resto de las localidades. En Huasco hay una marcada diferencia estacional, a diferencia de Isluga que permanece igual.
Zooplancton	Ascotán y Carcote son similares entre sí. Río Lauca, bofedal de Parinacota y lago Chungará son similares durante el verano, sin embargo, Chungará se diferencia durante el invierno. Isluga permanece similar durante el año, y es levemente diferente de Huasco durante el verano.

2.3. Historia geológica de los hábitats de las especies chilenas de *Orestias*.

La discrepancia en el grado de diferenciación entre las especies chilenas del género *Orestias* puede ser debido a la historia geológica de los ambientes que habitan. La cuenca del PNL se habría formado hace 12-8 kya, cuando el volcán Parinacota colapsó, invirtiendo el flujo de agua del río Lauca, dejándolo que fluya desde Chile hacia Bolivia y se habría alcanzado la configuración actual de la cuenca (Clavero R. *et al.* 2004; Sáez *et al.* 2007). Este evento sería el que habría dado inicio a la diferenciación genética de las especies, pero el corto tiempo de separación de las poblaciones no permitiría aún recuperarlas como linajes independientes (Vila *et al.* 2013; Guerrero-Jiménez *et al.* 2017). Por otra parte, el tiempo de divergencia de las especies que habitan los salares del extremo sur-oeste del Altiplano (salar Carcote y salar Ascotán) sería mucho más antiguo y estaría relacionado con los grandes paleolagos pleistocénicos que existieron en el Altiplano. El cuerpo de agua más grande descrito en el Altiplano corresponde al paleolago Tauca (18-14 kya, o hasta los 11 kya si se considera la fase Coipasa como la fase final de Tauca), el que habría conectado desde la cuenca del Poopó (Bolivia) hasta el salar Carcote, pero no habría alcanzado a inundar el salar Ascotán (Figura 6) (Sylvestre *et al.* 1999; Placzek *et al.* 2006, 2011; Blard *et al.* 2011). Dado que este lago, el más extenso que ha existido, no consiguió conectar estas dos cuencas, el lago más antiguo descrito hasta la actualidad, pero de menor extensión tampoco lo habría logrado (el paleolago Ouki, que permaneció entre los 98 y los 120 kya) (Placzek *et al.* 2006, 2011; Blard *et al.* 2011). Es decir, la conexión de estas dos especies debió haber sido previa a la existencia de estos paleolagos, es decir antes de los 100 kya. Scott (2010) estimó que el tiempo de divergencia entre *O. glorieae* y el clado compuesto por *O. ascotanensis*, *O. cf. agassii* de Huasco y Collacagua y *O. sp* de Umaqui (una localidad con presencia de *Orestias* sin catalogar aún) fue de entre 992-633 kya, dependiendo de la tasa de mutación de los marcadores mitocondriales utilizados para la construcción de los árboles filogenéticos (región control y dos tasas mutacionales de *cytb*).

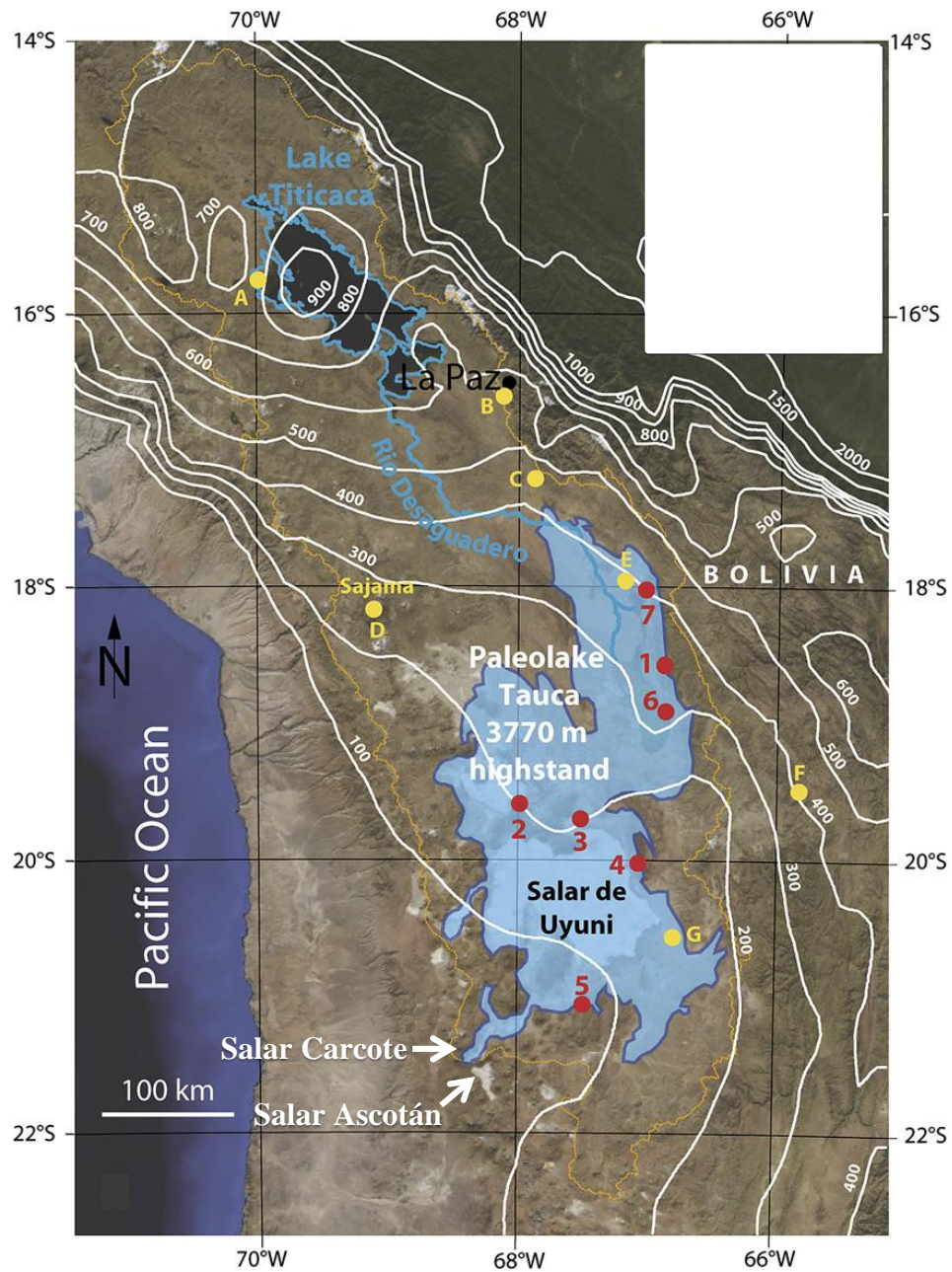


Figura 6. Reconstrucción del nivel más alto que habría alcanzado el paleolago Tauca (3.770 m sobre el nivel del mar). Se indica la ubicación del salar Carcote, que habría formado parte de este lago, mientras que el salar Ascotán (al sur del salar Carcote) no habría sido inundado por este paleolago. Modificado de Blard et al. (2011).

2.4. Pares de especies a comparar en este estudio.

En este estudio se consideraron cuatro de las siete especies chilenas de *Orestias*: *O. ascotanensis*, *O. gloriae*, *O. chungarensis* y *O. laucaensis*. Esta elección se hizo sobre la base del grado de diferenciación estimado a partir de los diferentes caracteres mencionados previamente. *O. chungarensis* y *O. laucaensis* son buenos representantes de una etapa inicial del proceso de especiación dado el bajo nivel de diferenciación entre estos taxa. Mientras que *O. ascotanensis* y *O. gloriae* son un par de especies característicos de una etapa avanzada en el proceso de especiación.

HIPÓTESIS

1)

Antecedentes:

El proceso de especiación es un proceso continuo en el cual la magnitud de la divergencia representa una medida que indica qué tan avanzado se está en este proceso.

Hipótesis:

A nivel genómico global y por locus, existe mayor divergencia entre *O. ascotanensis* y *O. gloriae*, que están en una etapa avanzada del proceso de especiación, que entre *O. chungarensis* y *O. laucaensis*, las especies que estén en una etapa incipiente de este proceso.

2)

Antecedentes:

Las cuatro especies de *Orestias* se originaron en un contexto de especiación alopátrica, donde los diferentes escenarios que ha experimentado cada una han provocado que la diferenciación ocurra de manera independiente en ellas.

Hipótesis:

Los patrones de diferenciación genómica entre especies en diferentes etapas del continuo de especiación presentarán bajos niveles de similitud, tanto entre especies como entre etapas de este continuo.

OBJETIVOS.

OBJETIVO GENERAL.

Caracterizar los patrones de divergencia genómica en dos pares de especies de *Orestias* que se encuentran en diferentes etapas del continuo de especiación.

OBJETIVOS ESPECÍFICOS.

1. Estimar la estructura genética a nivel interespecífico.
2. Describir y comparar el paisaje genómico de la diferenciación en pares de especies de *Orestias* en distintas etapas del continuo de especiación.
3. Establecer similitudes y diferencias entre especies en los SNPs detectados, regiones genómicas en las que se ubican y funciones biológicas en las que están involucradas.

RESULTADOS.

Los resultados de esta tesis se dividen en dos partes. La primera está descrita en el Capítulo II, donde se detallan los resultados de los análisis realizados para detectar y estimar el grado de diversidad y estructuración interespecífico en *Orestias*. La segunda parte, expuesta en el Capítulo III, describe la diferenciación global y en cada locus, obtenidas en dos etapas del continuo de especiación, además de la determinación de las regiones del genoma donde se han generado los polimorfismos y los procesos biológicos enriquecidos que estarían involucrados en la diferenciación de cada especie.

Ambos capítulos están redactados como manuscrito (el primero está escrito en inglés) para ser enviado a publicación, por lo tanto ambos incluyen una introducción del problema, descripción de los métodos utilizados, los resultados obtenidos y una discusión final.

CAPITULO II.

ESTRUCTURA GENÉTICA EN CUATRO ESPECIES DEL GÉNERO *ORESTIAS*.

Allopatric diversity in *Orestias* (Teleostei; Cyprinodontidae) at different stages of the speciation continuum

I. Abstract

The Andean pupfish of the genus *Orestias* are famous for the extensive species radiation that is endemic to Lake Titicaca. However, the group is much more widely distributed in the Andean Altiplano where they occupy a wide range of different habitats from streams and lakes to freshwater springs in salt pans left behind by ancient great lakes that dried up in the Holocene. The diversity and evolutionary history of *Orestias* in this system is not well known, yet it may be key to understanding the dramatic radiation in the Titicaca system. We sampled four species inhabiting isolated water bodies, two saltpan springs (*O. ascotanensis* in the Ascotan salt pan and *O. gloriae* in the Carcote salt pan), a freshwater lake (*O. chungarensis* in the Lake Chungara) and a stream (*O. laucaensis* in the Lauca River), for population genomic investigations based on RAD sequencing to estimate divergence between water bodies. The first two represent an advanced stage in the speciation continuum, while the last two correspond to a recent stage in this continuum.

After the genomic analysis we found a very deep divergence between the salt pan samples, Ascotán and Carcote, and between these and the Lake Chungará and Lauca River samples, and much less divergence between lake and stream. These results highlight the benefits of using genomic data to clarify the limits of taxonomic groups in order to unveil the mechanism of speciation.

2. Introduction

The process of speciation is often thought of as a continuous process where reproductive isolation gradually increases, starting from an initial stage represented by two fully compatible and randomly mating populations to a late stage where pre- or postzygotic reproductive isolation is complete between different species (Nosil *et al.* 2009b). To study pairs of closely related species that are at different stages of this continuum allows to indirectly infer how the differentiation and speciation processes happen (Coyne & Orr 2004; Nosil 2012; Nosil & Feder 2013; Cruickshank & Hahn 2014; Seehausen *et al.* 2014). Here, we follow this approach by documenting the divergence between populations at different stages of the speciation continuum: between allopatric but young taxonomic species and between very divergent allopatric species.

The fish genus *Orestias*, distributed in the Altiplano, the Andean plateau of Bolivia, Peru and Chile, offers several opportunities to study the speciation process because populations and species with different levels of differentiation can be found that may represent more than one mode of diversification. This genus is the most speciose genus of pupfish (Cyprinodontidae family) with 46 described species so far. Half of these are endemic to the Titicaca Lake (Parenti 1984a) representing an extensive species radiation, second in richness only to radiations of cichlid fish in African lakes (Fryer & Iles 1972; Kornfield & Smith 2000; Turner *et al.* 2001). However, the diversity and evolutionary history of *Orestias* is not well known. Importantly, the genus is also distributed on the Altiplano outside Lake Titicaca, and its distribution may be at least partly explained by the distribution of ancient lakes some of which were connected to Lake Titicaca (Sylvestre *et al.* 1999; Placzek *et al.* 2006, 2011; Blard *et al.* 2011). The study of diversity and diversification in this genus outside Titicaca may be key to understanding the dramatic radiation in the Titicaca system.

We studied four *Orestias* species inhabiting isolated water bodies in the Chilean Altiplano. We sampled these four aquatic systems, two salt pan springs, a freshwater lake and a nearby stream, each of them containing one nominal species, for population genomic research, by investigating genomic patterns divergence between water bodies. The four nominal species are *O. chungarensis* in the Lake Chungara, *O.*

laucaensis in the Lauca River, *O. ascotanensis* in the Ascotan salt pan, and *O. gloriae* in the Carcote salt pan. The first two taxa correspond to closely related, recently diverging species occupying different albeit geographically isolated habitats (lake versus stream) in the Lauca National Park (LNP). This study design allows us to compare the genomic signature of population structure between nominal species, that of a case of recent divergence between a lake and a stream species and that of an ancient speciation process represented by two species that occupy similar environments (nearby but fully isolated salt pans).

3. *Methods.*

In order to perform the genomic analyses on the four species mentioned above, previously we verified that the sampled individuals are representatives of these species, by testing (i) if they belong to the same clades as the species and (ii) if they represent the same or new genetic diversity compared to the previously described. For the first point, we reconstructed the phylogenetic relationships of Chilean *Orestias* using mtDNA control region as molecular marker, including these samples, the seven Chilean species and *Orestias* populations from recently discovered locations (Tabla 6). For the second goal, we obtained the haplotypic networks of the clades where the samples for genomic analyses are. Hereinafter, these individuals will be called “RAD individuals”, to differentiate them from those that only have mtDNA control region sequences.

3.1 *Sampling, DNA extraction, sequencing of mtDNA control region and RAD-Seq.*

For genomic analyses, 23 or 24 individuals of four species of *Orestias* inhabiting four different habitats were sampled (Tabla 6). These species are (i) *O. laucaensis* from the Lauca River, (ii) *O. chungarensis* from the Lake Chungara, (iii) *O. ascotanensis* from the spring 5 of the Ascotan salt pan, and (iv) *O. gloriae* from the spring 1 of the Carcote salt pan. These individuals were sent alive to the Faculty of Science at the University of Chile (Santiago, Chile). There, the individuals were euthanized with an overdose via immersion in anesthetic solution of tricaine methanesulfonate (MS-222) and were left in

there for 10 minutes after the cessation of opercular movement (AVMA 2013). Later, the right lateral fin of each individual was cut and preserved separately in RNAlater (Thermo Fisher Scientific), and stored at -80°C for later DNA extraction. The rest of the individual was preserved in analytical alcohol, also stored separately. The DNA extraction was performed with the DNeasy Blood & Tissue Kit (Qiagen), including the treatment with RNase. We measured the DNA concentration with Qubit dsDNA BR Assay Kit (Thermo Fisher Scientific), and also we checked the DNA integrity in a 2% agarose gel. The DNA concentrations were normalized to $25\text{ ng}/\mu\text{l}$ and then were sent to Floragenex, Inc. (<http://www.floragenex.com/index.php>) where one RAD (restriction site-associated DNA sequencing) library was prepared and single-end sequenced on an Illumina Hi-Seq 2000. The extracted DNA was also used to obtain mtDNA control region sequences of each individual using the primers described in Morales *et al.* (2011) and the PCR protocol described in Cruz-Jofré *et al.* (2016). The PCR products were sent to Macrogen Inc. (Seoul, South Korea) to be sequenced in both directions.

For phylogenetic and phylogeographic analysis, the sampled individuals from other Chilean locations (Tabla 6) were euthanized in the field as mentioned above. The individuals were preserved in analytical alcohol and stored in the Faculty of Science at the University of Chile (Santiago, Chile). The DNA extraction was performed with a salt-extraction method (Aljanabi & Martinez 1997). Later, we checked the DNA integrity in a 2% agarose gel. The mtDNA control region was amplified as mentioned above. All the sequences obtained were edited and aligned with the software ProSeq (Filatov 2009).

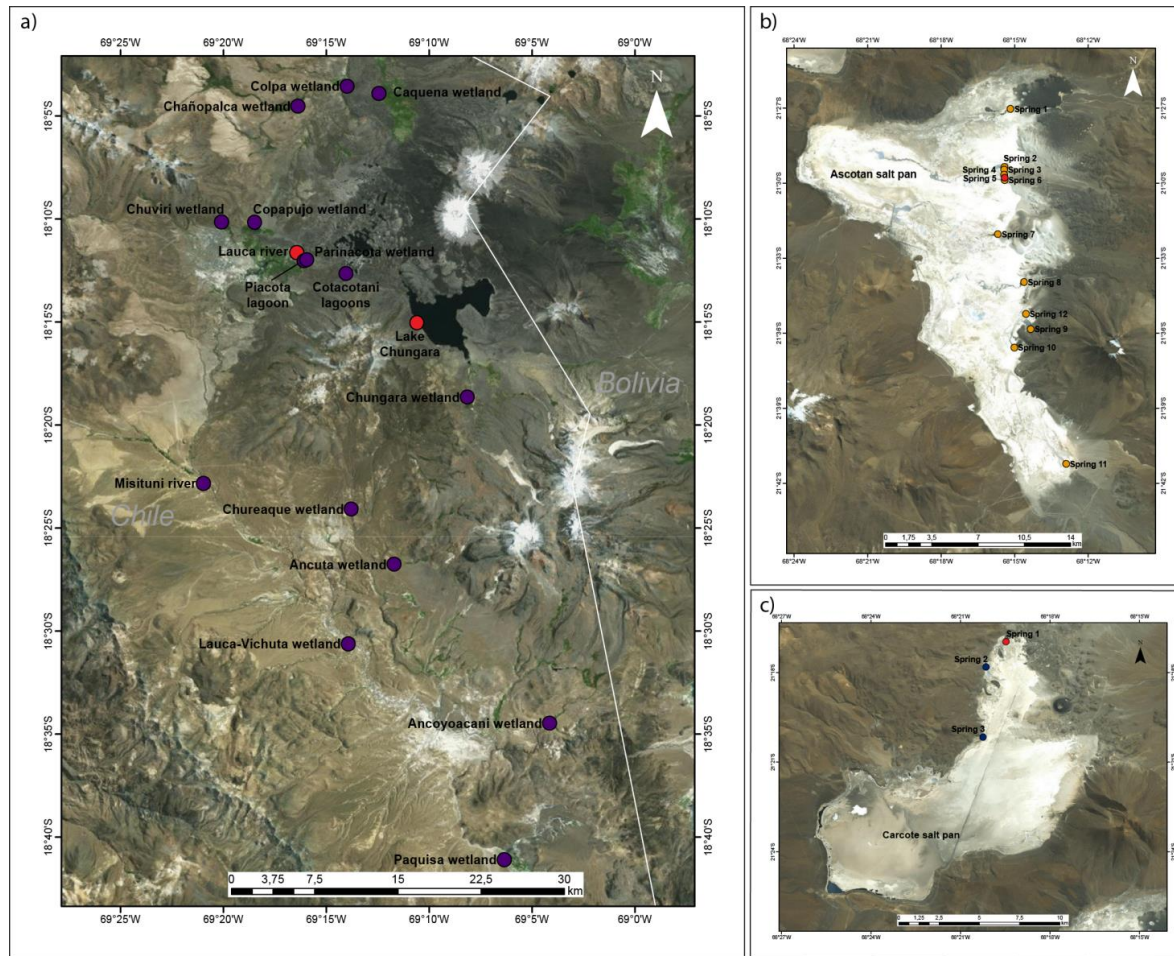


Figure 7. a) Sampled localities in the Lauca National Park (LNP) and nearby sites. b) Ascotán salt pan and the twelve springs with *O. ascotanensis* populations. c) Carcote salt pan and the three springs with *O. glorieae* populations. The red circles indicate the populations used for genomic analysis.

Table 6. Sampled sites in the Chilean distribution of *Orestias* that were included in the phylogenetic analysis, including the seven type localities and the sixteen recently discovered sites. a: Type locality; b: recently discovered site. * indicates the sample size for mtDNA control region sequences and then for RAD-Seq, respectively.

Species	Sample Size	Name of Locality	Geographic Coordinates
<i>O. chungarensis</i>	7 / 24 *	Lake Chungará ^a	18° 15' 02"S, 69° 10' 35"W
<i>O. laucaensis</i>	5 / 24 *	Lauca River ^a	18° 22' 50"S, 69° 20' 56"W
	5	Cotacotani Lagoons	18° 12' 39"S, 69° 14' 02"W
<i>O. piacotensis</i>	5	Piacota Lagoon ^a	18° 12' 01"S, 69° 16' 02"W
<i>O. parinacotensis</i>	5	Parinacota Wetland ^a	18° 12' 02"S, 69° 16' 04"W
<i>O. cf. agassii</i>	10	Isluga River ^a	19°15'22.9"S, 68°42'22.7"W
	7	Collacagua River	20°08'47.6"S 68°50'36.8"W
	7	Huasco salt pan	20°15'46.0"S 68°52'31.6"W
<i>O. ascotanensis</i>	8	Ascotán salt pan, Spring 1	21°27'02.0"S, 68°15'09.8"W
	19	Ascotán salt pan, Spring 2	21°29'21.2"S, 68°15'24.8"W
	6	Ascotán salt pan, Spring 3	21°29'27.8"S, 68°15'25.6"W
	8	Ascotán salt pan, Spring 4	21°29'39.8"S, 68°15'25.6"W
	2 / 23 *	Ascotán salt pan, Spring 5	21°29'47.2"S, 68°15'24.5"W
	5	Ascotán salt pan, Spring 6	21°29'52.8"S, 68°15'24.4"W
	16	Ascotán salt pan, Spring 7	21°32'02.7"S, 68°15'40.7"W
	16	Ascotán salt pan, Spring 8	21°33'57.2"S, 68°14'36.8"W
	7	Ascotán salt pan, Spring 9	21°35'50.6"S, 68°14'20.7"W
	8	Ascotán salt pan, Spring 10	21°36'35.0"S, 68°15'00.7"W
	8	Ascotán salt pan, Spring 11	21°41'13.90"S, 8°12'54.0"W
	5	Ascotán salt pan, Spring 12	21°35'13.88"S, 68°14'32.3"W
<i>O. gloriae</i>	11 / 24 *	Carcote salt pan, Spring 1 ^a	21°16'58.4"S, 68°19'28.3"W
	2	Carcote salt pan, Spring 2	21°17'49.1"S, 68°20'08.7"W
	6	Carcote salt pan, Spring 3	21°20'09.9"S, 68°20'14.5"W
<i>Orestias</i> sp	4	Chuviri Wetland ^b	18° 10' 09"S, 69° 20' 05"W
<i>Orestias</i> sp	5	Colpa Wetland ^b	18°03'33.5"S 69°13'58.6"W
<i>Orestias</i> sp	5	Lirima Wetland ^b	19°51'08.0"S 68°54'37.6"W
<i>Orestias</i> sp	10	Humaquilca Wetland ^b	17°44'12.6"S 69°23'26.2"W
<i>Orestias</i> sp	5	Ancoyoacani Wetland ^b	18° 34' 29"S, 69° 04' 08"W
<i>Orestias</i> sp	5	Chureaque Wetland ^b	18° 24' 06"S, 69° 13' 47"W
<i>Orestias</i> sp	6	Copapujo Wetland ^b	18° 10' 09"S, 69° 18' 28"W
<i>Orestias</i> sp	5	Paquiza Wetland ^b	18° 41' 06"S, 69° 06' 20"W

<i>Orestias</i> sp	5	Ancuta Wetland ^b	18° 26' 46"S, 69° 11' 41"W
<i>Orestias</i> sp	5	Caquena Wetland ^b	18° 03' 55"S, 69° 12' 26"W
<i>Orestias</i> sp	6	Chañopalca Wetland ^b	18° 04' 32"S, 69° 16' 22"W
<i>Orestias</i> sp	5	Misitune River ^b	18° 22' 50"S, 69° 20' 58"W
<i>Orestias</i> sp	5	Surapalca ^b	17° 48' 26"S, 69° 42' 53"W
<i>Orestias</i> sp	2	Chungará Wetland ^b	18° 18' 39"S, 69° 08' 08"W
<i>Orestias</i> sp	5	Cotaculco Wetland ^b	19°38'28,42"S, 68°45'47,6"W
<i>Orestias</i> sp	5	Lauca-Vichuta Wetland ^b	18°30'37,5"S 69°13'54,9"W

3.2 Phylogeny and phylogeography of mtDNA control region sequences.

For the phylogeny of Chilean species of *Orestias* we constructed a phylogenetic tree with the Maximum Likelihood approach (ML) and Bayesian inference (BI) for all the Chilean populations, including the seven species described for Chile, the sixteen newly discovered populations, and the RAD individuals (Tabla 6). As outgroups, we use three individuals of *O. puni* from the Titicaca Lake, and three individuals of *Aphanius sophiae* (Iran), as the sister group of *Orestias* (Parenti 1981). For the ML, we used RAxML (Stamatakis 2014) using a GTR + gamma as model of nucleotide substitution and 1,000 iterations for rapid bootstrap. For BI, we estimated the best substitution model with jModelTest 2.1.10 (Darriba *et al.* 2012). The model chosen was TPM1uf+I+G. We used MrBayes v3.2.6 (Ronquist *et al.* 2012) to run four MCMC in each of four independent runs, with 60 million of generations, trees sampled every 1,000 generations, and a burn-in of 25%. Both analyses were performed in CIPRES (Cyberinfrastructure for Phylogenic Research) Science Gateway V. 3.3, available in <https://www.phylo.org/> (Miller, M.A., Pfeiffer, W., and Schwartz, T., 2010).

The haplotype networks were performed per each clade where the RAD individuals were located in the phylogenetic trees: *O. ascotanensis* located within the CHAL clade, *O. glorieae* in its own clade, and LNP clade, where both *O. laucaensis* and *O. chungarensis* are included. We used the software PopART v. 1.7 (Leigh & Bryant 2015) with the median-joining algorithm (Bandelt *et al.* 1999) to constructed the networks.

3.3 Data preparation and variant, genotype and SNP calling.

All the analysis of RAD-Seq data were performed in the Genetic Diversity Centre (GDC), ETH Zurich.

The raw data of the sequenced library contained 173 million of 100 bp reads. First, we removed the reads of the bacteriophage Phix (8.71% of the total raw reads) that was added to calibrate sequence quality scores. Then, we extracted all reads that have a complete *SbfI* site within the first 16 nucleotides (10 nucleotides of barcodes plus the 6 nucleotides of the restriction site); these corresponded to 145 million of reads. Later, we demultiplexed these reads using the `process_radtags` v1.26 from Stacks (Catchen *et al.* 2011, 2013) in order to obtain the reads of each individual. And, before the mapping to the genome, we preprocessed the reads based on their quality: we kept only reads with 100% of the bases with quality score of at least 10 and removed all reads with at least 5% of the bases with quality score below 20, by applying the FASTQ Quality Filter tool of FASTX-Toolkit (available in http://hannonlab.cshl.edu/fastx_toolkit/index.html). After this step, we retrieve 118 million of high quality reads considering the 95 individuals. In parallel, we index the annotated reference genome of *O. ascotanensis* (<http://orestias-crg.cmm.uchile.cl/>) with SAMtools v0.1.19 (Li *et al.* 2009) and the `faidx` option, and then with BOWTIE 2 v2.2.6 (Langmead *et al.* 2009) and the option `bowtie2-build`. Next, we aligned the high quality reads of each individual to the indexed genome using BOWTIE 2 v2.2.6 performing an end-to-end alignment and the default parameters. SAMtools v0.1.19 (Li *et al.* 2009) was used again to transform the alignments into binary format. Later, we performed a base-quality recalibration, which included two steps: first, we calculated an empirical error estimation based on 1.5 billion base pairs of PhiX reads, using the tool BaseRecalibrator of GATK v.2.7 (McKenna *et al.* 2010; DePristo *et al.* 2011; Van der Auwera *et al.* 2013). Second, the recalibration table produced was used to recalibrate each individual with the tool PrintReads of GATK v.2.7. The next step was the variant and genotype calling, which was performed with the tool UnifiedGenotyper of GATK v.2.7. Additionally, we considered a contamination of 3% to be removed, a minimum base quality score of 20 to be considered for calling, and the genotype likelihoods calculation mode which allows calling for both SNPs and insertions/deletions (indels). Then, we filtered the resulting dataset to get good quality SNPs by using

VCFtools v0.1.12b (Danecek *et al.* 2011) applying the following filters: minimum quality value of sites of 30 (--minQ 30), maximum proportion of missing data per site of 80% (--max-missing 0.2), two alleles per site as maximum (--max-alleles 2) to remove multiallelic sites, sites with minimum mean depth (over all included individuals) greater than or equal to 10 (--min-meanDP 10), minimum genotype quality of 20 (--minGQ 20), minimum genotype sequencing depth of 10 (--minDP 10), minor allele count greater than or equal to 1 (--mac 1), and we removed indels and surrounding SNPs in a distance of 10 bases. Finally, we filtered out those sites with mean depth averaged across all individuals higher than 60, because they could be paralogous regions.

We performed the same procedure, from the variant and genotype calling step, to the datasets conformed by the individuals of each species by separated, and to the dataset comprising the individuals of *O. chungarensis* and *O. laucaensis* (48 individuals in total). The maximum mean depth per site averaged across all individuals was 65 for these dataset, except for the *O. glorieae* dataset (60) and for the dataset comprising the young species (50).

Additionally, we applied filters for specific analysis, which are described in the following sections.

3.4 RAxML.

In order to work with an appropriate SNP dataset to perform this phylogenetic analysis, we applied the following filters with VCFtools v0.1.12b to the dataset obtained in the previous section: we fixed a genotype depth cutoff at 3 (--minDP 3), we considered sites with Minor Allele Count greater than 1 (--mac 1) to keep alleles of single divergent individual, and we did not apply the --max-alleles filter to retain the multiallelic sites. The vcf file was transform into the Phylip format with a custom python script.

We constructed a phylogeny for the 95 RAD individuals with RAxML version 8.2.4, using 100 rapid bootstrap inferences and thereafter a thorough ML search with the GTR + gamma model.

3.5 Principal Component Analysis (PCA) between species.

We applied the following filters to the vcf file with VCFtools v0.1.12b: minor allele frequency filter of 0.05 (--maf 0.05), we kept sites with quality higher than 30 (--minQ 30) and excluded sites with a proportion of missing data greater than 70% (--max-missing 0.3). The PCA was performed in R (R Core Team 2015) with the package SNPrelate (Zheng *et al.* 2012).

3.6 Hierarchical Structure Analysis.

For this analysis we worked with two data files: one comprising the four species, a second one with the individuals of the LNP species only, *O. chungarensis* and *O. laucaensis*. The initial dataset was processed with VCFtools v0.1.12b to keep sites with quality higher than 30 (--minQ 30), to exclude sites with a proportion of missing data higher than 20% (--max-missing 0.8) and to remove sites that are closer than 1,000 base pairs, hence sites that were strongly linked were not considered (--thin 1000). Then, the vcf files were transformed into a *structure* format files with PGDSpider2 (Lischer & Excoffier 2012).

We use the Bayesian clustering of the software *structure* version 2.3 (Pritchard *et al.* 2000). We checked from two to seven groups to estimate the most likely number of clusters that fit the model. 100,000 iterations were run for ten Markov chain Monte Carlo after a burn-in period of 10,000, under the admixture model.

To handle with the differences in the clustering outcomes we use CLUMPP (Jakobsson & Rosenberg 2007) to find the optimal alignment of the ten replicates, in such a way that all replicates have as close a match as possible. Additionally, we obtained a mean of the permuted matrices across replicates. The algorithm used was *FullSearch*. The most likely number of clusters was determined by consensus among replicates. To visualize the results of *structure* and CLUMPP we use the program *distruct* version 1.1 (Rosenberg 2003).

To estimate the genetic structure in these species, we follow a hierarchical strategy as in Roy *et al.* (2015): first, we identify the number of genetic cluster in the whole dataset. Then, the individuals assigned to each cluster were independently re-assessed. This process was repeated until there was no further subdivision.

3.7 Pairwise F_{ST} between species.

We estimated the pairwise F_{ST} values between pair of species, i.e., between salt pan species, *O. ascotanensis* and *O. gloriae*, and between LNP species, *O. chungarensis* and *O. laucaensis*. Pairwise F_{ST} values were calculated in Arlequin version 3.5.2.2 (Excoffier & Lischer 2010), and their significance was estimated after 10.000 permutations.

4. Results

4.1 Phylogeny and phylogeography of Chilean *Orestias*.

We sequenced the mtDNA control region of 356 individuals in total, including 178 individuals of the seven described species in Chile, the 95 RAD individuals and 83 individuals from sixteen Chilean, newly discovered localities (Tabla 6, Figura 7). For the phylogenetic reconstruction of Chilean *Orestias* populations we kept only non-redundant sequences, but we do retained all the sequences of the RAD individuals, hence this analysis was based on 290 sequences of the mtDNA control region of 844 bp length. We found three mayor clades in the ingroup (Figura 8), CHAL clade (Collacagua, Huasco, Ascotán and Lirima populations), *O. gloriae* clade and LNP clade. All the RAD individuals are recovered in one of these lineages (Figura 8): all the RAD individuals from Carcote salt pan, Spring 1, are located in the *O. gloriae*, Spring 1, clade (Figure S1a), the RAD individuals from Lauca River and Lake Chungará are within the LNP clade (Figure S1b), and all the RAD individuals sampled in Ascotán salt pan are located in the CHAL clade, closely related to other *O. ascotanensis* individuals (Figure S1c). This phylogeny shows a very deep divergence between the salt pan samples, Ascotán and Carcote, and between

these and the Lake Chungará and Lauca River samples, and much less deep divergence between lake and stream.

Within the *O. glorieae* clade (Figure S1a), there are two lineages, one with the individuals from Spring 1 and the other with the individuals from Spring 2 and 3. The described species from the LNP are all located in the same well-supported monophyletic group, the LNP clade (Figure S1b). *O. chungarensis* is the only species that is recovered in one lineage (that also contains one individual of *O. laucaensis*), yet without statistical support; the other three LNP species (*O. laucaensis*, *O. parinacotensis* and *O. piacotensis*) are not recovered as monophyletic groups. Within this clade there are also seven newly discovery localities: Chuviri, Copapujo, Ancoyoacani, Caquena, Chañopalca and Chungará wetlands and Misitune River. The CHAL clade contains individuals from Collacagua River, Huasco salt pan, Ascotán salt pan and Lirima wetland (Figure S1c). This clade is unsupported, but it recovered the same populations as clade D in Vila et al. (2013), except for Lirima which was not included in that publication, and Humaquilca (called Umaqui in Vila et al. (2013)). Collacagua and Huasco contain *O. cf. agassii*, and they are closely related within the CHAL clade. This species is also described in Isluga River (but see Cruz-Jofré et al. 2014), however the individuals from this site are out of the CHAL clade and they are found in two different lineages which are both distantly related to each other and to the CHAL clade (Figura 8). Additionally, this clade is closely related to five recently discovered sites: Paquisá, Ancuta, Lauca-Vichuta, Chureaque, and Cotaculco wetlands. On the other hand, the phylogenetic position of Humaquilca, Colpa and Surapalca wetlands is undefined. There is a lineage closely related to the *O. glorieae* clade, but there also are two more clades unrelated to each other and the mentioned lineage.

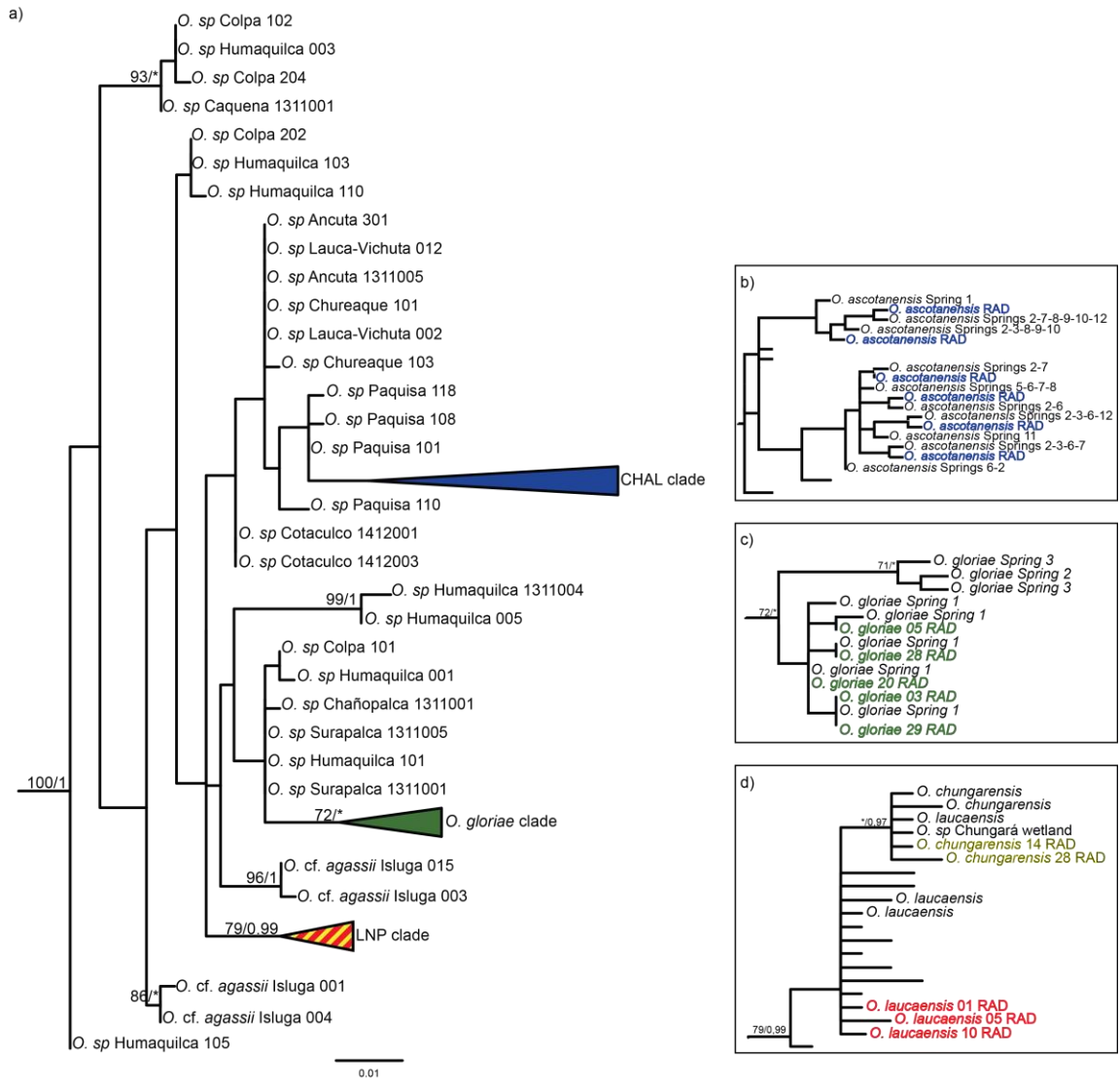


Figura 8. Maximum likelihood phylogenetic reconstruction for the mtDNA control region sequences of Chilean populations of *Orestias*. Bootstrap values >70% (ML) and posterior probabilities (BI) >0.95 of each node is indicated (* : no support). In a) it is shown the ingroup, presenting the three main collapsed clades: b) the CHAL clade, c) the *O. glorieae* clade, and d) the LNP clade. b), c), and d) are schematic representations of each of these clades, showing the close relationship between the RAD individuals and each of their own species (for simplicity, there are not indicated any other species or population within each of these clades; see the Figure S1 for the complete uncollapsed clades).

The haplotype networks (Figura 9) shows that the RAD individuals belong to the same haplogroups where there are individuals sampled in the same sites in previous field works. In the case of *O. ascotanensis* (Figura 9a), twelve RAD individuals share haplotypes with individuals from their same spring (spring 5) of the Ascotán salt pan, four RAD individuals share haplotypes with individuals from springs 2 to 7 (except spring 5), and seven RAD individuals represent six new haplotypes. In the case of *O. gloriae* (Figura 9b), the four haplotypes found in the RAD individuals are only shared with individuals that were sampled in the same spring 1 of Carcote salt pan. Similarly, the haplotype network of the LNP clade (Figura 9c) show that the RAD individuals of *O. laucaensis* and *O. chungarensis*, only share haplotypes with individuals from their same species. Additionally, it shows that these species are well differentiated (pairwise $\Phi_{ST} = 0.499$, p-value = 0.001).

These results indicate that the RAD individuals are representatives of each of the four species. And also, confirms the genetic patterns found in these species in earlier works (Morales *et al.* 2011; Vila *et al.* 2013; Cruz-Jofré *et al.* 2016; Guerrero-Jiménez *et al.* 2017).

4.2 Genomic differentiation in pairs of *Orestias* species.

The analysis of RAD-Seq data from the four species of *Orestias* shows that they are highly differentiated species. The RAxML analysis based on the RAD-Seq data (Figura 10) reveals strong reciprocal monophyly of the individuals in each of the four water bodies. The PCA analysis (Figura 11) shows that the three first principal components (65.69% of the explained variance) differentiated the four species: the PC1 clearly separated the recent species pair of the ancient species pair, the PC2 differentiated the salt pan species (Figura 11a), while the incipient LNP species were separated in the PC3 (Figura 11b). Both the RAxML and PCA analyses show that each species is well-differentiated, and that the differentiation is higher between the salt pan species than between the LNP species. This is supported by the pairwise Φ_{ST} value between the salt pan species (pairwise $\Phi_{ST} = 0.66972$, p-value < 0.00001), which is almost three times higher than the pairwise Φ_{ST} value for the LNP species (pairwise $\Phi_{ST} = 0.24164$, p-value < 0.00001).

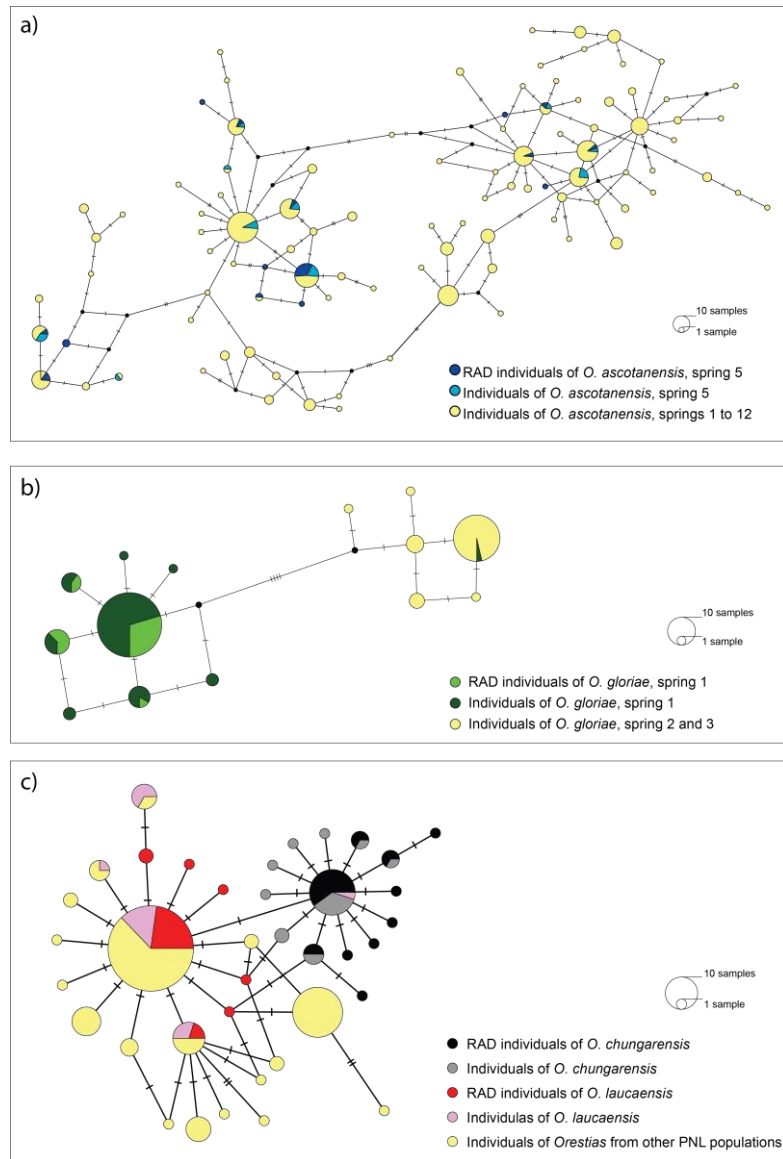


Figura 9. Haplotype networks of the control region mtDNA sequences of a) *O. ascotanensis*, b) *O. gloriæ* and c) *O. chungarensis* and *O. laucaensis* along all the LNP populations.

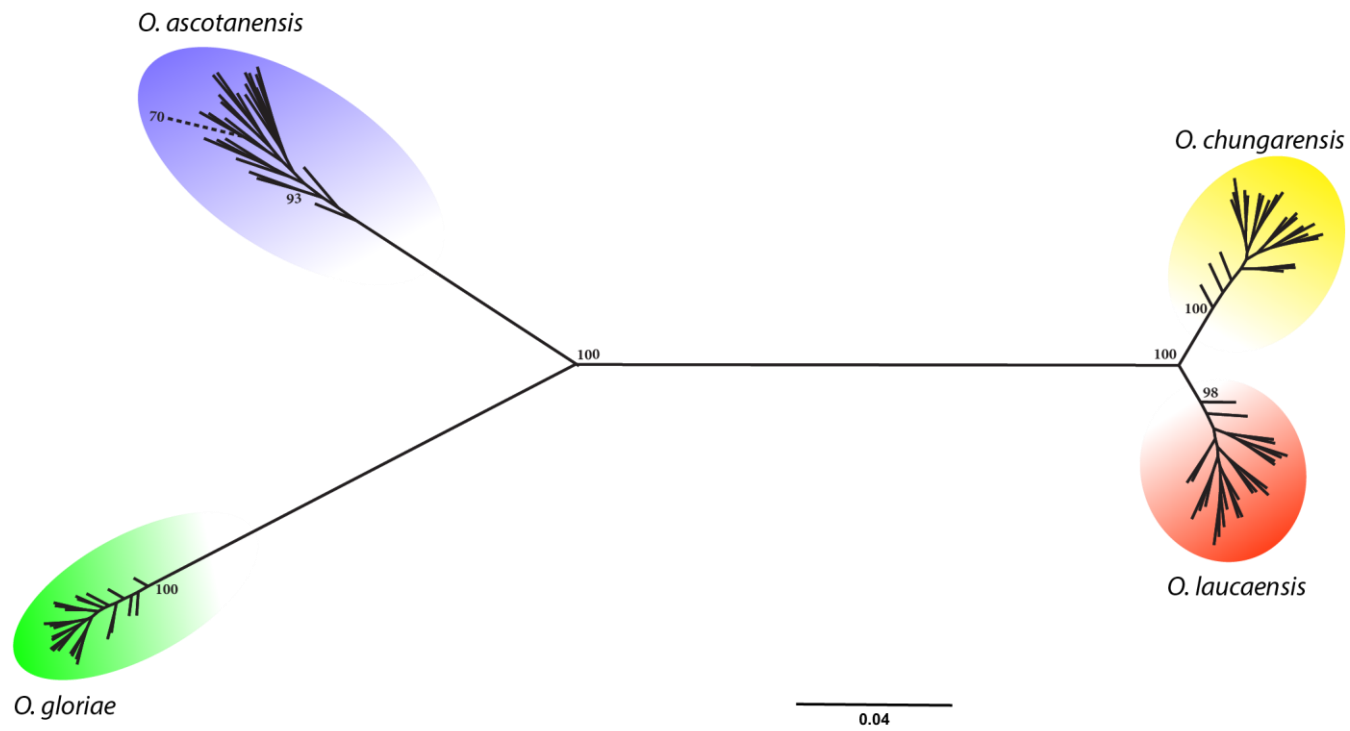


Figura 10. RAxML phylogenetic reconstruction of the RAD individuals, comprising *O. ascotanensis* (blue), *O. gloriae* (green), *O. chungarensis* (yellow) and *O. laucaensis* (red).

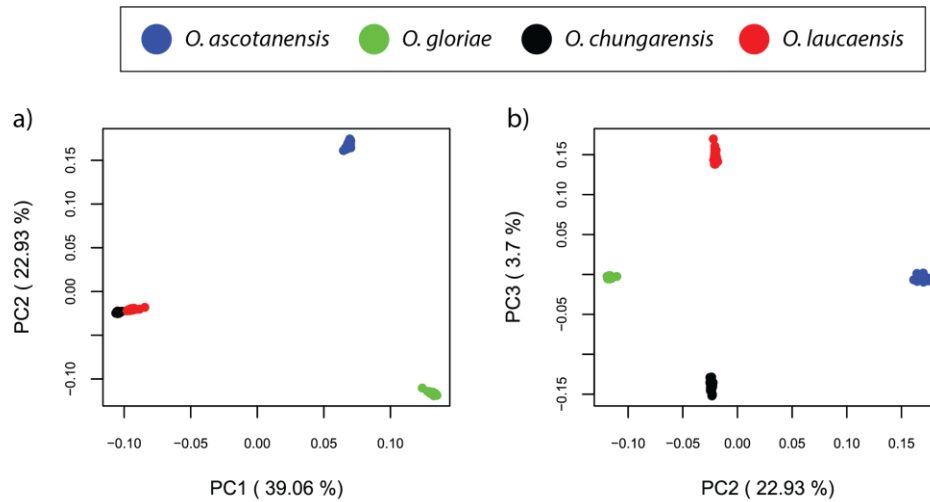


Figura 11. Principal Component Analysis (PCA) for the RAD individuals of the four species of *Orestias*. Variance explained by each principal component (PC) is indicated in parenthesis. a) PC1 versus PC2, and b) PC2 versus PC3.

Hierarchical *structure* analysis (Figura 12) confirms a very strong differentiation between the four water bodies. The best subdivision for the entire dataset was in $K = 3$, where the two salt pan species were separated in different cluster, while the third one put together the LNP species, with an assignment probability of 100% for each individual to their respective cluster. Subsequently, the analysis of a dataset comprising only the two LNP species shown that the best subdivision is in $K = 2$, where each species formed a genetic cluster, again with a 100% of assignment probability for each individual to their respective cluster.

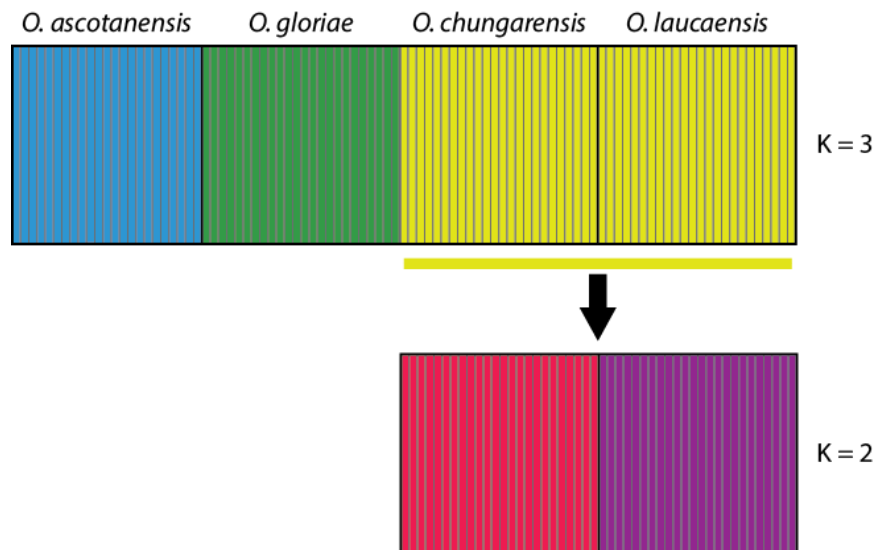


Figura 12. Hierarchical *structure* analysis, showing the number of genetic cluster at each step of the analysis. Each individual is represented as a bar and the different color represents the assignment probability to different clusters. First, the complete dataset was analyzed, founding 3 clusters. Then, the genetic cluster comprising the LPN species was analyzed, obtaining a best grouping of 2, where each cluster was formed by each species.

5. Discussion.

In this work we compared the genetic structure at different stages of the speciation continuum considering four allopatric species of *Orestias*, two of them are ancient, deeply divergent species and the other two have recently diverged.

Phylogenetic and phylogeographic analysis based on the mtDNA control region allowed us to recognize and to assign the RAD individuals to the correct species: they were recovered in the clades where there are other individuals of the same species, and showed similar haplotypic diversity to the same populations where they were sampled (Figura 8 y 9). In the case of *O. ascotanensis* (Figura 9a), four RAD individuals share haplotypes with individuals from springs 2 to 7, except spring 5 (their collection site), and seven RAD individuals are six new haplotypes. Previous studies showed that the individuals from springs 2 to 7

form one genetic group, and the presence of new haplotypes in the RAD individuals is expected due the high diversity observed in these springs (Morales *et al.* 2011; Cruz-Jofré *et al.* 2016). Consequently the RAD individuals are a representative sample of each species. Additionally, and despite the use of a highly variable marker as the mtDNA control region that caused several polytomies and nodes without statistical support, we recovered mainly the same groups as in Vila *et al.* (2013): the LNP clade, the *O. gloriae* clade and the CHAL clade, are equivalent to the Clade B, Lauca, the Clade C, Carcote, and the Clade D, Ascotán/Huasco/Collacagua in Vila *et al.* (2013), respectively. Future phylogenetic analysis of *Orestias* should consider the whole distribution of the genus in order to reveal the relationship of the described species and recently discovered populations, especially for unveils the connections between distant populations: it is likely that those populations whose phylogenetic positions are undetermined could be more related to Bolivian populations. At the present we cannot assure which Chilean and Bolivian populations are or were connected, mainly due to the complex orogeny of the region and the lack of geological and hydrological information beyond the paleolakes descriptions. To expand the number of molecular markers would also help to improve the resolution of the lineages, by using additional mtDNA marker as *cytb* and *ND2* that has shown to be useful in the resolution of these lineages (Vila *et al.* 2013) or SNPs data obtained from applying genomic technologies, that have proved to be useful even to shallow phylogenies (Wagner *et al.* 2013; Eaton & Ree 2013; Takahashi & Moreno 2015; Harvey *et al.* 2016).

The analysis of the RAD data showed a clear differentiation between the four species, where the divergence was larger between the ancient species than those that recently diverged. The clear differentiation between *O. chungarensis* and *O. laucaensis* is an important result of this work, because support previously obtained results: they were described as different species based on their different chromosomal number (Arratia 1982), and then also on their chromosomal formulae (Araya 2009; Vila *et al.* 2010). Then, in the phylogenetic analysis of Vila *et al.* (2013) there was not enough resolution to differentiate the species from the LNP in different lineages. However, Guerrero-Jiménez *et al.* (2017) described these species as differentiated groups but with a low pairwise Φ_{ST} value based on nucleotide pairwise differences of mtDNA control region sequences (pairwise $\Phi_{ST} = 0.49$, p-value < 0.0001; similar value to that of this study pairwise $\Phi_{ST} = 0.499$, p-value < 0.0001) and on eight microsatellites markers

(pairwise $F_{ST} = 0.09$, p-value < 0.0001). Our results of RAD-Seq support the idea that they are not just slightly different, but well-differentiated groups: the RAxML phylogeny show them as two different, well-supported lineages, the PCA separated them in differentiated clusters, *structure* assigned their individuals in two different genetic clusters with 100% of assignment probabilities, and there is a significant Φ_{ST} value between them (Φ_{ST} pairwise = 0.24164, p-value < 0.0001). Even when the Φ_{ST} pairwise value is smaller than that estimated with mtDNA control region (only one marker), the estimation with RAD-Seq implied the use of 3,576 SNPs. This suggest that the molecular markers used in previous works had limited resolution (Vila *et al.* 2013; Guerrero-Jiménez *et al.* 2017), and it was necessary a genomic scan to reveal the strong genetic divergence between these species.

The RAD-Seq technique showed to be more resolute than other genetic markers in identifying differences in two recently diverging groups. For this reason, this technique could be applied to every described species and newly discovered population of *Orestias*, helping to unfold the phylogenetic relationships and to determine the actual diversity in this genus, which would be an important contribution to the systematic of this group. Additionally, it could help to understand how the speciation process has occurred in *Orestias*, and to recognize the evolutionary mechanisms that has act at different stages of the continuum speciation.

6. Supporting Information



Figure S1. a) The *O. gloriæ* clade. The RAD individuals of *O. gloriæ* are shown in green. b) LNP clade comprising the populations from the Lauca National Park; the RAD individuals of *O. chungarensis* are shown in yellow, while the RAD individuals of *O. laucaensis* are shown in red. c) CHAL clade including the populations from Collacagua river, Huasco salt pan, Ascotán salt pan and Lirima wetland; the RAD individuals of *O. ascotanensis* are shown in blue.

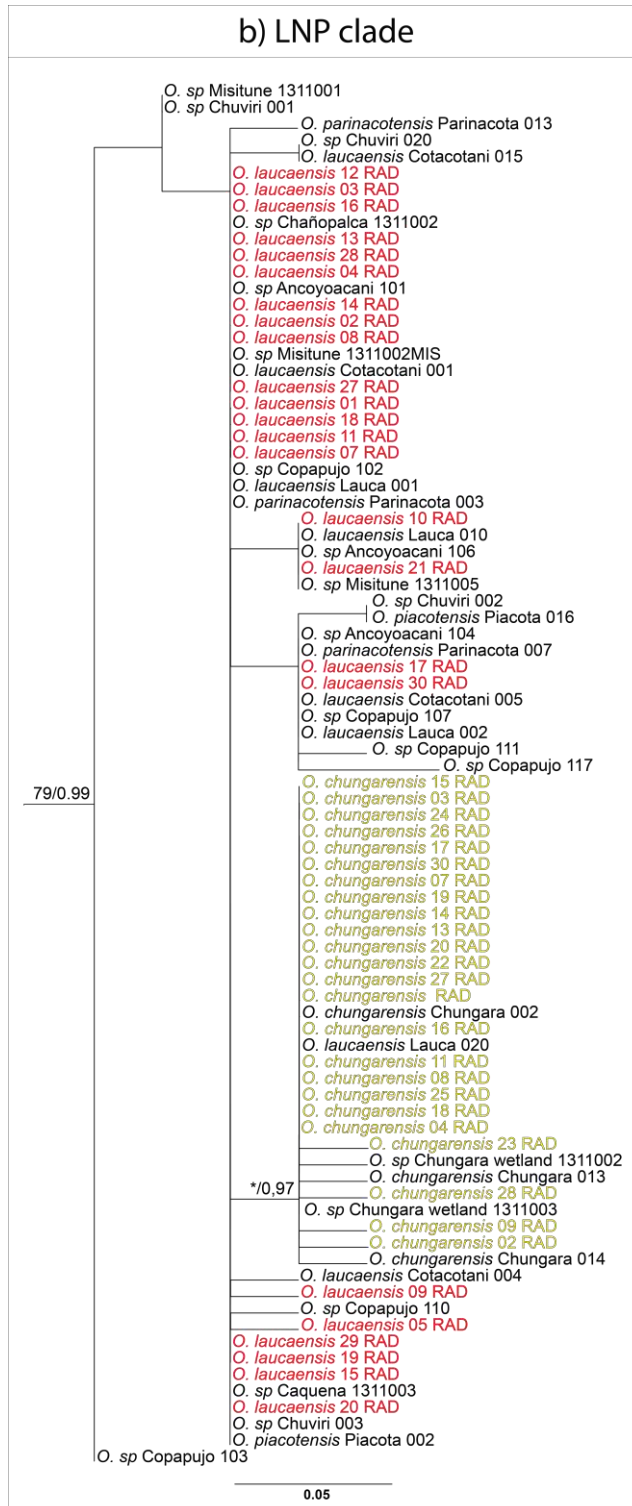


Figure S1. (continuation)

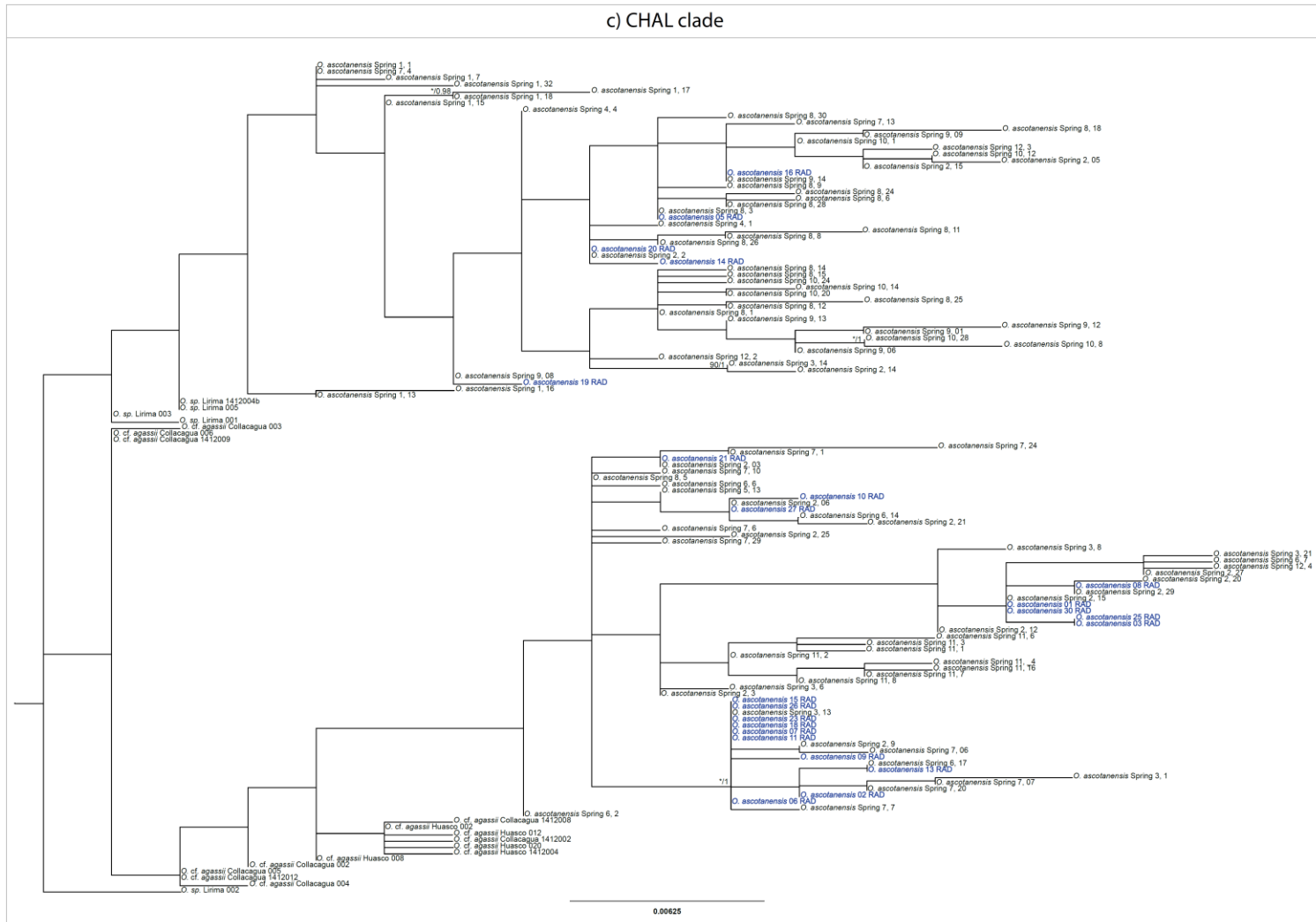


Figure S1. (continuation)

CAPITULO III.

**ANÁLISIS COMPARATIVO DE LOS PATRONES DE DIFERENCIACIÓN GENÓMICA EN
CUATRO ESPECIES DE *ORESTIAS*.**

1. Resumen

La descripción de los patrones genómicos de diferenciación en distintas etapas del continuo de especiación ha ayudado a determinar la proporción del genoma que contribuye a la divergencia y la naturaleza de los genes o zonas genómicas involucradas. En este trabajo se describen los patrones de diferenciación a partir de datos de RAD-Seq en dos etapas del continuo de especiación considerando cuatro especies de *Orestias*: a nivel de especies que divergieron recientemente, *O. laucaensis* y *O. chungarensis*, y a nivel de especies muy divergentes, *O. ascotanensis* y *O. gloriae*. Estas cuatro especies se distribuyen de forma alopátrica y en ausencia de flujo génico. Los resultados indicaron que tanto los niveles de diferenciación global como el número de loci con F_{ST} significativo aumentan a medida que se avanza en el continuo de especiación. En cuanto al número total de SNPs, se detectó (i) una baja proporción de éstos compartidos entre especies, siendo esta proporción mayor entre especies recientes que entre especies lejanas, y (ii) que éstos se ubican principalmente en zonas génicas. Por otra parte, los loci que más contribuyen a la diferenciación de las especies también serían específicos para cada par de especies y la mayoría también se ubican en regiones génicas. Estos resultados indicarían que el proceso de diferenciación ha seguido un camino independiente en cada etapa del continuo de especiación.

2. *Introducción*

Una pregunta central del estudio de la especiación es conocer la secuencia en la cual las barreras reproductivas emergen a lo largo del continuo de especiación. Debido a que no es posible observar directamente el proceso de especiación en poblaciones naturales de organismos multicelulares de reproducción sexual (Seehausen *et al.* 2014), es que se han realizado estudios de poblaciones en diferentes estados del continuo de especiación para conocer la cronología y la importancia relativa de los procesos evolutivos actuando en cada etapa (Seehausen *et al.* 2014; Wolf & Ellegren 2017). De forma más específica, la descripción de los patrones genómicos de diferenciación en estas distintas etapas ha ayudado a determinar la proporción del genoma que contribuye a la divergencia y la naturaleza de los genes o zonas genómicas involucradas (Strasburg *et al.* 2012; Seehausen *et al.* 2014).

En un escenario de especiación alopátrica se espera que la divergencia afecte a la mayoría del genoma debido a la ausencia del efecto homogeneizador del flujo génico (Coyne & Orr 2004), a diferencia de especiación en presencia de éste donde la divergencia ocurriría sólo en algunas regiones que contienen genes que están bajo un régimen de selección divergente, mientras que el resto de genoma es homogeneizado por el flujo génico (Nosil *et al.* 2012). Es así que la diferenciación genética presentará una distribución de frecuencias “tipo L”, es decir, la mayoría de los loci presentan bajos niveles de F_{ST} (Nosil *et al.* 2012; Ryan *et al.* 2017). En alopatría, prácticamente cualquier fuerza que causa divergencia puede eventualmente provocar especiación (Coyne & Orr 2004). Entre estos factores se encuentran selección divergente, una respuesta diferencial a presiones selectivas similares, deriva génica, o una combinación de éstos (Coyne & Orr 2004; Nosil *et al.* 2009a, 2012; Strasburg *et al.* 2012). De esta manera, las poblaciones se diferenciarían de manera independiente unas de otras, las diferencias genéticas se acumularían a una tasa más rápida que una acumulación a tasa lineal, efecto conocido como “bola de nieve” que genera aislamiento reproductivo (Orr 1995), con lo que la probabilidad de especiación aumentaría medida que aumenta el tiempo de divergencia (Orr 1995; Wolf & Ellegren 2017). La forma más tradicional de determinar el grado de diferenciación entre poblaciones es a través de la estimación del índice de estructuración F_{ST} (Wright 1931, 1949), ya sea a nivel global como a nivel de cada locus o SNP. En este trabajo se describen los patrones de diferenciación en dos etapas del continuo de especiación, una en que

dos especies divergieron recientemente (12-8 ka; Guerrero-Jiménez *et al.* 2017), *O. laucaensis* y *O. chungarensis*, versus dos especies cuyo tiempo de divergencia es mayor (> 300 ka; Scott 2010), *O. ascotanensis* y *O. gloriae*; estas cuatro especies se distribuyen de forma alopátrica y en ausencia de flujo génico. Esta descripción se realizó sobre la base de diferenciación global, es decir el nivel de diferenciación total considerando todos los sitios de cada par de especies, y la diferenciación local o de cada SNP. Luego, conociendo las similitudes y diferencias entre los patrones de diferenciación de cada especie, se determinaron las regiones del genoma donde se han generado los polimorfismos y los procesos biológicos enriquecidos que estarían involucrados en la diferenciación de cada especie. Para esto se llevó a cabo una comparación interespecífica del número total de SNPs, en busca de aquellos que son únicos en cada especie, entendiendo cada uno de estos sitios como polimorfismo o diversidad que se ha generado de forma exclusiva en cada especie y que la diferencia de las otras. Finalmente, y una vez que ya se conocían las diferencias y similitudes entre los patrones de diferenciación, la ubicación en el genoma de los SNPs exclusivos y los procesos biológicos implicados en la diferenciación de cada especie, la siguiente búsqueda se centró en los SNPs que más contribuyen a la diferenciación de cada par de especies, ya sea los sitios presenten en una o en ambas etapa del continuo de especiación, y se identificaron los procesos biológicos en los que dichos sitios estarían involucrados. De esta manera, logramos describir detalladamente los patrones de diferenciación y los procesos biológicos implicados en cada una de las etapas del continuo de especiación.

3. Metodología

En este capítulo se utilizó sólo el set de datos de RAD-Seq. Los filtros aplicados fueron los mismos que se aplicaron en el capítulo anterior, excepto que se utilizó una proporción máxima de missing data por sitio de 20% (--max-missing 0.8), y se eliminaron todos los sitios cuya profundidad media por sitio (mean depth per site) promediada en todos los individuos fue mayor que 60. Estos filtros se aplicaron utilizando VCFtools v0.1.12b (Danecek *et al.* 2011).

3.1 Estimación de los valores de F_{ST} global y F_{ST} por cada sitio.

Para estimar el nivel de divergencia genómica primero se estimaron los niveles de F_{ST} global en cada etapa del continuo de especiación. Luego, se estimaron los valores de F_{ST} para cada locus en cada una de estas etapas: especies recientes (*O. chungarensis* y *O. laucaensis*) y especies muy divergentes (*O. ascotanensis* y *O. gloriae*). Ambos análisis se realizaron en el programa Arlequin 3.5 (Excoffier & Lischer 2010) mediante un análisis de AMOVA estándar y para cada locus por separado, incluyendo la variación intra-individual. Se utilizaron 10.000 permutaciones para estimar la significancia del valor de F_{ST} ; los valores de p se corrigieron con el método FDR (False Discovery Rate; Benjamini & Hochberg 1995) implementado en la función *p.adjust* del paquete *sva* (Leek *et al.* 2012) de R (R Core Team 2015). Se graficaron la distribución de los valores de F_{ST} de cada locus y el valor de F_{ST} de cada locus ubicado dentro de su respectivo scaffold, utilizando los paquetes *gtools* (<https://CRAN.R-project.org/package=gtools>), *ggplot2* (Wickham 2009) y *reshape* (Wickham 2007) de R (R Core Team 2015). Las distribuciones de los valores de F_{ST} se compararon mediante una prueba de los rangos con signo de Wilcoxon.

3.2 Estimación de SNPs y categorías de Gene Ontology comunes y únicas por especie.

En este análisis se utilizaron los set de datos de cada una de las especies por separado. Se eliminaron todos aquellos sitios que se ubicaban a una distancia menor a 1.000 pb (--thin 1000), de manera de utilizar sólo sitios que no debieran estar ligados, utilizando VCFtools v0.1.12b (Danecek *et al.* 2011). Se realizaron diagramas de Venn para estimar el número de sitios únicos y comunes entre especies dentro de cada etapa del continuo de especiación, con el paquete *gplots* (Warnes *et al.* 2016) en R (R Core Team 2015). Se espera que el número de sitios comunes entre especies esté directamente relacionado con el tiempo de divergencia de las especies, mientras que los SNPs únicos a cada especie representan polimorfismo que distingue a cada una de ellas. Para verificar si los SNPs únicos representan vías o funciones génicas similares o diferentes, éstos fueron mapeados en el genoma de referencia de *O. ascotanensis* para su

identificación, mediante el uso de un script personalizado, obteniendo sitios en genes anotados y sitios en zonas intergénicas. Con el número de sitios que fueron clasificados en genes anotados se realizaron dos análisis: (i) en PANTHER (Protein ANalysis THrough Evolutionary Relationships; Mi *et al.* 2013, 2017) se realizó un análisis de lista de genes, utilizando el símbolo o nombre de los genes como identificador (ID), para estimar el número de genes con y sin ID conocido; (ii) con el conjunto de genes anotados con ID conocido y sus términos de Gene Ontology (GO) se realizó un análisis de enriquecimiento génico con la herramienta GOrilla (Gene Ontology enRIchment anaLysis and visuaLizAtion tool; Eden *et al.* 2007, 2009). En ambas bases de datos se utilizó la anotación del genoma humano.

3.3 Detección de sitios más diferenciados entre especies.

Para detectar cuáles son los sitios que más contribuyen a la diferenciación de las especies se realizó un análisis discriminante de componentes principales (Discriminant Analysis of Principal Components, DAPC; Jombart *et al.* 2010) utilizando el paquete adegenet (Jombart *et al.* 2008) en R (R Core Team 2015). DAPC es un método multivariado libre de supuestos, que primero realiza un análisis de componentes principales transformando las variables originales (alelos) a variables sintéticas llamadas funciones discriminantes (combinaciones lineales de las variables originales que muestran la mayor varianza entre grupos y la menor varianza dentro de grupos), y luego realiza un análisis discriminante de los componentes principales retenidos. De esta manera, maximiza las diferencias entre grupos, mientras minimiza las diferentes dentro de cada grupo.

Este análisis se corrió para tres set de datos diferentes que incluían (i) las dos especies recientes, (ii) las dos especies muy divergentes, y (iii) las cuatro especies. También se eliminaron todos aquellos sitios que se ubicaban a una distancia menor a 1.000 pb (--thin 1000) utilizando VCFtools v0.1.12b (Danecek *et al.* 2011). Para los dos primeros set de datos se retuvieron los primeros 15 componentes principales que explicaron el 50,4% y 79,5% de la varianza total, respectivamente; sólo una función discriminante fue retenida en cada uno de estos set de datos. Para el set de datos que comprende las cuatro especies, se

retuvieron los primeros 30 componentes principales, que explicaron el 84,3% de la varianza total, y tres funciones discriminantes. A partir de la distribución de las contribuciones de los alelos a la primera función discriminante, se extrajeron aquellos SNPs del percentil 90, considerándolos como aquellos SNPs que más contribuyen a la separación de los grupos. Al igual que en el análisis anterior, se analizaron estos sitios para verificar si están involucrados en las mismas vías o funciones génicas. Para ello, estos sitios fueron ubicados en el genoma de referencia de *O. ascotanensis* para clasificarlos en genes anotados o no anotados, utilizando el mismo script personalizado mencionado en el punto anterior, y se realizaron los mismo dos análisis mencionados arriba con el número de sitios clasificados en genes anotados: (i) en PANTHER (Protein ANalysis THrough Evolutionary Relationships; Mi *et al.* 2013, 2017) se realizó un análisis de lista de genes, utilizando el nombre de los genes como identificador (ID), para estimar el número de genes con y sin ID conocido; (ii) con el conjunto de genes anotados con ID conocido y sus términos de Gene Ontology (GO) se realizó un análisis de enriquecimiento génico con la herramienta GOrilla (Gene Ontology enRIchment anaLysis and visuaLizAtion tool; Eden *et al.* 2007, 2009). En ambas bases de datos se utilizó la anotación del genoma humano.

4. Resultados

4.1 Valores de F_{ST} a distintos niveles de diferenciación.

El número de total de loci obtenidos por cada set de datos para estimar los valores de F_{ST} promedió $15.210,5 \pm 1.820,5$, donde el set de datos de las especies muy divergentes presenta un 21% más de SNPs que el set de datos de las especies incipientes (Tabla 7).

Los patrones de divergencia interespecífica se estimaron mediante (i) el cálculo de F_{ST} global y (ii) el cálculo de los valores de F_{ST} por cada locus. En la Tabla 7 se observa que el valor de F_{ST} global entre especies fue 3 veces más alto en las especies divergentes que en las especies recientes, y ambos valores fueron significativamente diferentes de cero. El análisis de F_{ST} locus por locus estimó que el número de

loci con valores de F_{ST} significativos al 5% fue casi un quinto del total de loci en especies recientes, mientras que esta cantidad aumentó a casi la mitad de los loci entre especies divergentes (Tabla 7).

La distribución de valores de F_{ST} para las especies incipientes (Figura 13a) es una curva “tipo L”, es decir, está fuertemente sesgada a la izquierda debido a que la mayoría de los valores son iguales a cero, y el valor máximo de $F_{ST} = 1$ es alcanzado por sólo 4 loci (Tabla 7). Por otro lado, la distribución de valores de F_{ST} para el par de especies divergentes (Figura 13b) se asemeja a una curva tipo “J invertida”, dada la alta proporción de valores cercanos a cero, al igual que la curva anterior, pero además presenta un elevado número de loci con $F_{ST} = 1$ (814 en total; Tabla 7), además de presentar una mayor representación de valores intermedios; ambas distribuciones son significativamente diferentes (prueba de los rangos con signo de Wilcoxon, $W = 69.314.000$, $p < 0,05$). Las Figura 13c y Figura 13d muestran las distribuciones de los valores de F_{ST} significativos al 5% de cada par de especies. Como se espera, ambas curvas están desplazadas hacia la derecha con respecto a la distribución de los valores de F_{ST} totales. Además, los valores máximos son $F_{ST} = 1$, pero en la etapa incipiente de especiación la representación de estos valores es mucho menor que en la etapa avanzada (Tabla 7). La curva entre especies divergentes está invertida con respecto a la curva de los valores totales indicando que la mayoría de los valores de F_{ST} significativos son los valores cercanos a 1. Finalmente, se observa que la mediana de la distribución aumenta a medida que se aumenta en el continuo de especiación, tanto para los valores de totales de F_{ST} como en las curvas que consideran sólo los loci con F_{ST} significativo.

Los valores de F_{ST} para cada locus localizados en el scaffold del genoma de referencia de *O. ascotanensis* en el que se ubican (Figura 14), muestran (i) que el número de loci diferenciados y fijados aumenta a medida que se avanza en el continuo de especiación y (ii) que los loci diferenciados se encuentran distribuidos a lo largo de todo el genoma. De esta manera, los patrones de diferenciación observados se podrían clasificar de acuerdo a su grado de diferenciación: diferenciación media y diferenciación alta. En el primero, describe el patrón detectado en el par de especies que divergió recientemente, donde se observa que cerca del 20% de los loci están diferenciados, pero muy pocos de estos están fijados (menos del 0,0003% del total de loci). El segundo caso describiría el patrón detectado entre especies con tiempo

de divergencia muy extenso, cuya principal característica es que cerca del 50% de los loci están diferenciados, y de éstos, el 10% está fijado.

Tabla 7. Valores de F_{ST} globales y resumen de los resultados de los valores de F_{ST} locus por locus, en cada par especies de *Orestias*. N: tamaño muestral. ^a: p-value < 0.05. ^b: entre paréntesis se indica el porcentaje de loci con respecto al total de loci. ^c: se aplicó la corrección FDR para estimar el número de loci con F_{ST} significativo.

Especies	N	N° loci	F_{ST} global	N° loci con F_{ST} significativo al 5% ^c	N° loci con $F_{ST} = 0$	N° loci con $F_{ST} = 1$
<i>O. chungarensis</i> / <i>O. laucaensis</i>	48	13390	0,21525 ^a	2659 (19,86%) ^b	5902 (44,1%) ^b	4 (0,03%) ^b
<i>O. ascotanensis</i> / <i>O. gloriae</i>	47	17031	0,63604 ^a	7678 (45,08%) ^b	2908 (17,1%) ^b	814 (4,8%) ^b

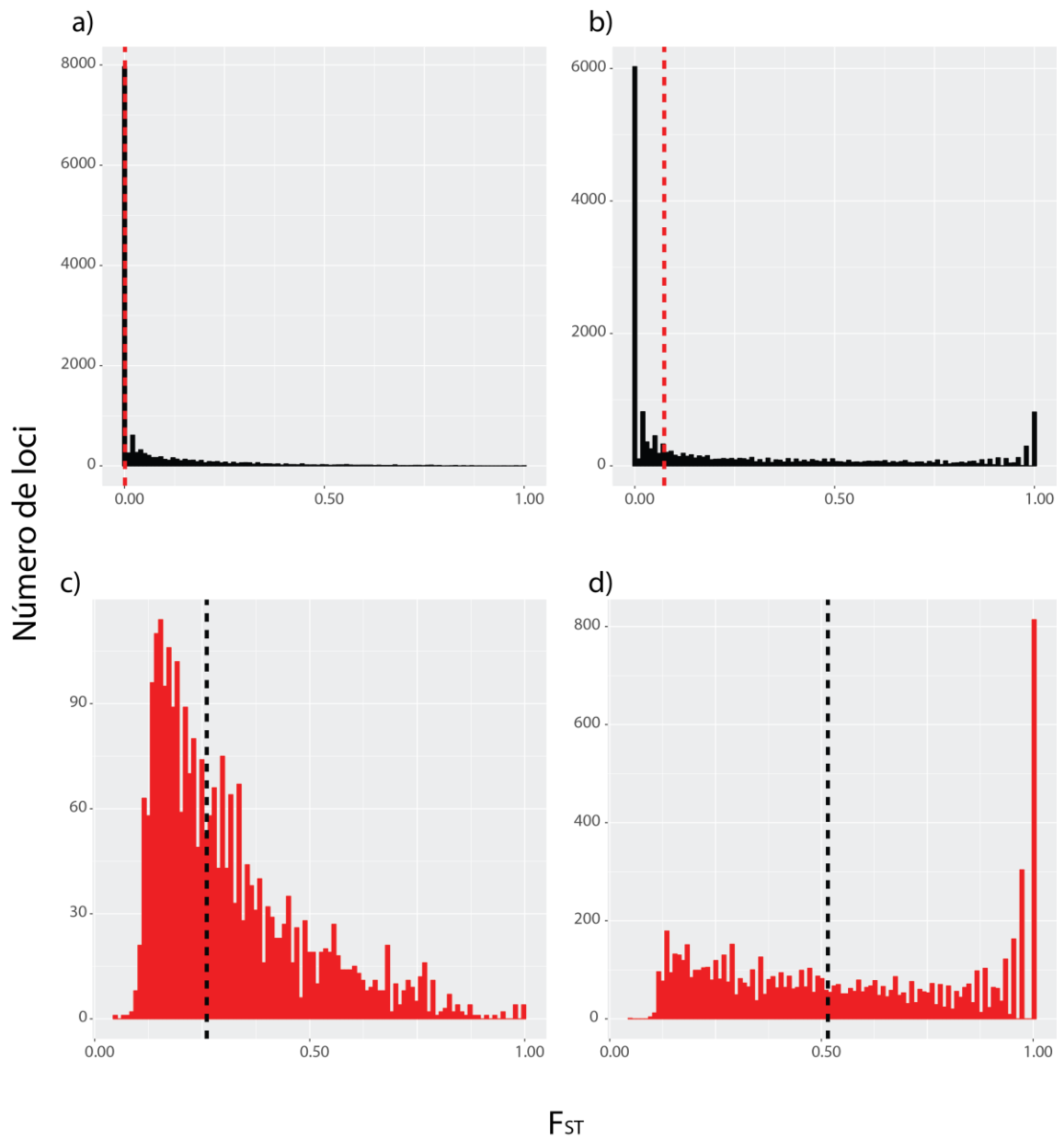


Figura 13. Distribuciones de los valores de F_{ST} entre especies de *Orestias* a) entre las especies del PNL, que recientemente divergieron, *O. laucaensis* y *O. chungarensis*, y b) entre las especies de los salares, que presentan un largo tiempo de divergencia, *O. ascotanensis* y *O. gloriae*. También se indican las distribuciones de los valores de F_{ST} significativos (después de aplicar la corrección de FDR) c) entre especies del PNL y d) entre especies de los salares. Las líneas punteadas (rojas y negras) indican la mediana de cada distribución: en a) es 0,00095, en b) es 0,07, en c) es 0,26y en d) es 0,52.

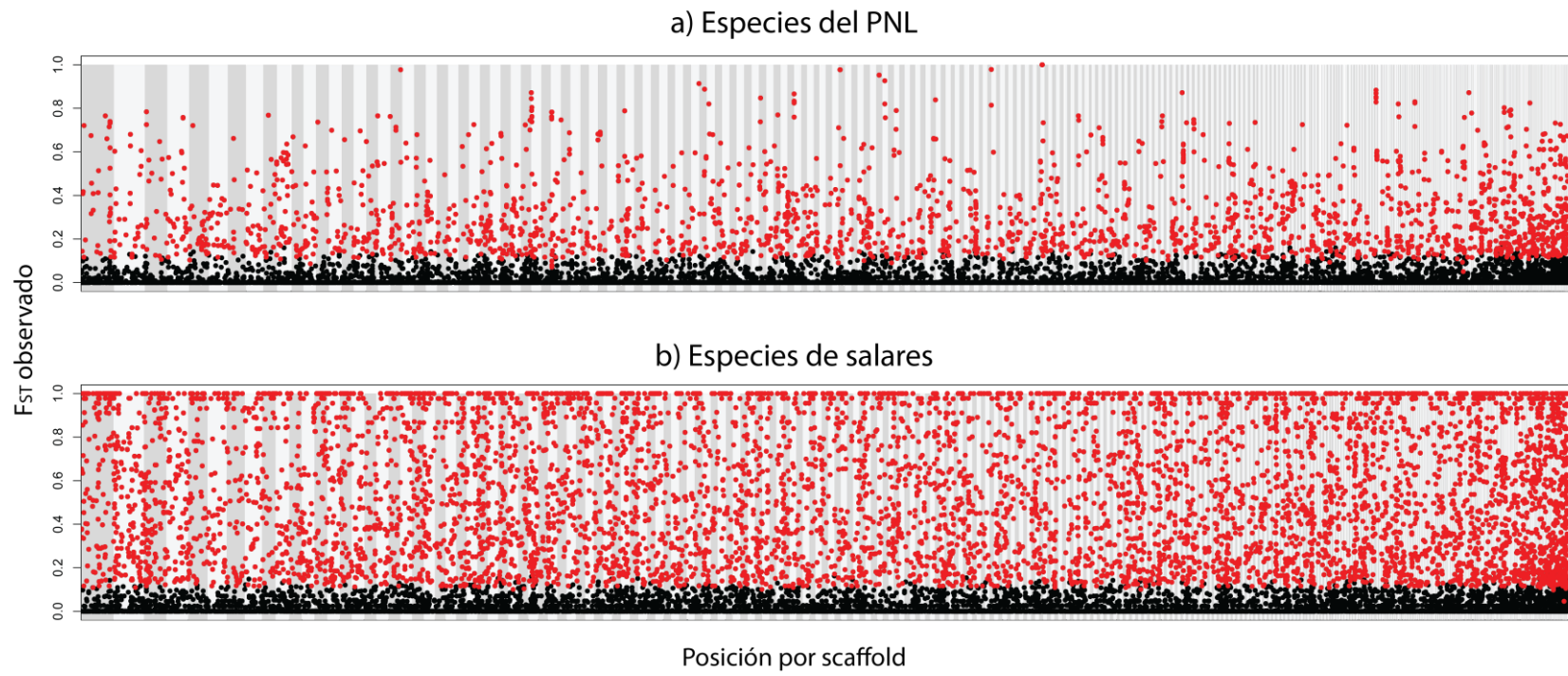


Figura 14. Niveles de diferenciación para cada locus en cada par de especies de *Orestias*. En a) se indican los valores de F_{ST} de cada locus entre las especies del PNL, *O. chungarensis* y *O. laucaensis*, que han divergido recientemente, y en b) entre las especies de los salares, *O. ascotanensis* y *O. gloriata*, con mayor tiempo de divergencia. En el eje x se indican las posiciones de cada locus dentro del scaffold al que pertenecen (indicados en gris claro y gris oscuro para diferenciar scaffolds adyacentes), y en el eje y se indican los valores de F_{ST} para cada locus. Los puntos negros indican los loci con valores de F_{ST} con un valor de $p > 0,05$ y los puntos rojos indican los loci con valores de F_{ST} significativos al 5%, después de la corrección de FDR.

4.2 Número de SNPs y categorías GO comunes y únicos por especie.

Para este análisis se eliminaron los loci que estaban a una distancia menor de 1.000 pb, obteniendo el total de loci por set de datos indicados en la Figura 15 y en la Tabla 8. En la Figura 15 se indica el número de SNPs totales resultantes por especie y el diagrama de Venn que relaciona el número de SNPs compartidos entre especies divergentes (especies de los salares) y entre especies recientes (especies del PNL), además del número de loci que son exclusivos para cada especie en esas combinaciones. El número total de genes por especies (Figura 15a) es mayor en *O. ascotanensis* (con 7.791 SNPs), mientras que *O. gloriae* presenta el menor número de sitios (4.273 sitios). Es decir existe una diferencia de 3.518 sitios entre las especies de los salares, mientras que entre las especies del PNL es de 2.000 SNPs. Por otra parte, el número de sitios comunes en los pares de especies es menor en las de salares (5,7% de 12.064 SNPs en total entre ambas especies; Figura 15b), que en las del PNL (14,4% de 11.296 SNPs en total considerando ambas especies; Figura 15c).

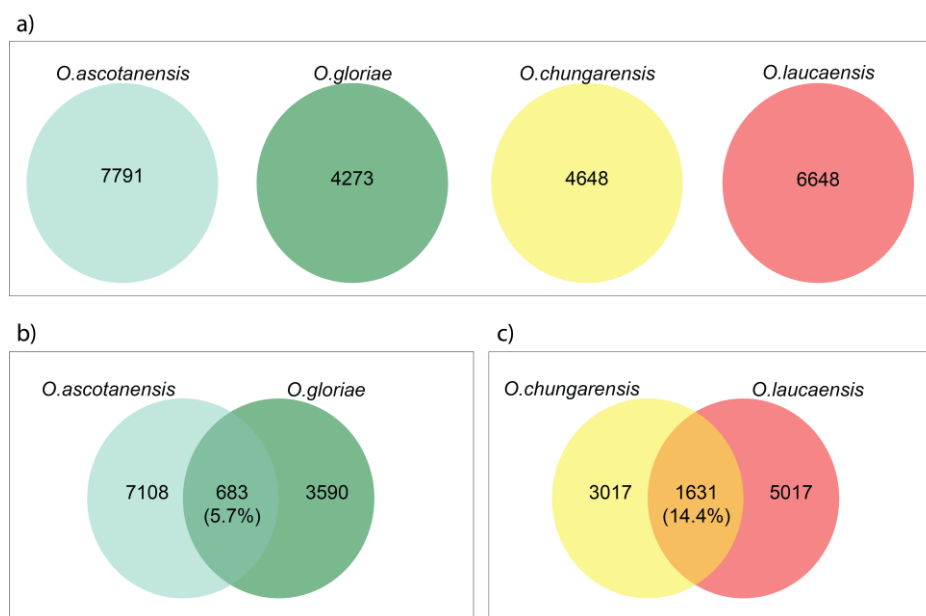


Figura 15. a) Número de sitios (SNPs) totales por cada especie. b) y c) Diagramas de Venn relacionando el número de SNPs de b) las especies de los salares y c) las especies del PNL. En la intersección de cada diagrama se indica el número de sitios que son comunes entre especies y el porcentaje que representa, considerando el total de sitios como la suma de SNPs de ambas especies.

Luego, los SNPs únicos de cada especie obtenidos a partir de los diagramas de Venn (Figura 15b y Figura 15c), fueron mapeados en el genoma de referencia encontrándose que la mayoría de estos sitios se ubican en zonas génicas, mientras que cerca del 32% de los sitios únicos de cada especie se encontrarían en regiones intergénicas (Tabla 8). Esto coincide con que el sitio que reconoce la enzima de restricción utilizada para preparar la librería que fue secuenciada, *Sbf I*, se encuentra con mayor frecuencia en regiones codificantes (Amores *et al.* 2011). Posteriormente, el análisis de enriquecimiento realizado a partir de los genes donde se encontraron SNPs arrojó que la mayor cantidad de categorías GO enriquecidas son las de Procesos Biológicos, siendo las especies divergentes las que tienen más de este tipo de categorías: *O. glorieae* es la especie que muestra el mayor número (36), mientras que *O. chungarensis* es la que presenta el menor número (sólo 5), a pesar de que ambas especies poseen un número similar de SNPs anotados y únicos para cada una (2.439 sitios en *O. glorieae*, y 2.068 en *O. chungarensis*). *O. laucaensis*, a diferencia de *O. chungarensis* (ambas especies del PNL) muestra 17 categorías GO de Procesos Biológicos enriquecidas, lo que es levemente menor al de las especies de los salares. Por otra parte, hay mayor cantidad de categorías enriquecidas de Función Molecular en las especies de salares (10 para *O. ascotanensis* y 12 para *O. glorieae*) que en las del PNL (1 para *O. chungarensis* y 3 para *O. laucaensis*). Y en cuanto a las categorías enriquecidas de Componente Celular, se detectó un número muy bajo para las cuatro especies: sólo se detectaron 3 de estas categorías para *O. ascotanensis* y 1 para *O. laucaensis*. En las Tabla MS 1 y Tabla MS 2 se detallan las categorías GO enriquecidas para cada especie. Es importante tener en cuenta que estos resultados de enriquecimientos de categorías GO pueden estar sesgados a lo que existe en la base de datos de GOrilla, ya que una lista de genes puede ser más reducida que otra, pero puede estar mejor representada en esta base de datos, y por lo tanto generará mayor cantidad de categorías GO enriquecidas.

Finalmente, a partir de los SNPs únicos de cada especie obtenidos a partir de los diagramas de Venn (Figura 15b y 15c) y sus categorías GO enriquecidas (Tabla 8), se advirtió que muy pocas de éstas fueron comunes entre especies (Tabla 9). Entre las especies de los salares se observó que sólo una categoría de Proceso Biológico es común (*multicellular organismal reproductive process*) y tres categorías de Función

Molecular (*protein complex binding, zinc ion binding y macromolecular complex binding*), mientras que ni entre las especies del PNL ni entre las cuatro especies se detectaron procesos comunes.

En resumen, se observó que sólo un bajo porcentaje de los SNPs corresponde a polimorfismo compartido entre especies (Figura 15b y 15c), y que éste es menor entre especies lejanas, que entre especies que han divergido recientemente. Por otra parte, los SNPs únicos de cada especie se ubicarían en genes que enriquecen categorías GO que en general también son únicas por especie (las excepciones se indican en las Tabla 9, MS 1 y MS 2). Es decir, es más frecuente que los SNPs y las categorías GO enriquecidas sean exclusivos de cada especie, a que exista polimorfismo y funciones o vías génicas comunes entre ellas.

Tabla 8. Detalle del número de SNPs por especies, número de SNPs que más contribuyen a la diferenciación de las especies, y el número de SNPs que más contribuyen a la diferenciación y que son únicos y comunes entre set de datos. Se indican los SNPs que se encuentran en genes anotados (y los que no) en el genoma de referencia, número de los genes anotados que tienen un ID conocido (y los que no) en la base de datos de PANTHER y el número de categorías GO enriquecidas que se obtienen a partir de éstos.

Análisis	Especies	N° de SNPs							N° de Categorías GO enriquecidas		
		SNPs totales	SNPs únicos	SNPs en percentil 90	SNPs en genes	SNPs en zonas intergénicas	IDs encontrados	IDs no encontrados	Proceso Biológico	Función Molecular	Componente Celular
SNPs por especie	<i>O. ascotanensis</i>	7791	7108	-	4681	2427	3239	91	21	10	3
	<i>O. glorieae</i>	4273	3590	-	2439	1151	1985	45	36	12	0
	<i>O. chungarensis</i>	4648	3017	-	2068	949	1708	39	5	1	0
	<i>O. laucaensis</i>	6648	5017	-	3330	1687	2558	67	17	3	1
SNPs que más contribuyen a la diferenciación de las especies	Todas las especies	13552	-	1355	473	882	793	21	9	2	0
	Antiguas	10089	-	1009	320	689	603	17	1	1	0
	Recientes	8574	-	857	338	519	458	13	5	0	0
SNPs que más contribuyen a la diferenciación de las especies y que son únicos o comunes por set de datos	Sólo Antiguas	-	921	-	630	291	557	15	0	0	0
	Sólo Recientes	-	769	-	472	297	418	13	6	0	0
	Comunes entre Antiguas y Recientes	-	88	-	60	28	59	2	0	0	0

Tabla 9. Número de categorías GO enriquecidas comunes entre especies de Procesos Biológicos (bajo la diagonal principal) y Función Molecular (sobre la diagonal principal).

	<i>O. ascotanensis</i>	<i>O. gloriae</i>	<i>O. chungarensis</i>	<i>O. laucaensis</i>
<i>O. ascotanensis</i>	-	3	0	1
<i>O. gloriae</i>	1	-	0	2
<i>O. chungarensis</i>	2	0	-	0
<i>O. laucaensis</i>	3	6	0	-

4.3 Identificación de los SNPs de mayor contribución a la diferenciación entre especies

El análisis DAPC permitió identificar los sitios que más contribuyen a la diferenciación de las especies, ya sea considerando las cuatro especies, así como entre especies recientes y antiguas (Tabla 8). De la misma manera que en la sección anterior, se observó que la mayoría (~60%) de los SNPs que contribuyen a la diferenciación de las especies se encuentran en regiones génicas. A partir de esos loci se realizó un análisis en GOrilla que indicó pocas categorías enriquecidas (Tabla MS 3): en el set de datos que contempla las cuatro especies hay un enriquecimiento de nueve categorías GO de Procesos Biológicos, de las cuales dos son categorías hijas: (i) *negative regulation of signal transduction*, y (ii) *regulation of voltage-gated calcium channel activity*. Mientras que hay un enriquecimiento de dos Funciones Moleculares: (i) *cofactor binding* y (ii) *lyase activity*. Considerando el set de datos de las especies divergentes, se detectó el enriquecimiento de sólo un Proceso Biológico, *negative regulation of DNA metabolic process*, y de una Función Molecular, *ATPase activity, coupled*. Para el caso del set de datos de las especies recientes, sólo se detectó el enriquecimiento de cinco Procesos Biológicos, de los cuales dos son categorías hijas, (i) *antigen processing and presentation of exogenous peptide antigen*, y (ii) *antigen processing and presentation of peptide antigen via MHC class I*. No se encontró enriquecimiento para las categorías Función Molecular y en ninguno de los tres set de datos se detectó enriquecimiento de las categorías de Componente Celular. Por último, no se encontró ninguna categoría enriquecida común entre los set de datos.

Posteriormente, se relacionaron los SNPs que contribuyen más a la diferenciación en las especies recientes y las especies antiguas, mediante un diagrama de Venn (Figura 16), para identificar aquellos sitios que son únicos y los que son comunes en la diferenciación de cada par de especies. De la misma manera que en la sección anterior, la intersección de los SNPs de cada set de datos representa una baja proporción (4,9%) de los SNPs totales (suma de los SNPs que más contribuyen en la diferenciación de especies recientes y especies antiguas = 1.778 SNPs). Enseguida, a partir del número de SNPs que son únicos en la diferenciación de cada par de especies (769 en el par “recientes” y 921 en el par “divergentes” como se indica en la Figura 16) se identificaron las categorías GO enriquecidas en cada etapa del continuo de especiación (Tabla 8). Nuevamente, el número de sitios que se encuentran en genes anotados son mayores que el número de sitios que se encuentran en regiones intergénicas. Luego, estos genes que contienen SNPs permitieron reconocer sólo seis categorías enriquecidas de Procesos Biológicos, todas detectadas únicamente en el set de datos que diferencia a las especies recientes (Tabla MS 4). Es interesante notar que si bien el número de SNPs que diferencia a las especies divergentes es mayor (921 SNPs), no se detectó ninguna categoría GO enriquecida a partir de ellos. Por lo tanto, y de manera similar al resultado anterior, tanto el polimorfismo como las funciones génicas serían exclusivos para cada etapa del continuo de especiación.

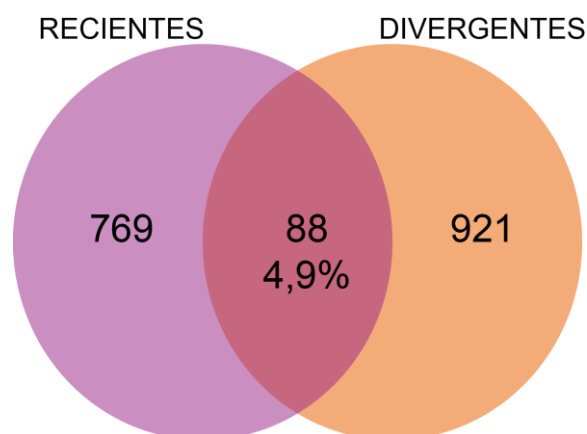


Figura 16. Diagrama de Venn que relaciona el número de SNPs que más contribuyen a la diferenciación entre especies recientes y divergentes.

5. *Discusión.*

En este tercer capítulo se describieron los patrones genómicos de divergencia a lo largo del continuo de especiación considerando dos pares de especies del género *Orestias*, un par de especies incipientes y un par de especies muy divergentes. Esta descripción se realizó a nivel de diferenciación global, de diferenciación de cada locus, por número de SNPs en cada etapa del continuo y número de SNPs que más diferencian a cada par de especies. Los resultados permitieron observar empíricamente cómo el grado de divergencia genómica entre especies alopátricas aumenta a medida que se avanza en el continuo de especiación. Entre especies recientes y entre especies divergentes se logró detectar diferenciación tanto a nivel global como a nivel de locus entre estas especies. La diferenciación global fue casi tres veces mayor entre las especies más distantes, y el número de loci diferenciados, estandarizado por el número de loci totales, fue dos veces mayor entre especies distantes que entre especies recientes. Adicionalmente, la distribución de frecuencias de la diferenciación fue diferente (Figura 13), ya que entre especies cercanas se obtuvo una curva tipo “L”, mientras que entre las especies lejanas se obtuvo una curva tipo “J invertida”: en ambos histogramas se encontró que la mayoría de los loci presentaron valores de F_{ST} cercanos a cero, pero se diferencian principalmente en que el número de loci fijados y la cantidad de loci diferenciados es mucho mayor entre especies lejanas. Adicionalmente, la mayoría del polimorfismo encontrado también es diferente y particular de cada especie (Figura 15b y 15c), al igual que los sitios que diferencian a cada par de especies (Figura 16), es decir, el polimorfismo en una etapa inicial es diferente al detectado en una etapa avanzada del continuo de especiación. Y estos sitios están asociados con funciones o vías génicas que también son diferentes en cada especie (Tabla 9). La principal diferencia entre los patrones interespecíficos estaría dada por el tiempo de divergencia: se ha estimado que las especies del PNL habrían divergido hace 12-8 mil años atrás (Clavero R. *et al.* 2004; Sáez *et al.* 2007; Guerrero-Jiménez *et al.* 2017), lo cual habría permitido la diferenciación de sólo el 20% del total de loci. Por otra parte, se ha estimado un tiempo de divergencia entre las especies antiguas cercano a los 300 ka (Scott 2010), lo que habría permitido la diferenciación de casi el 50% de los loci y la fijación del 5% de los loci (Tabla 7). En esta misma línea, el tiempo de divergencia también explicaría que el polimorfismo común entre especies recientes sea mayor que entre especies antiguas (Figura 15b y 15c). Por otra parte, la presencia de alelos

fijados entre especies divergentes sugiere un origen temprano de la diferenciación en esos sitios (Loh *et al.* 2013). Sin embargo, y dado el corto tiempo de divergencia entre las especies recientes, la diferenciación máxima detectada en cuatro sitios podría ser indicio de algún tipo de régimen selectivo sobre ellos, o estar ligados a genes bajo selección que no fueron secuenciados. Estos resultados sugieren que el proceso de diferenciación de estas cuatro especies ha seguido un camino independiente en cada una de ellas, lo cual es esperable en un proceso de especiación alopátrico. Sin embargo, entre las especies de salares se esperaba detectar similitudes en la diferenciación debido a que los ambientes de estas dos especies son similares. Esta falta de similitud puede ser interpretada como (i) que el largo tiempo de divergencia pudo haber borrado cualquier patrón similar, (ii) que la técnica de RAD-Seq no permitió la secuenciación de aquellas zonas que albergan las similitudes, o (iii) que los procesos adaptativos han seguido rutas distintas en cada caso a pesar de la similitud de ambientes que se observan hoy.

En conclusión, a partir de estos resultados se podría inferir el proceso de diferenciación en especies del género *Orestias* distribuidas alopátricamente: después de un corto tiempo de separación se diferencia una proporción importante de loci pero no logran fijarse (incomplete sorting). A medida que el tiempo de divergencia aumenta, también aumenta el número de loci diferenciados y comienzan a aparecer loci fijados, tanto en regiones génicas como intergénicas, lo que conllevaría efectos sobre la regulación o la transcripción génica y eventualmente sobre el fenotipo de cada especie.

Los patrones de divergencia detectados en este capítulo son similares a los descritos en trabajos previos, pero detectados en escenarios de divergencia en presencia de flujo génico o de especiación ecológica. Se han detectado distribuciones de frecuencia de la diferenciación genética de tipo L, tal como los de la Figura 13a, en especies de peces (en cíclidos [Loh *et al.* 2013] y en pupfish del Caribe [McGirr *et al.* 2017]), en insectos (como en los palotes del género *Timema* [Nosil *et al.* 2012a] y en las mariposas de los géneros *Heliconius* [Martin *et al.* 2013] y *Papilio* [Ryan *et al.* 2017]) y en aves (del género *Ficedula* [Burri *et al.* 2015] y en cuervos [Vijay *et al.* 2016]). Sin embargo, estas distribuciones habrían sido el resultado de procesos de divergencia en presencia de flujo génico, donde sólo algunos loci adaptativos se resisten al efecto homogeneizador del flujo génico y presentan niveles altos de divergencia, al igual que aquellos loci ligados a éstos. La etapa de diferenciación entre especies incipientes de *Orestias* es

comparable a la de especies de *Cyprinodon* de la isla San Salvador, en las Bahamas, las cuales habrían sido el resultado de un proceso de radiación adaptativa reciente ligada a una rápida diversificación morfológica, principalmente en el tamaño de la mandíbula, que les habría permitido ocupar diferentes nichos tróficos (Martin & Wainwright 2011; Martin 2016). En estos pupfish del Caribe, cuyos hábitat son tan jóvenes como los de *Orestias* del PNL (<10.000 años atrás), se detectaron valores de diferenciación similares a los de *Orestias*, además de similitudes en la forma de la curva de la distribución de F_{ST} (Figura 13a), con la mayoría de los valores cercanos a cero y una pequeña proporción de los loci fijados (Martin & Feinstein 2014; McGirr & Martin 2017). Los bajos niveles de diferenciación y las pocas variantes fijadas, que además están asociadas al tamaño de la mandíbula, habrían permitido la variabilidad fenotípica de estas especies (McGirr & Martin 2017). Por otra parte, entre especies divergentes, la curva de frecuencias de diferenciación detectada en las especies de *Orestias* de los salares (Figura 13b) es similar a la detectada en un locus que determina el tamaño del pico en especies de pinzones de Darwin de morfología divergente en este carácter (Lamichhaney *et al.* 2016). Lamichhaney *et al.* (2016) estimaron los valores de F_{ST} entre especies de distinto tamaño corporal y de pico, dentro de la región HMGA2 (de 525.000 bases nucleotídicas de largo), observándose una curva tipo “J invertida” como resultado de que la gran mayoría de los loci presentaron valores cercanos a cero, pero una proporción importante de loci tiene valores cercanos a 1 (1.327 SNPs con valores de $F_{ST} > 0,8$), de los cuales la mayoría se ubicaron en regiones regulatorias. Por lo tanto, este locus tendría un efecto sobre el tamaño del pico (o sobre los huesos o músculos craneofaciales asociados) lo cual a su vez tendría un efecto sobre el fitness, que se habría hecho evidente durante un evento severo de sequía, en el que disminuyó la disponibilidad de alimento, aumentando la competencia y las variantes génicas asociadas a tamaño grande de pico en pinzones medianos habrían estado en desventaja selectiva frente a pinzones grandes también de pico grande, provocando el desplazamiento de carácter ecológico en los pinzones medianos hacia picos más pequeños (Grant & Grant 2006; Lamichhaney *et al.* 2016). A diferencia de estos trabajos, las curvas de frecuencias de diferenciación entre especies obtenidas en este capítulo se explicarían principalmente por el tiempo de divergencia en cada una de las etapas que permitiría actuar a la deriva génica durante más o menos tiempo. Es decir, los patrones de divergencia detectados en presencia de flujo génico o de selección, no

son patrones exclusivos de esos escenarios evolutivos (Southcott & Kronforst 2017), sino que también se pueden observar en escenarios de divergencia alopatrica considerando diferentes tiempos de divergencia. Southcott & Kronforst (2017) simularon la cantidad del genoma que diverge en el tiempo bajo diferentes niveles de flujo génico (incluyendo sin flujo génico o alopatría), y en presencia o ausencia de selección; observaron que la proporción del genoma que diverge aumenta de manera más rápida que un aumento lineal (específicamente, de manera exponencial) cuando (i) no hay selección ni flujo génico o cuando hubo flujo génico sólo en una etapa temprana después de la divergencia, (ii) no hay selección y hay bajos niveles o no hay flujo génico, y (iii) en presencia de selección sin flujo génico, con flujo génico en una etapa temprana después de la divergencia o con bajos niveles de flujo génico.

Adicionalmente, se ha detectado que en el caso de estados más avanzados de diferenciación, las distribuciones de los valores de F_{ST} entre especies suelen tener una media y varianza más alta, tal como se observó en la Fig. 13 en que la mediana aumentó a medida que se avanza en el continuo de especiación.

Otro resultado interesante fue la mayor detección de SNPs en zonas génicas que intergénicas (Tabla 8), pero se debe tener en cuenta que la enzima de restricción *Sbf I* utilizada para la preparar la librería corta más frecuentemente en zonas codificantes (Amores *et al.* 2011). Todavía es necesario poner a prueba si los sitios de restricción de esta enzima son efectivamente más abundantes en zonas génicas que intergénicas, por ejemplo cuantificando aquellos sitios en ambas zonas en el genoma de referencia de *O. ascotanensis*. Algunos de los aspectos que aún quedan por explorar son la identificación de la localización exacta de estos polimorfismos en el genoma (en intrones, zonas de splicing, exones, o regiones UTR), y si los cambios que generan son sinónimos o no sinónimos. Esto permitirá comprender si los SNPs detectados pueden tener un efecto directo sobre la transcripción génica al estar ubicado en exones, o tener un efecto regulador de la transcripción génica o afectar los sitios de *splicing* de los intrones, si se ubican en intrones o en regiones UTR (Cooper 2010), para en definitiva inferir el posible efecto que pueden tener estas variantes a nivel fenotípico, y a la vez estimar el nivel de aislamiento postcigótico generado después del aislamiento geográfico.

6. Material Suplementario

Tabla MS 1. Categorías GO enriquecidas de Proceso Biológico por especie. Se indican las categorías compartidas por pares de especies: ^a: *O. ascotanensis* y *O. gloriæ*; ^b: *O. ascotanensis* y *O. chungarensis*; ^c: *O. ascotanensis* y *O. laucaensis*; ^d: *O. gloriæ* y *O. laucaensis*. *O. chungarensis* no comparte categorías enriquecidas con *O. laucaensis* ni con *O. gloriæ*.

Término GO	Descripción
<i>O. ascotanensis</i>	
GO:0048513	animal organ development ^b
GO:0006928	movement of cell or subcellular component
GO:0001568	blood vessel development
GO:0044763	single-organism cellular process ^c
GO:0048562	embryonic organ morphogenesis
GO:0044699	single-organism process ^c
GO:0043297	apical junction assembly ^c
GO:0065007	biological regulation
GO:0043523	regulation of neuron apoptotic process
GO:0048609	multicellular organismal reproductive process ^a
GO:0050789	regulation of biological process
GO:0043524	negative regulation of neuron apoptotic process
GO:0048598	embryonic morphogenesis
GO:0007166	cell surface receptor signaling pathway
GO:2001031	positive regulation of cellular glucuronidation
GO:2001029	regulation of cellular glucuronidation
GO:0035668	TRAM-dependent toll-like receptor signaling pathway
GO:0035669	TRAM-dependent toll-like receptor 4 signaling pathway
GO:0060193	positive regulation of lipase activity
GO:0007215	glutamate receptor signaling pathway
GO:0048565	digestive tract development ^b

Tabla MS 1. (continuación)

Término GO	Descripción
<i>O. glorieae</i>	
GO:0043278	response to morphine
GO:0014072	response to isoquinoline alkaloid
GO:0048609	multicellular organismal reproductive process ^a
GO:0048386	positive regulation of retinoic acid receptor signaling pathway
GO:0045944	positive regulation of transcription from RNA polymerase II promoter
GO:0038093	Fc receptor signaling pathway
GO:0032526	response to retinoic acid
GO:0010628	positive regulation of gene expression
GO:0048385	regulation of retinoic acid receptor signaling pathway
GO:0019098	reproductive behavior
GO:0002768	immune response-regulating cell surface receptor signaling pathway
GO:0051099	positive regulation of binding
GO:1902680	positive regulation of RNA biosynthetic process
GO:0045893	positive regulation of transcription, DNA-templated
GO:1903508	positive regulation of nucleic acid-templated transcription
GO:0006175	dATP biosynthetic process ^d
GO:0006167	AMP biosynthetic process ^d
GO:0006169	adenosine salvage ^d
GO:0046060	dATP metabolic process
GO:0046086	adenosine biosynthetic process
GO:0046033	AMP metabolic process
GO:0046385	deoxyribose phosphate biosynthetic process
GO:0044209	AMP salvage ^d
GO:0009202	deoxyribonucleoside triphosphate biosynthetic process ^d
GO:0009216	purine deoxyribonucleoside triphosphate biosynthetic process ^d
GO:0009265	2'-deoxyribonucleotide biosynthetic process
GO:0009153	purine deoxyribonucleotide biosynthetic process
GO:0002757	immune response-activating signal transduction
GO:0002429	immune response-activating cell surface receptor signaling pathway
GO:0051254	positive regulation of RNA metabolic process
GO:0043401	steroid hormone mediated signaling pathway
GO:0009755	hormone-mediated signaling pathway
GO:0002253	activation of immune response
GO:0002764	immune response-regulating signaling pathway
GO:0098693	regulation of synaptic vesicle cycle
GO:0010565	regulation of cellular ketone metabolic process
<i>O. chungarensis</i>	
GO:0048565	digestive tract development ^b
GO:0050994	regulation of lipid catabolic process
GO:0048513	animal organ development ^b
GO:0009950	dorsal/ventral axis specification
GO:0050768	negative regulation of neurogenesis
<i>O. laucaensis</i>	
GO:0044699	single-organism process ^c
GO:0045835	negative regulation of meiotic nuclear division
GO:0044763	single-organism cellular process ^c
GO:0006175	dATP biosynthetic process ^d
GO:0006167	AMP biosynthetic process ^d
GO:0006169	adenosine salvage ^d
GO:0044209	AMP salvage ^d
GO:0009202	deoxyribonucleoside triphosphate biosynthetic process ^d
GO:0009216	purine deoxyribonucleoside triphosphate biosynthetic process ^d
GO:0010763	positive regulation of fibroblast migration
GO:0030073	insulin secretion
GO:0048617	embryonic foregut morphogenesis
GO:0043297	apical junction assembly ^c
GO:0034763	negative regulation of transmembrane transport
GO:0034766	negative regulation of ion transmembrane transport
GO:0010165	response to X-ray
GO:0000187	activation of MAPK activity

Tabla MS 2. Categorías GO enriquecidas de Función Molecular por cada una de las especies. Se indican las categorías compartidas por pares de especies: ^a: *O. ascotanensis* y *O. gloriae*; ^b: *O. ascotanensis* y *O. laucaensis*; ^c: *O. gloriae* y *O. laucaensis*. *O. chungarensis* no comparte categorías enriquecidas con ninguna de las otras tres especies.

Término GO	Descripción
<i>O. ascotanensis</i>	
GO:0099600	transmembrane receptor activity
GO:0038023	signaling receptor activity
GO:0071933	Arp2/3 complex binding
GO:0032403	protein complex binding ^{a,b}
GO:0004872	receptor activity
GO:0060089	molecular transducer activity
GO:0004871	signal transducer activity
GO:0008270	zinc ion binding ^a
GO:0044877	macromolecular complex binding ^a
GO:0035276	ethanol binding
<i>O. gloriae</i>	
GO:0019904	protein domain specific binding
GO:0008270	zinc ion binding ^a
GO:0035257	nuclear hormone receptor binding
GO:0032403	protein complex binding ^{a,c}
GO:0051427	hormone receptor binding
GO:0035258	steroid hormone receptor binding
GO:0046914	transition metal ion binding
GO:0044877	macromolecular complex binding ^a
GO:0050681	androgen receptor binding
GO:0004001	adenosine kinase activity ^c
GO:0019206	nucleoside kinase activity
GO:0030898	actin-dependent ATPase activity
<i>O. chungarensis</i>	
GO:0016597	amino acid binding
<i>O. laucaensis</i>	
GO:0034713	type I transforming growth factor beta receptor binding
GO:0004001	adenosine kinase activity ^c
GO:0032403	protein complex binding ^{b,c}

Tabla MS 3. Categorías GO enriquecidas detectadas a partir del 10% de loci que más contribuyen a la diferenciación de las especies en tres set de datos: todas las cuatro especies, las especies antiguas (de los salares) y especies recientes (del PNL).

Término GO	Descripción
PROCESOS BIOLÓGICOS	
Todas las especies	
GO:0048585	negative regulation of response to stimulus
GO:0090303	positive regulation of wound healing
GO:1903036	positive regulation of response to wounding
GO:0010648	negative regulation of cell communication
GO:0023057	negative regulation of signaling
GO:0009968	negative regulation of signal transduction
GO:0061041	regulation of wound healing
GO:1901385	regulation of voltage-gated calcium channel activity
GO:1903034	regulation of response to wounding
Especies Antiguas	
GO:0051053	negative regulation of DNA metabolic process
Especies Recientes	
GO:0002474	antigen processing and presentation of peptide antigen via MHC class I
GO:0048002	antigen processing and presentation of peptide antigen
GO:0002478	antigen processing and presentation of exogenous peptide antigen
GO:0019882	antigen processing and presentation
GO:0019884	antigen processing and presentation of exogenous antigen
FUNCIÓN MOLECULAR	
4 especies	
GO:0048037	cofactor binding
GO:0016829	lyase activity
Especies Antiguas	
GO:0042623	ATPase activity, coupled

Tabla MS 4. Categorías GO enriquecidas de Procesos Biológicos detectadas a partir de los 769 SNPs que sólo diferencian a las especies recientes. Ninguna otra categoría presentó subcategorías enriquecidas. Ninguno de los otros dos set de datos presentó categorías enriquecidas.

Término GO	Descripción
GO:0002474	antigen processing and presentation of peptide antigen via MHC class I
GO:0048002	antigen processing and presentation of peptide antigen
GO:0002478	antigen processing and presentation of exogenous peptide antigen
GO:0019882	antigen processing and presentation
GO:0019884	antigen processing and presentation of exogenous antigen
GO:0034629	cellular protein complex localization

DISCUSIÓN GENERAL

1. Aportes a la comprensión del proceso de especiación en *Orestias*.

En el Capítulo II de esta tesis se estimaron los diferentes grados de estructuración entre cuatro especies chilenas del género *Orestias*. En el Capítulo III se exploraron los patrones de diferenciación entre estas especies en mayor detalle, en búsqueda de las regiones del genoma donde ha ocurrido la diferenciación. Uno de los primeros resultados que aparecieron en el Capítulo II fue la clara diferenciación entre las especies incipientes, uniéndose a la evidencia cromosómica que las indica como entidades diferentes. En cuanto al nivel de diferenciación detectado con la técnica de representación reducida, RAD-Seq (F_{ST} pareado = 0.24, p-value < 0.001), éste fue un valor intermedio considerando los valores de diferenciación detectados previamente con microsatélites (F_{ST} pareado = 0.09, p-value < 0.001) y secuencias de la región control del ADN mitocondrial (F_{ST} pareado = 0.048, p-value < 0.001). Hasta este punto, el uso de una técnica de secuenciación masiva no representó un avance importante al momento de estimar los niveles de F_{ST} entre especies. Sin embargo, la ventaja de utilizar esta técnica se evidenció al momento de detectar entidades o cluster: con los marcadores previamente mencionados no había sido posible diferenciar a las especies del PNL, pero la alta cantidad de marcadores SNPs obtenidos a partir de RADSeq permitió separar a las cuatro especies claramente en cluster o linajes diferentes, especialmente en el caso de las especies del PNL que pudieron ser asignadas a dos entidades diferenciadas. De este resultado se desprenden dos posibles aplicaciones al estudio de las relaciones filogenéticas en el género: (i) realizar un estudio filogenómico, aplicando esta técnica a todo el género, tanto a las especies descritas como a las

poblaciones que cuyo estatus taxonómico aún no ha sido establecido, o (ii) realizar un estudio filogenético basado en la genotipificación de aquellos SNPs que más diferenciaron a las especies considerándolos como marcadores candidatos que podrían ser capaces de distinguir grupos que han divergido recientemente, incluso menos de 10 mil años atrás, como es el caso de las especies del PNL. Esto permitiría robustecer los análisis filogenéticos existentes permitiendo poner a prueba las hipótesis filogenéticas propuestas para este género, incrementando también el poder estadístico de las inferencias obtenidas, y observar en mayor detalle los eventos de especiación recientes que puedan haber ocurrido, por ejemplo, después de la extinción del último gran paleolago, Tauca, hace 14 mil años atrás (Sylvestre *et al.* 1999; Placzek *et al.* 2006, 2011; Blard *et al.* 2011). Más aún, estos marcadores podrían resolver las relaciones filogenéticas dentro de los cuatro grandes clados/complejos encontrados por Esquer Garrigos *et al.* (2013), permitiendo avanzar en la sistemática de este grupo, sin tener que recurrir necesariamente a la obtención de datos de genomas completos, dado el alto costo que significan aún.

Por otra parte, el uso de técnicas de secuenciación masiva, aunque sean de representación reducida como RAD-Seq (o cualquiera de sus variantes como ddRAD, ezRAD, 2b-RAD o nextRAD), permiten una mayor comprensión del paisaje genómico de la diferenciación (Ellegren 2014). En el Capítulo III de esta tesis se abordó esta temática y se logró responder algunas de las preguntas planteadas al comienzo de este trabajo: (i) ¿Dónde ocurre la divergencia?, de forma homogénea a lo largo de todo el genoma; (ii) ¿cuál es la proporción del genoma que contribuye a la divergencia?, dependiendo de la etapa del continuo de divergencia puede ser desde el 20% de los loci diferenciados (con sólo 4 loci fijados), hasta cerca del 50% de los loci diferenciados de los cuales el 10% está fijado; y (iii) ¿cuál es la naturaleza de los genes o zonas genómicas involucradas?, esta respuesta es variable y depende caso a caso, concordante a un modelo de especiación alopátrica.

Finalmente, y a modo de proyección de estos resultados, dado que estos patrones fueron obtenidos a partir de especies alopátricas, podrían compararse con los que se pueden obtener de especies distribuidas en simpatria, de manera tal de observar patrones de diferenciación genómica en estos distintos escenarios, en busca de los factores que determinan a cada uno. Así, los resultados obtenidos en esta tesis podrían

conformar el modelo nulo en el estudio de especiación con flujo génico (Southcott & Kronforst 2017). Por otra parte, las especies de *Orestias* que habitan en el Lago Titicaca podrían ser consideradas como el modelo de estudio en simpatria, principalmente porque es posible encontrar especies tanto de un mismo complejo como de diferentes complejos en un mismo ambiente, lo que también permitiría incluir en el estudio diferentes etapas del continuo de especiación.

2. Crítica a la utilización de RAD-Seq en estudios de especiación.

Si bien el objetivo de esta tesis fue comparar patrones de diferenciación genómicos, éstos fueron resueltos con un método de secuenciación reducida, como lo es RAD-Seq. Y es importante tener en cuenta dos aspectos de esta técnica, (i) el sesgo de la obtención de secuencia hacia sitios de restricción, y (ii) qué tan representativos son los datos obtenidos del genoma total. El primer punto se refiere a que esta técnica logra una representación reducida o sesgada del genoma, pues sólo secuencia regiones asociadas a los sitios de corte de la enzima que se utiliza para la fabricación de la librería. Y por lo tanto, cualquier resultado que se obtenga estará directamente relacionado con secuencias asociadas a esos sitios de corte, que no necesariamente pueden ser representativos del genoma completo. Un ejemplo de esto es que en esta tesis se detectaron una cantidad importante de SNPs en zonas génicas, debido a que los sitios de restricción de la enzima utilizada para preparar la librería a secuenciar, la *sbfI*, se encuentran preferentemente en estas zonas. Sin embargo, si se realizaran nuevamente estos análisis, pero ahora a partir de datos de genomas completos, es esperable que la cantidad de SNPs en zonas intergénicas sea mayor que en el zonas génicas. Mientras que el segundo aspecto de esta técnica radica en que la cantidad de SNPs que se obtienen puede ser una porción muy pequeña de lo que realmente existe. Por ejemplo, el máximo de SNPs por especie obtenido con RAD-Seq (después de una serie de filtros) fue 7.791 SNPs en total para *O. ascotanensis* (Figura 15). Por otra parte, la secuenciación del genoma completo de esta especie, utilizado como genoma de referencia en este trabajo, permitió determinar una tasa de 1 SNP cada 621 bases nucleotídicas, y un tamaño total de este genoma de 756 Mb, lo que equivale a un poco más de 1.200.000 SNPs en total en esta especie (datos no publicados). Esto significa que con RAD-Seq se logró

obtener menos del 1% de SNPs totales de *O. ascotanensis* (este valor puede variar dependiendo de los restrictivo que sean los filtros que se apliquen). En otras palabras, todos los resultados de esta tesis están basados en información parcial y reducida del genoma de estas especies, lo que puede o no ser un problema de acuerdo a la pregunta que se quiera abordar. Si el objetivo es estimar diferenciación genética global entre especies, el avance con respecto a lo que ya se sabía no fue significativo. Si el objetivo es delimitar especies (*sensu lato*), esta técnica fue suficientemente poderosa para separar a las especies del PNL como dos entidades claramente diferentes. Por otra parte, si el objetivo es estimar el paisaje genómico de diferenciación, esta técnica representa una primera aproximación a la comprensión de éste, pero siempre teniendo en cuenta la limitada información genómica con la que se obtuvo, y que el patrón podría cambiar si se utilizara una técnica más intensiva de secuenciación. En ese mismo sentido, los polimorfismos obtenidos y los sitios que más diferencian a estas especies y las categorías GO enriquecidas podrían ser o no los mismo que se obtienen si se realizara el mismo análisis con genomas completos.

Finalmente, la técnica de RAD-Seq mostró ser muy útil para comenzar a explorar los patrones de diferenciación genómica entre estas especies, tal como lo ha sido en etapas iniciales de proyectos de investigación que intentan comprender el proceso de especiación, sobre todo cuando se estudian organismos no modelo en los que no existen datos genómicos previos. Pero también ha quedado en evidencia que para una comprensión integral de este proceso es necesario trabajar con genomas completos, no sólo para observar sitios específicos, sino que para comprender la dinámica completa del genoma.

Tabla 10. Resultados y conclusiones obtenidas en esta tesis luego de poner a prueba cada hipótesis planteada.

Antecedentes	Hipótesis	Resultado	Conclusión
<p>El proceso de especiación es un proceso continuo en el cual la magnitud de la divergencia representa una medida que indica qué tan avanzado se está en este proceso.</p>	<p>A nivel genómico global y por locus, existe mayor divergencia entre <i>O. ascotanensis</i> y <i>O. gloriae</i>, que están en una etapa avanzada del proceso de especiación, que entre <i>O. chungarensis</i> y <i>O. laucaensis</i>, las especies que estén en una etapa incipiente de este proceso.</p>	<p>La diferenciación global del par de especies en la etapa avanzada del continuo de especiación fue tres veces mayor que en el par de especies recientes. La proporción de loci diferenciados significativamente aumentó desde un 20% del total de loci en la etapa reciente, hasta casi la mitad de los loci diferenciados en la etapa más avanzada, de los cuales 10% se encuentra fijado.</p>	<p>Los resultados obtenidos están de acuerdo con la hipótesis planteada.</p>
<p>Las cuatro especies de <i>Orestias</i> se originaron en un contexto de especiación alopátrica, donde los diferentes escenarios que ha experimentado cada una han provocado que la diferenciación ocurra de manera independiente en ellas.</p>	<p>Los patrones de diferenciación genómica entre especies en diferentes etapas del continuo de especiación presentarán bajos niveles de similitud, tanto entre especies como entre etapas de este continuo.</p>	<p>Se observó que el proceso de diferenciación ha sido diferente en cada etapa del continuo de especiación, ya sea en términos de polimorfismo compartido entre pares de especies, de SNPs que más contribuyen a la diferenciación y vías génicas enriquecidas por estos SNPs.</p>	<p>Los resultados obtenidos están de acuerdo con la hipótesis planteada.</p>

BIBLIOGRAFIA

- Aljanabi SM, Martinez I (1997) Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. *Nucleic Acids Research*, **25**, 4692–4693.
- Amores A, Catchen J, Ferrara A, Fontenot Q, Postlethwait JH (2011) Genome evolution and meiotic maps by massively parallel DNA sequencing: spotted gar, an outgroup for the teleost genome duplication. *Genetics*, **188**, 799–808.
- Andrew RL, Rieseberg LH (2013) Divergence is focused on few genomic regions early in speciation: incipient speciation of sunflower ecotypes. *Evolution*, **67**, 2468–82.
- Araya CA (2009) Estudio de la variabilidad cariotípica en *Orestias* (Teleostei: Cyprinodontidae): Caracterización de la heterocromatina constitutiva, mapeo de genes ribosomales y determinación de la cantidad de DNA nuclear. Universidad de Chile.
- Araya-Jaime C, Lam N, Vila Pinto I, Méndez MA, Iturra P (2017) Chromosomal organization of four classes of repetitive DNA sequences in killifish *Orestias ascotanensis* Parenti, 1984 (Cyprinodontiformes, Cyprinodontidae). *Comparative Cytogenetics*, **11**, 463–475.
- Arratia G (1982) Peces del altiplano de Chile. In: *El hombre y los ecosistemas de montaña*, pp. 93–133.
- Van der Auwera GA, Carneiro MO, Hartl C *et al.* (2013) From fastQ data to high-confidence variant calls: The genome analysis toolkit best practices pipeline. *Current Protocols in Bioinformatics*, 11.10.1-11.10.33.

- AVMA (2013) *AVMA Guidelines for the Euthanasia of Animals: 2013 Edition*. American Veterinary Medical Association, Schaumburg, IL.
- Baldo L, Santos ME, Salzburger W (2011) Comparative transcriptomics of eastern african cichlid fishes shows signs of positive selection and a large contribution of untranslated regions to genetic diversity. *Genome Biology and Evolution*, **3**, 443–455.
- Bandelt H, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, **16**, 37–48.
- Barron EJ, Peterson WH (1989) Model simulation of the cretaceous ocean circulation. *Science*, **244**, 684–6.
- Barton NH (2000) Genetic hitchhiking. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **355**, 1553–62.
- Barton NH (2001) Speciation. *Trends in Ecology & Evolution*, **16**, 325.
- Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, **57**, 289–300.
- Bernardi G (2013) Speciation in fishes. *Molecular Ecology*, **22**, 5487–5502.
- Bernatchez L, Renaut S, Whiteley AR *et al.* (2010) On the origin of species: insights from the ecological genomics of lake whitefish. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **365**, 1783–1800.
- Blard P-H, Sylvestre F, Tripathi a. K *et al.* (2011) Lake highstands on the Altiplano (Tropical Andes) contemporaneous with Heinrich 1 and the Younger Dryas: new insights from ^{14}C , U–Th dating and $\delta^{18}\text{O}$ of carbonates. *Quaternary Science Reviews*, **30**, 3973–3989.
- Bradic M, Teotónio H, Borowsky RL (2013) The population genomics of repeated evolution in the blind cavefish *Astyanax mexicanus*. *Molecular Biology and Evolution*, **30**, 2383–2400.
- Burri R, Nater A, Kawakami T *et al.* (2015) Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of *Ficedula*

- flycatchers. *Genome Research*, **25**, 1656–1665.
- Butlin R, DeBelle A, Kerth C *et al.* (2012) What do we need to know about speciation? *Trends in Ecology & Evolution*, **27**, 27–39.
- Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH (2011) Stacks: building and genotyping Loci de novo from short-read sequences. *G3 (Bethesda, Md.)*, **1**, 171–82.
- Catchen J, Hohenlohe P a, Bassham S, Amores A, Cresko W a (2013) Stacks: an analysis tool set for population genomics. *Molecular Ecology*, **22**, 3124–3140.
- Chapman MA, Hiscock SJ, Filatov DA (2013) Genomic divergence during speciation driven by adaptation to altitude. *Molecular Biology and Evolution*, **30**, 2553–2567.
- Clavero R. JE, Sparks SJ, Polanco E, Pringle MS (2004) Evolution of Parinacota volcano, Central Andes, Northern Chile. *Revista Geológica de Chile*, **31**, 317–347.
- Cooper DN (2010) Functional intronic polymorphisms: Buried treasure awaiting discovery within our genes. *Human Genomics*, **4**, 284.
- Coyne J, Orr H (2004) *Speciation*. Sinauer Associates, Inc., Sunderland, Massachusetts.
- Cruikshank TE, Hahn MW (2014) Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology*, **23**, 3133–3157.
- Cruz-Jofré F, Morales P, Esquer-Garrigos Y *et al.* (2014) Taxonomic identity of the forms of *Orestias agassii* (Teleostei: Cyprinodontidae) from Chile: A morphological comparison with the syntypes of the species of Peru and Bolivia. *Gayana*, **78**, 25–29.
- Cruz-Jofré F, Morales P, Vila I *et al.* (2016) Geographical isolation and genetic differentiation: The case of *Orestias ascotanensis* (Teleostei: Cyprinodontidae), an Andean killifish inhabiting a highland salt pan. *Biological Journal of the Linnean Society*, **117**, 747–759.
- Cruz-Jofré F, Valladares MA, Vila I, Mendez MA (2013) The genus *Orestias* (Teleostei: Cyprinodontidae): nomenclatural errors in the assignation of species names. *Zootaxa*, **3746**, 597–599.

- Cui R, Schumer M, Kruesi K *et al.* (2013) Phylogenomics reveals extensive reticulate evolution in *Xiphophorus* fishes. *Evolution*, **67**, 2166–79.
- Danecek P, Auton A, Abecasis G *et al.* (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156–2158.
- Darriba D, Taboada GL, Doallo R, Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods*, **9**, 772–772.
- DePristo MA, Banks E, Poplin R *et al.* (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, **43**, 491–498.
- Eaton DAR, Ree RH (2013) Inferring phylogeny and introgression using RADseq data: An example from flowering plants (Pedicularis: Orobanchaceae). *Systematic Biology*, **62**, 689–706.
- Eden E, Lipson D, Yogev S, Yakhini Z (2007) Discovering motifs in ranked lists of DNA sequences. *PLoS Computational Biology*, **3**, 0508–0522.
- Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z (2009) GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics*, **10**, 48.
- Ellegren H (2008) Sequencing goes 454 and takes large-scale genomics into the wild. *Molecular Ecology*, **17**, 1629–1631.
- Ellegren H (2014) Genome sequencing and population genomics in non-model organisms. *Trends in Ecology & Evolution*, **29**, 51–63.
- Elmer KR, Fan S, Gunter HM *et al.* (2010) Rapid evolution and selection inferred from the transcriptomes of sympatric crater lake cichlid fishes. *Molecular Ecology*, **19**, 197–211.
- Esquer Garrigos YS (2013) Multi-scale evolutionary analysis of a high altitude freshwater species flock: diversification of the *agassizii* complex (*Orestias*, Cyprinodontidae, Teleostei) across the Andean Altiplano. Museum National D'Histoire Naturelle.
- Esquer Garrigos Y, Hugueny B, Koerner K *et al.* (2013) Non-invasive ancient DNA protocol for fluid-preserved specimens and phylogenetic systematics of the genus *Orestias* (Teleostei:

- Cyprinodontidae). *Zootaxa*, **3640**, 373–394.
- 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**, 564–567.
- Feder JL, Gejji R, Yeaman S, Nosil P (2012) Establishment of new mutations under divergence and genome hitchhiking. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **367**, 461–74.
- Feulner PGD, Chain FJJ, Panchal M *et al.* (2013) Genome-wide patterns of standing genetic variation in a marine population of three-spined sticklebacks. *Molecular Ecology*, **22**, 635–649.
- Filatov DA (2009) Processing and population genetic analysis of multigenic datasets with ProSeq3 software. *Bioinformatics*, **25**, 3189–3190.
- Franchini P, Fruciano C, Spreitzer ML *et al.* (2014) Genomic architecture of ecologically divergent body shape in a pair of sympatric crater lake cichlid fishes. *Molecular Ecology*, **23**, 1828–1845.
- Fryer GT, Iles D (1972) *The Cichlid Fishes of the Great Lakes of Africa: Their Biology and Evolution*. Oliver and Boyd, Edinburgh.
- Gagnaire P-A, Pavey S a, Normandeau E, Bernatchez L (2013) The genetic architecture of reproductive isolation during speciation-with-gene-flow in lake whitefish species pairs assessed by RAD sequencing. *Evolution*, **67**, 2483–97.
- Grant PR, Grant BR (2006) Evolution of Character Displacement in Darwin's Finches. *Science*, **313**, 224–226.
- Guerrero CJ (2013) Diferenciación genética, morfológica y ecológica en el complejo *agassizii* del género *Orestias* en las cuencas de la Puna del noreste del Chile. Universidad de Chile.
- Guerrero-Jiménez CJ, Peña F, Morales P *et al.* (2017) Pattern of genetic differentiation of an incipient speciation process: The case of the high Andean killifish *Orestias*. *PLOS ONE*, **12**, e0170380.
- Haq BU, Hardenbol J, Vail PR, Hardenbol JAN (1987) Chronology of fluctuating sea levels since the triassic. *Science*, **235**, 1156–1167.

- Harr B (2006) Genomic islands of differentiation between house mouse subspecies. *Genome Research*, **16**, 730–737.
- Harvey MG, Smith BT, Glenn TC, Faircloth BC, Brumfield RT (2016) Sequence capture versus restriction site associated dna sequencing for shallow systematics. *Systematic Biology*, **65**, 910–924.
- Hendry AP (2009) Ecological speciation! Or the lack thereof? *Canadian Journal of Fisheries and Aquatic Sciences*, **66**, 1383–1398.
- Hendry AP, Peichel CL, Matthews B, Boughman JW, Nosil P (2013) Stickleback research: The now and the next. *Evolutionary Ecology Research*, **15**, 111–141.
- Henning F, Jones JC, Franchini P, Meyer A (2013) Transcriptomics of morphological color change in polychromatic Midas cichlids. *BMC genomics*, **14**, 171.
- Hey J (2001) The mind of the species problem. *Trends in Ecology & Evolution*, **16**, 326–329.
- Hohenlohe PA, Bassham S, Etter PD *et al.* (2010) Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics*, **6**, e1000862.
- Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics (Oxford, England)*, **23**, 1801–6.
- Johnson T, Scholz C, Talbot M (1996) Late Pleistocene desiccation of Lake Victoria and rapid evolution of cichlid fishes. *Science*, **273**, 1091–1093.
- Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*, **11**, 94.
- Jombart T, Lyon D, Biome L De (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, **24**, 1403–1405.
- Jones JC, Fan S, Franchini P, Scharl M, Meyer A (2013) The evolutionary history of *Xiphophorus* fish and their sexually selected sword: a genome-wide approach using restriction site-associated DNA sequencing. *Molecular Ecology*, **22**, 2986–3001.

- Jones FC, Grabherr MG, Chan YF *et al.* (2012) The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*, **484**, 55–61.
- Joyce DA, Lunt DH, Bills R *et al.* (2005) An extant cichlid fish radiation emerged in an extinct Pleistocene lake. *Nature*, **435**, 90–95.
- Keller B, Soto D (1998) Hydrogeologic influences on the preservation of *Orestias ascotanensis* (Teleostei: Cyprinodontidae), in Salar de Ascotán, northern Chile. *Revista Chilena de Historia Natural*, **71**, 147–156.
- Keller I, Wagner CE, Greuter L *et al.* (2013) Population genomic signatures of divergent adaptation, gene flow and hybrid speciation in the rapid radiation of Lake Victoria cichlid fishes. *Molecular Ecology*, **22**, 2848–2863.
- Kornfield I, Smith P (2000) African cichlid fishes: model systems for evolutionary biology. *Annual Review of Ecology and Systematics*, **31**, 163–196.
- Lamichhaney S, Han F, Berglund J *et al.* (2016) A beak size locus in Darwins finches facilitated character displacement during a drought. *Science*, **352**, 470–474.
- Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, **10**, R25.
- Lauzanne L (1982) Les *Orestias* (Pisces, Cyprinodontidae) du Petit lac Titicaca. *Revista Hydrobiologia Tropical*, **15**, 39–70.
- Lauzanne L (1991) Especies nativas Los *Orestias*. In: *El lago Titicaca: síntesis del conocimiento limnológico actual* (eds Claude D, Iltis A), pp. 409–423. HISBOL, La Paz, Bolivia.
- Lawniczak MKN, Emrich SJ, Holloway AK *et al.* (2010) Widespread divergence between incipient *Anopheles gambiae* species revealed by whole genome sequences. *Science*, **330**, 512–4.
- Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD (2012) The SVA package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics*, **28**, 882–883.
- Leigh JW, Bryant D (2015) POPART: Full-feature software for haplotype network construction. *Methods*

- in Ecology and Evolution*, **6**, 1110–1116.
- Li H, Handsaker B, Wysoker A *et al.* (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.
- Lischer HEL, Excoffier L (2012) PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics*, **28**, 298–299.
- Loh Y-HE, Bezault E, Muenzel FM *et al.* (2013) Origins of shared genetic variation in african cichlids. *Molecular Biology and Evolution*, **30**, 906–917.
- Loubens G (1989) Observations sur les poissons de la partie bolivienne du lac Titicaca. IV. *Orestias* spp., *Salmo gairdneri* et problèmes d'aménagement. *Revue d'Hydrobiologie Tropicale*, **22**, 157–177.
- Mallet J (2008) Hybridization, ecological races and the nature of species: empirical evidence for the ease of speciation. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **363**, 2971–86.
- Mallet J, Beltrán M, Neukirchen W, Linares M (2007) Natural hybridization in heliconiine butterflies: the species boundary as a continuum. *BMC Evolutionary Biology*, **7**, 28.
- Mann G (1954) *Vida de los peces en aguas chilenas*. Ministerio de Agricultura y Universidad de Chile, Santiago de Chile.
- Márquez-García M, Vila I, Hinojosa LF *et al.* (2009) Distribution and seasonal fluctuations in the aquatic biodiversity of the southern Altiplano. *Limnologica - Ecology and Management of Inland Waters*, **39**, 314–318.
- Martin CH (2016) Context dependence in complex adaptive landscapes: frequency and trait-dependent selection surfaces within an adaptive radiation of Caribbean pupfishes. *Evolution*, **70**, 1265–1282.
- Martin SH, Dasmahapatra KK, Nadeau NJ *et al.* (2013) Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome Research*, **23**, 1817–1828.
- Martin CH, Feinstein LC (2014) Novel trophic niches drive variable progress towards ecological speciation within an adaptive radiation of pupfishes. *Molecular Ecology*, **23**, 1846–1862.

- Martin CH, Wainwright PC (2011) Trophic novelty is linked to exceptional rates of morphological diversification in two adaptive radiations of *Cyprinodon* pupfish. *Evolution*, **65**, 2197–2212.
- McGaugh SE, Noor M a F (2012) Genomic impacts of chromosomal inversions in parapatric *Drosophila* species. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **367**, 422–9.
- McGirr JA, Martin CH (2017) Novel candidate genes underlying extreme trophic specialization in caribbean pupfishes. *Molecular Biology and Evolution*, **34**, 873–888.
- McKenna A, Hanna M, Banks E *et al.* (2010) The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, **20**, 1297–1303.
- Metzker ML (2010) Sequencing technologies — the next generation. *Nature Reviews Genetics*, **11**, 31–46.
- Mi H, Huang X, Muruganujan A *et al.* (2017) PANTHER version 11: Expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. *Nucleic Acids Research*, **45**, D183–D189.
- Mi H, Muruganujan A, Casagrande JT, Thomas PD (2013) Large-scale gene function analysis with the PANTHER classification system. *Nature Protocols*, **8**, 1551–1566.
- Miyagi R, Terai Y, Aibara M *et al.* (2012) Correlation between nuptial colors and visual sensitivities tuned by opsins leads to species richness in sympatric Lake Victoria cichlid fishes. *Molecular Biology and Evolution*, **29**, 3281–3296.
- Morales P, Cruz-Jofré F, Vila I *et al.* (2013) Phylogenetic and morphometric analysis of the species of *Orestias* (Teleostei; Cyprinodontidae) of the southern Chilean Altiplano support the hypothesis of a recent origin of the species of the Lauca National Park modulated by the collapse of the Parinacota volcano. In: *Congress of the European Society of Evolutionary Biology* , p. 1145. Lisbon.
- Morales P, Vila I, Poulin E (2011) Genetic structure in remnant populations of an endangered cyprinodontid fish, *Orestias ascotanensis*, endemic to the Ascotán salt pan of the Altiplano. *Conservation Genetics*, **12**, 1639–1643.

- Nadeau NJ, Whibley A, Jones RT *et al.* (2012) Genomic islands of divergence in hybridizing *Heliconius* butterflies identified by large-scale targeted sequencing. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **367**, 343–353.
- Nelson JS (2006) *Fishes of the World*. John Wiley and Sons, Hoboken, New Jersey.
- Nosil P (2012) *Ecological speciation*. Oxford University Press.
- Nosil P, Feder JL (2012a) Genomic divergence during speciation: causes and consequences. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **367**, 332–342.
- Nosil P, Feder JL (2012b) Widespread yet heterogeneous genomic divergence. *Molecular Ecology*, **21**, 2829–2832.
- Nosil P, Feder JL (2013) Genome evolution and speciation: toward quantitative descriptions of pattern and process. *Evolution*, **67**, 2461–7.
- Nosil P, Funk DJ, Ortiz-Barrientos D (2009a) Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*, **18**, 375–402.
- Nosil P, Gompert Z, Farkas TE *et al.* (2012) Genomic consequences of multiple speciation processes in a stick insect. *Proceedings of the Royal Society B: Biological Sciences*, **279**, 5058–5065.
- Nosil P, Harmon LJ, Seehausen O (2009b) Ecological explanations for (incomplete) speciation. *Trends in Ecology & Evolution*, **24**, 145–156.
- Orr HA (1995) The population genetics of speciation: The evolution of hybrid incompatibilities. *Genetics*, **139**, 1805–1813.
- Parenti LR (1981) A phylogenetic and biogeographic analysis of cyprinodontiform fishes (Teleostei, Atherinomorpha). *Bulletin of the American Museum of Natural History*, **168**, 335–557.
- Parenti LR (1984a) A taxonomic revision of the Andean killifish genus *Orestias* (Cyprinodontiformes, Cyprinodontidae). *Bulletin of the American Museum of Natural History*, **178**, 107–214.
- Parenti LR (1984b) Biogeography of the Andean killifish genus *Orestias* with comments on the species

- flock concept. In: *Evolution of fish species flocks* (eds Echelle AA, Kornfield I), pp. 85–92. University of Maine Press.
- Placzek C, Quade J, Patchett PJ (2006) Geochronology and stratigraphy of late Pleistocene lake cycles on the southern Bolivian Altiplano: Implications for causes of tropical climate change. *Geological Society of America Bulletin*, **118**, 515–532.
- Placzek CJ, Quade J, Patchett PJ (2011) Isotopic tracers of paleohydrologic change in large lakes of the Bolivian Altiplano. *Quaternary Research*, **75**, 231–244.
- Powell THQ, Hood GR, Murphy MO *et al.* (2013) Genetic divergence along the speciation continuum: the transition from host race to species in *Rhagoletis* (Diptera: Tephritidae). *Evolution*, **67**, 2561–76.
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.
- R Core Team (2015) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- Recknagel H, Elmer KR, Meyer A (2013) A hybrid genetic linkage map of two ecologically and morphologically divergent Midas cichlid fishes (*Amphilophus* spp.) obtained by massively parallel DNA sequencing (ddRADSeq). *G3 (Bethesda, Md.)*, **3**, 65–74.
- Renaut S, Maillet N, Normandeau E *et al.* (2012) Genome-wide patterns of divergence during speciation: the lake whitefish case study. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **367**, 354–363.
- Renaut S, Nolte AW, Bernatchez L (2010) Mining transcriptome sequences towards identifying adaptive single nucleotide polymorphisms in lake whitefish species pairs (*Coregonus* spp. Salmonidae). *Molecular Ecology*, **19**, 115–131.
- Rice W, Hostert E (1993) Laboratory experiments on speciation: what have we learned in 40 years? *Evolution*, 1637–1653.
- Risacher F, Alonso H, Salazar C (2003) The origin of brines and salts in Chilean salars: a hydrochemical

- review. *Earth-Science Reviews*, **63**, 249–293.
- Roesti M, Hendry AP, Salzburger W, Berner D (2012) Genome divergence during evolutionary diversification as revealed in replicate lake-stream stickleback population pairs. *Molecular Ecology*, **21**, 2852–2862.
- Roesti M, Moser D, Berner D (2013) Recombination in the threespine stickleback genome—patterns and consequences. *Molecular Ecology*, **22**, 3014–3027.
- Ronquist F, Teslenko M, Van Der Mark P *et al.* (2012) MrBayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, **61**, 539–542.
- Rosenberg NA (2003) DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Notes*, **4**, 137–138.
- Roy D, Lucek K, Walter RP, Seehausen O (2015) Hybrid “superswarm” leads to rapid divergence and establishment of populations during a biological invasion. *Molecular Ecology*, **24**, 5394–5411.
- Ryan SF, Fontaine MC, Scriber JM *et al.* (2017) Patterns of divergence across the geographic and genomic landscape of a butterfly hybrid zone associated with a climatic gradient. *Molecular Ecology*.
- Sáez A, Valero-Garcés BL, Moreno A *et al.* (2007) Lacustrine sedimentation in active volcanic settings: the Late Quaternary depositional evolution of Lake Chungará (northern Chile). *Sedimentology*, **54**, 1191–1222.
- Salzburger W, Meyer A (2004) The species flocks of East African cichlid fishes: recent advances in molecular phylogenetics and population genetics. *Die Naturwissenschaften*, **91**, 277–90.
- Salzburger W, Meyer A, Baric S, Verheyen E, Sturmbauer C (2002) Phylogeny of the Lake Tanganyika cichlid species flock and its relationship to the Central and East African haplochromine cichlid fish faunas. *Systematic Biology*, **51**, 113–135.
- Schartl M, Walter RB, Shen Y *et al.* (2013) The genome of the platyfish, *Xiphophorus maculatus*, provides insights into evolutionary adaptation and several complex traits. *Nature Genetics*, **45**, 567–

572.

- Scott S (2010) Sistemática y filogenia de *Orestias* del complejo *agassizii* (Teleostei: Cyprinodontidae) de la Puna. Universidad de Chile.
- Seehausen O, Butlin RK, Keller I *et al.* (2014) Genomics and the origin of species. *Nature Reviews Genetics*, **15**, 176–192.
- Seehausen O, Terai Y, Magalhaes IS *et al.* (2008) Speciation through sensory drive in cichlid fish. *Nature*, **455**, 620–626.
- Smadja CM, Butlin RK (2011) A framework for comparing processes of speciation in the presence of gene flow. *Molecular Ecology*, **20**, 5123–5140.
- Sousa V, Hey J (2013) Understanding the origin of species with genome-scale data: modelling gene flow. *Nature Reviews Genetics*, **14**, 404–414.
- Southcott L, Kronforst MR (2017) A neutral view of the evolving genomic architecture of speciation. *Ecology and Evolution*, **7**, 6358–6366.
- St-Cyr J, Derome N, Bernatchez L (2008) The transcriptomics of life-history trade-offs in whitefish species pairs (*Coregonus* sp.). *Molecular Ecology*, **17**, 1850–1870.
- Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, **30**, 1312–1313.
- Strasburg JL, Sherman N a, Wright KM *et al.* (2012) What can patterns of differentiation across plant genomes tell us about adaptation and speciation? *Philosophical Transactions of the Royal Society B: Biological Sciences*, **367**, 364–373.
- Sturmbauer C, Baric S, Salzburger W, Rüber L, Verheyen E (2001) Lake level fluctuations synchronize genetic divergences of cichlid fishes in African lakes. *Molecular Biology and Evolution*, **18**, 144–54.
- Sturmbauer C, Meyer A (1992) Genetic divergence, speciation and morphological stasis in a lineage of African cichlid fishes. *Nature*.

- Sylvestre F, Servant M, Servant-Vildary S *et al.* (1999) Lake-Level chronology on the Southern Bolivian Altiplano (18°–23°S) during late-glacial time and the early Holocene. *Quaternary Research*, **51**, 54–66.
- Takahashi T, Moreno E (2015) A RAD-based phylogenetics for *Orestias* fishes from Lake Titicaca. *Molecular Phylogenetics and Evolution*, **93**, 307–317.
- Tchernavin D (1944) A Revision of the Subfamily Orestiinae. *Proceedings of the Zoological Society of London*, **114**, 140–233.
- Turner TL, Hahn MW, Nuzhdin S V (2005) Genomic islands of speciation in *Anopheles gambiae*. *PLoS Biology*, **3**, e285.
- Turner G, Seehausen O, Knight M, Allender C, Robinson R (2001) How many species of cichlid fishes are there in African lakes? *Molecular Ecology*, 793–806.
- Vandervoort D, Jordan T, Zeitler P, Alonso R (1995) Chronology of internal drainage development and uplift, southern Puna plateau, Argentine central Andes. *Geology*, **23**, 145–148.
- Verheyen E, Salzburger W, Snoeks J, Meyer A (2003) Origin of the superflock of cichlid fishes from Lake Victoria, East Africa. *Science*, **300**, 325–9.
- Vijay N, Bossu CM, Poelstra JW *et al.* (2016) Evolution of heterogeneous genome differentiation across multiple contact zones in a crow species complex. *Nature Communications*, **7**, 13195.
- Vila I (2006) A new species of killifish in the genus *Orestias* (Teleostei: Cyprinodontidae) from the southern high Andes, Chile. *Copeia*, 471–476.
- Vila I, Morales P, Scott S *et al.* (2013) Phylogenetic and phylogeographic analysis of the genus *Orestias* (Teleostei: Cyprinodontidae) in the southern Chilean Altiplano: the relevance of ancient and recent divergence processes in speciation. *Journal of Fish Biology*, **82**, 927–43.
- Vila I, Pinto M (1986) A new species of killifish (Pisces, Cyprinodontidae) from the Chilean Altiplano. *Revue d'hydrobiologie tropicale*, **19**, 233–239.
- Vila I, Scott S, Lam N *et al.* (2010) Karyological and morphological analysis of divergence among species

- of the killifish genus *Orestias* (Teleostei: Cyprinodontidae) from the Southern Altiplano. In: *Origin and Phylogenetic Interrelationships of Teleosts* (eds Nelson JS, Schultze H-P, Wilson MVH), pp. 471–480. Verlag Dr. Friedrich Pfeil, Munchen, Germany.
- Vila I, Scott S, Mendez M (2011) *Orestias gloriae*, a new species of cyprinodontid fish from saltpan spring of the southern high Andes (Teleostei: Cyprinodontidae). *Ichthyological Exploration of Freshwaters*, **22**, 345–353.
- Wagner CE, Keller I, Wittwer S *et al.* (2013) Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. *Molecular Ecology*, **22**, 787–798.
- Warnes GR, Bolker B, Bonebakker L *et al.* (2016) gplots: Various R Programming Tools for Plotting Data. R package version 3.0.1.
- Wickham H (2007) Reshaping Data with the reshape Package. *Journal of Statistical Software*, **21**, 1–20.
- Wickham H (2009) *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, New York.
- Willis B (1910) Principles of paleogeography. *Science*, **31**, 241–260.
- Wolf JBW, Ellegren H (2017) Making sense of genomic islands of differentiation in light of speciation. *Nature Reviews Genetics*, **18**, 87–100.
- Wright S (1931) Evolution in Mendelian Populations. *Genetics*, **16**, 97–159.
- Wright S (1949) The genetical structure of populations. *Annals of Eugenics*, **15**, 323–354.
- Wu C-I (2001) The genic view of the process of speciation. *Journal of Evolutionary Biology*, **14**, 851–865.
- Zheng X, Levine D, Shen J *et al.* (2012) A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, **28**, 3326–3328.

