

# **Tesis de Maestría**

**PEDECIBA**

**Opción Zoología**

**Ensamblado del transcriptoma branquial del Pejerrey y  
búsqueda de selección positiva a nivel molecular en  
Actinopterigios: aportes al estudio de la  
osmorregulación en peces.**

**Lic. Javier Calvelo**

Tutor: Alejandro D'Anatro

Co-tutores: Matías Feijoo y Enrique P. Lessa

Tribunal: Andrés Iriarte, Héctor Romero y María José Arezo

## Resumen:

El grupo de los peces actinopterigios comprende la mitad de las especies de vertebrados existentes, con representantes habitando en prácticamente todos los medios acuícolas disponibles en el planeta. A lo largo de su evolución se han adaptado a vivir en condiciones dulceacuícolas, marinas y estuarinas. En tiempos recientes, el estudio de la evolución del grupo se ha beneficiado enormemente por el desarrollo de las nuevas técnicas de secuenciación masiva y su abrumadora capacidad para la generación de datos. En esta tesis se hace uso de estas nuevas capacidades para contribuir a mejorar nuestra comprensión de la osmorregulación de peces actinopterigios mediante dos estudios: La secuenciación del transcriptoma branquial de *Odontesthes argentinensis*, una especie eurihalina con potencialidad de convertirse en un modelo de estudio para la osmorregulación de peces teleósteos; y con una búsqueda de evidencia de selección direccional, actuando a lo largo de la diversificación de los actinopterigios, sobre genes involucrados en la osmorregulación. Se lograron secuenciar y anotar exitosamente los transcriptomas branquiales para 6 individuos de *O. argentinensis* colectados en distintos puntos de la costa Uruguay, información que será útil en futuros estudios. La búsqueda de selección por su lado encontró evidencia de selección direccional actuando sobre genes importantes en la respuesta osmorregulatoria de actinopterigios, asociada a cambios climáticos que dieron forma a la biota de la Tierra.

## **Agradecimientos**

Le agradezco a la gente del Piso 6 por su apoyo y consejos, especialmente cierto informático por salvarme en más de una ocasión. A mi familia y amigos por soportar más de una calentura. A mi novia por mantenerme semi-cuerdo durante la redacción de esta tesis. Y a la ANII por conceder la beca de maestría que me permitió dedicarme a la biología estos últimos años.

# Tabla de Contenido

Capítulo 1: Introducción general .....	5
1.1 Introducción a la Tesis.....	5
1.2 Esquema de trabajo del presente estudio .....	9
Capítulo 2: Ensamblado del transcriptoma de <i>O. argentinensis</i> .....	11
2.1 Introducción.....	11
2.1.1 Aspectos Biológicos de <i>O. argentinensis</i> .....	11
2.1.2 Secuenciadores de Próxima Generación.....	12
2.1.3 Ensamblado de los transcriptomas.....	12
2.2 Metodología.....	13
2.3 Resultados.....	16
2.4 Discusión .....	23
2.5 Conclusiones finales Capítulo 2 .....	25
Capítulo 3: Evidencia de selección en genes implicados en osmorregulación.....	26
3.1 Introducción.....	26
3.1.1 Mecanismos moleculares de la osmorregulación.....	26
3.1.2 Análisis de selección basados en las tasas de sustitución .....	29
3.2 Metodología.....	31
3.2.1 Filogenia utilizada.....	31
3.2.2 Obtención de datos.....	31
3.2.3 Alineamiento de secuencias y curación manual .....	34
3.2.3 Análisis de selección.....	35
3.3 Resultados.....	37
3.3.1 Selección a nivel de linajes.....	46
3.3.2 Selección a nivel de sitios.....	50
3.4 Discusión .....	60
3.4.1 Sitios bajo selección.....	62
3.4.2 Linajes bajo selección.....	63
3.4.3 Observaciones sobre la metodología utilizada.....	68
3.5 Conclusiones finales Capítulo 3 .....	71
Conclusiones finales.....	72
Perspectivas futuras .....	72
Anexos: .....	73
Bibliografía.....	74

# Capítulo 1: Introducción general

## 1.1 Introducción a la Tesis

Uno de los principales objetivos de la biología evolutiva es el estudio de cómo las presiones selectivas han moldeado la evolución de los seres vivos, que mecanismos llevan a la adaptación de las especies al medio en el que habitan (Freeman & Herron 2014) y que consecuencias que tienen en el la diversificación de las especies (Butlin et al. 2012; Via 2009; Weissing et al. 2011). Cuando los organismos se ven expuestos a nuevas presiones selectivas, la selección natural favorece la fijación de nuevos fenotipos adaptativos (Freeman & Herron 2014), una situación más que frecuente considerando el gran número de cambios geológicos y climáticos que ha experimentado nuestro planeta a lo largo de su historia (ej. Holz 2015; Seton et al. 2012). Ocasionalmente, estos cambios ocurren de forma drástica a una escala global, generando cambios permanentes en la biota del planeta (Hull 2015). No es sorprendente que se haya encontrado evidencia de la selección positiva todo a lo largo de todo el árbol de la vida, una vez que las herramientas utilizadas adquirieran un grado de refinamiento suficiente (ej. Bhatt et al. 2011; Cao & Shi 2012; Demogines et al. 2010; Oellermann et al. 2015; Xu et al. 2016).

Uno de los fenómenos a los que deben adaptarse todos los seres vivos (independiente de su tamaño, ascendencia filogenética o ambiente en el que habiten) es la tendencia de las moléculas a difundir desde regiones de alta concentración a regiones de baja concentración (Nelson & Cox 2008). Cuando dos soluciones con distintas concentraciones de soluto (tienen una osmolaridad diferente) están separadas por una membrana semipermeable que limita significativamente el flujo del soluto, el solvente tiende a desplazarse desde la solución con menor osmolaridad a la de mayor (Nelson & Cox 2008). En el caso de los seres vivos, este fenómeno ocurre con el agua y las membranas celulares. Dependiendo de la relación de osmolaridad del citoplasma y el medio externo se identifican 3 situaciones (Nelson & Cox 2008): 1) Isotónicas: cuando el medio y el organismo tienen la misma osmolaridad, 2) Hipotónicas: cuando el medio tiene menor osmolaridad que el organismo, y 3) Hipertónicas: cuando el medio tiene mayor osmolaridad que el organismo.

Los peces actinopterigios no son una excepción. Desde los orígenes a finales del Paleozoico (Betancur-R. et al. 2013; Near et al. 2012), se han diversificado hasta representar el grupo más diverso de vertebrados (Nelson 2006), adaptándose a vivir en la gran mayoría de los ambientes acuáticos existentes en el planeta, desde lagos en cordilleras montañosas a varios miles de metros de altura a las fosas oceánicas (Nelson 2006). Una de las claves del éxito de éste grupo es su capacidad para mantener una

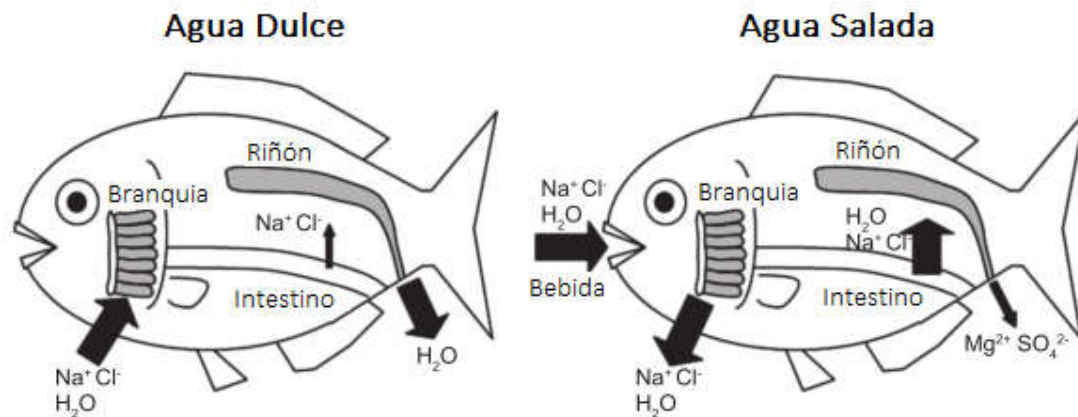
osmolaridad interna diferente a la del medio (Takei et al. 2014; Cerdà & Finn 2010; Hwang et al. 2011; Evans 2008). Las especies dulceacuícolas y marinas mantienen osmolaridades en el entorno de los 300 y los 400 mmol/kg respectivamente, a pesar de que el agua dulce tiene una osmolaridad de <1mmol/kg y la marina supera los 1000mmol/kg (Evans 2008).

Mientras que la mayoría de las especies de teleósteos se limitan a vivir únicamente en condiciones dulceacuícolas o marinas (Nelson 2006), algunas se han adaptado a lidiar con ambas situaciones, cambiando de medio en distintas etapas de su ciclo de vida (ej. peces anádromos como *Salmo salar*, o catádromos como *Anguilla anguilla*), o incluso en escalas temporales tan cortas como de días, incluso horas (ej. especies estuarinas como: *Fundulus heteroclitus*) (Cerdà & Finn 2010; Evans 2008; Hwang et al. 2011; Nelson 2006; Takei et al. 2014).

Estas pocas especies eurihalinas ofrecen la oportunidad de analizar experimentalmente los cambios fisiológicos asociados con la aclimatación a estas dos condiciones tan dispares (Hwang et al. 2011; Takei et al. 2014), facilitando asociar las diferencias observadas (comportamientos, cambios hormonales, niveles de expresión genética, etc.) a las condiciones del medio (Hwang et al. 2011; Marshall & Grosell 2005; Takei et al. 2014). Desafortunadamente, estos estudios se han visto limitados a un acotado grupo de especies; aunque de diversa procedencia filogenética, como por ejemplo *A. anguilla*, *Dicentrarchus labrax*, *Oreochromis mossambicus*, *S. salar*, *Sparus auratus*, entre otras (Cerdà & Finn 2010; Evans 2008; Hwang et al. 2011; Takei et al. 2014).

Dentro de lo que permiten conocer este limitado número de ejemplos, los actinopterigios mantienen una osmolaridad corporal constante mediante movimientos de iones y agua en tres órganos principales: Branquia, Intestino y Riñón (Evans 2008; Hwang et al. 2011; Marshall & Grosell 2005; Takei et al. 2014). Expuestos al agua dulce, los actinopterigios compensan la pérdida constante de iones y la ganancia de agua con la captura activa de iones por las branquias (Evans 2008; Hwang et al. 2011; Takei et al. 2014) y mediante la excreción de abundantes cantidades de agua en la orina (Marshall & Grosell 2005). Mientras que en agua marina, la ganancia de iones es compensada mediante la secreción activa de estos por las branquias (Evans 2008; Hwang et al. 2011; Takei et al. 2014) y la deshidratación por la ingesta de importantes cantidades de agua que es desalinizada en el tracto digestivo y luego absorbida (Grosell 2010; Grosell 2011; Marshall & Grosell 2005); esto acompañado por la excreción de pequeñas cantidades de orina isotónica (Marshall & Grosell 2005). Estos movimientos de agua e iones se esquematizan en la Figura 1, modificada de Takei et al. (2014).

**Fig. 1:** Movimientos de agua e iones en Branquias, Intestino y Riñones involucrados en la osmorregulación de peces Actinopterigios. Imagen modificada de Takei et al. (2014).



De los órganos involucrados en la osmorregulación, las branquias son las que han recibido más atención por parte de los Biólogos, en ellas ocurre la secreción y captura de iones en agua marina y dulce respectivamente (Evans 2008; Hwang et al. 2011; Marshall & Grosell 2005; Takei et al. 2014). Esto se realiza a través de células especializadas denominadas ionocitos, también llamadas en la literatura como células clorhídricas o células ricas en mitocondrias (Hwang et al. 2011; Takei et al. 2014). Los ionocitos branquiales han sido clasificadas en varios tipos dependiendo de los transportadores de iones que son co-expresados en su superficie (Hwang et al. 2011; Takei et al. 2014), aunque han sido caracterizados en muy pocas especies de actinopterigios y difícilmente sean representativos de todo el grupo. Sin embargo la identidad de los genes implicados directamente en el movimiento de iones y agua parece estar conservada en branquias (Cerdà & Finn 2010; Hwang et al. 2011; Takei et al. 2014), y aparentemente también en el intestino (Cerdà & Finn 2010; Grosell 2010; Grosell 2011) y riñones (Kato et al. 2011; Teranishi et al. 2013).

Desde el punto de vista de la biología evolutiva, es de gran interés estudiar y entender como el linaje de los Teleosteoromorpha, principal grupo actual de Actinopterigios, pasó de ser un grupo minoritario a comienzos del mesozoico (Romano et al. 2016) a ser el mayor grupo de vertebrados existente en la actualidad (Nelson 2006), logrando colonizar tal diversidad de ambientes (Evans 2008; Hwang et al. 2011; Marshall & Grosell 2005; Nelson 2006; Takei et al. 2014). El abordaje de esta clase de preguntas se ha beneficiado enormemente con el desarrollo y refinamiento de varias herramientas. En particular el diseño de mejores algoritmos para la detección de selección positiva a partir de alineamientos de secuencias codificantes de ADN (Kosakovsky et al. 2011; Murrell et al. 2012; Lu & Guindon 2013; Yang 2007), así como los métodos de secuenciación de próxima generación (NGS) con los que obtener cantidades sin precedentes de

información (Bahassi & Stambrook 2014; Liu et al. 2012; Metzker 2010). Estos avances han abierto las puertas para la búsqueda de evidencia de selección natural a una escala genómica, incluyendo organismos de los que previamente no se disponía de datos genéticos (Roux et al. 2014; Tsagkogeorga et al. 2015).

En esta tesis se propone contribuir al estudio de la osmorregulación en actinopterigios haciendo uso de estos avances tecnológicos. Por un lado se presenta el ensamblado y anotación del transcriptoma branquial de la especie conocida comúnmente como Pejerrey *O. argentinensis* (Teleostei: Atheriniformes), una especie eurihalina (Dyer 2006; González Castro et al. 2009), a partir de individuos capturados en distintas localidades con diferentes salinidades. Al igual que otras especies eurihalinas comúnmente estudiadas en el marco de la osmorregulación (ej. *O. mossambicus* y *F. heteroclitus* (Takei et al. 2014)), esta especie enfrenta cambios en la osmolaridad del medio. Esta condición, sumada a su fácil colecta en el campo y crianza en cautiverio (González Castro et al. 2009; Sampaio 2006), hacen de esta especie un buen modelo para el estudio de los mecanismos implicados en la osmorregulación. En segundo lugar, se realiza un análisis en busca de evidencia de selección natural en genes implicados en la osmorregulación de peces actinopterigios. Para ello, se emplearon los datos disponibles en la base de datos Ensembl, así como Protein y SRA (ambas a cargo del NCBI (NCBI Resource Coordinators 2015)).

Hipótesis de trabajo:

- 1) El transcriptoma branquial de *O. argentinensis* contiene genes con un rol conocido en la osmorregulación.
- 2) Los genes implicados en la osmorregulación en peces actinopterigios han estado bajo selección positiva en varios momentos a lo largo de la diversificación del grupo.

Predicciones asociadas a la Hipótesis 1):

- a) Se encontrarán genes con participación conocida y/o potencial en la osmorregulación en el transcriptoma ensamblado.

Predicciones asociadas a la Hipótesis 2):

- a) Se encontrará evidencia de selección positiva en los genes que participen en la osmorregulación en peces actinopterigios, ya sea constante o episódica.
- b) Las señales de selección positiva encontradas a nivel de linajes corresponderán a cambios climáticos importantes en la historia geológica de la Tierra y/o a la



colonización de por parte de un grupo particular de un nuevo ambiente con un régimen salino diferente al de su procedencia.

Objetivo general:

Contribuir al estudio de la osmorregulación en actinopterigios mediante la obtención, y anotación, del transcriptoma branquial de *O. argentinensis* y mediante un análisis para detectar evidencia de selección natural a nivel molecular en genes implicados en la osmorregulación de peces actinopterigios.

Objetivos específicos:

- a) Ensamblado del transcriptoma branquial de *O. argentinensis*.
- b) Anotación del transcriptoma branquial de *O. argentinensis*.
- c) Búsqueda bibliográfica de posibles genes vinculados a la osmorregulación en peces.
- d) Obtención, selección, alineamiento y edición de las secuencias de genes vinculados en la osmorregulación de actinopterigios en las bases de datos.
- e) Identificación de linajes bajo selección positiva, detectados consistentemente por varios métodos distintos.
- f) Identificación de sitios encontrados bajo selección positiva por múltiples métodos.
- g) Búsqueda bibliográfica de los posibles cambios climáticos que pudieron afectar la evolución de los linajes encontrados bajo selección positiva, así como cambios en su historia de vida.
- h) Búsqueda bibliográfica de la posible importancia funcional de los sitios encontrados bajo selección positiva.

## 1.2 Esquema de trabajo del presente estudio

En el Capítulo 2 de esta tesis se presenta el ensamblado y anotación del transcriptoma branquial de la especie *O. argentinensis*. Primero se analiza brevemente la biología de la especie, así como la tecnología y herramientas computacionales pertinentes. Luego se detalla la metodología utilizada en la toma de muestras, secuenciación, procesamiento de los *reads*, ensamblado y anotación de los transcritos reconstruidos. Finalmente se presentan los resultados y se discuten las posibles aplicaciones de los datos obtenidos.

En el Capítulo 3 se presenta el análisis en busca de evidencia de selección natural en genes implicados en la osmorregulación de peces actinopterigios. Primero se analizan los mecanismos utilizados por los peces teleósteos en respuesta a la salinidad del medio,

con énfasis en los genes involucrados, y los métodos existentes para la detección de selección positiva a nivel de secuencias codificantes. Luego se detalla la metodología utilizada y los resultados obtenidos. Finalmente, se busca asociar los sitios encontrados bajo selección positiva con regiones funcionales conocidas de cada gen; así como asociar los linajes encontrados bajo selección con cambios climáticos, ocurridos a lo largo de la historia de la Tierra.

## Capítulo 2: Ensamblado del transcriptoma de *O. argentinensis*

### 2.1 Introducción

#### 2.1.1 Aspectos Biológicos de *O. argentinensis*

Los Atheriniformes en general, incluyendo al género *Odontesthes*, son de particular interés para el estudio de la osmorregulación debido a que han invadido medios estuarinos y dulceacuícolas repetidamente a lo largo del cenozoico (Campanella et al. 2015). El género *Odontesthes* está ampliamente distribuido en regiones tropicales y templadas de Sudamérica, ocupando ambientes costeros marinos y dulceacuícolas (Dyer 2006; Llompart et al. 2013). Varias especies del grupo son consideradas de importancia económica para las pesquerías locales y pisciculturas, incluyendo *O. argentinensis* (Barros et al. 2004; Llompart 2011; Sampaio 2006; Somoza et al. 2008).

*Odontesthes argentinensis* habita ambientes marinos y estuarinos a lo largo de la costa atlántica sudamericana, desde São Paulo (25°S) en Brasil a Chubut (43°2'S) en Argentina (Dyer 2006). Estudios con marcadores genéticos sugieren que la estructuración poblacional de la especie parece estar determinada en gran medida por la distancia geográfica entre localidades marinas (Beheregaray & Sunnucks 2001), con pequeñas poblaciones que experimentan un flujo genético limitado con sus contrapartes marinas en lagunas costeras como las Lagunas Custodia, de los Patos y Rocha (Beheregaray & Levy 2000; Beheregaray & Sunnucks 2001).

Las larvas de *O. argentinensis* se alimentan principalmente de presas planctónicas como larvas de insectos, crustáceos y algas filamentosas (Bemvenuti 2006). Cambiando a una dieta bentónica compuesta por crustáceos, poliquetos, oligoquetos y moluscos en su etapa adulta (Bemvenuti 2006; Cristobo 2014).

La temporada reproductiva varía a lo largo de su distribución, desde dos meses en la bahía de San Blas (Llompart et al. 2013) a 5 meses en regiones subtropicales de Brasil (Moresco 2006), siempre comenzando cuando la temperatura se encuentra por debajo de los 20°C, independientemente de la región geográfica (Llompart et al. 2013). Como en otros miembros de la familia Atherinopsidae, la determinación sexual es dependiente de la temperatura del agua durante el desarrollo embrionario (Strüssmann et al. 2010), siendo la proporción 1:1 entre sexos cuando la temperatura se encuentra entre 18 y 21°C (Strüssmann et al. 1996).

### 2.1.2 Secuenciadores de Próxima Generación

El proyecto del genoma humano (Collins et al. 2003; Liu et al. 2012) desencadenó un gran avance en las tecnologías de secuenciación, impulsando el desarrollo de herramientas capaces de abordar preguntas biológicas fuera del alcance de la mayoría de los científicos pocos años atrás (Bahassi & Stambrook 2014; Liu et al. 2012; Metzker 2010). El gran número de secuencias generado por estas tecnologías es aprovechado en campos como la Genómica (Zhanga et al. 2011), Metagenómica (Thomas et al. 2012), y Transcriptómica (Haas & Zody 2010).

Dentro de la transcriptómica, los estudios de RNAseq consisten en la secuenciación masiva de cDNAs mediante NGS (Haas et al. 2013; Wang et al. 2009). Esta aproximación ofrece una alternativa a los microarrays para la caracterización de transcriptomas y estudios de expresión diferencial, libre de limitaciones como la dependencia del conocimiento previo sobre las secuencias de los genes para el diseño de las sondas (Wang et al. 2009). De igual forma permite la detección de SNPs, *splicing* alternativo, nuevos transcriptos, así como el estudio de ARN no codificantes como los micro ARN, entre otros (Ozsolak & Milos 2011; Wang et al. 2009).

Muchos de los secuenciadores disponibles en el mercado han sido utilizados con éxito en RNAseq (Wang et al. 2009). En particular destacan los secuenciadores por síntesis Illumina, estos permiten generar gran número de *reads* sobre los que estimar la expresión genética, y además presentan una alta reproductibilidad de las re-secuenciaciones de las muestras (Ekblom & Galindo 2011; Marioni et al. 2008). En estos dispositivos, los fragmentos de ADN a secuenciar son inmovilizados en una fase sólida y se generan cluster de copias de cada fragmento por secuenciación en puente (Liu et al. 2012; Taylor et al. 2011). Finalmente se elimina una de las hebras de los fragmentos y se realiza la secuenciación mediante la observación de la incorporación de nucleótidos marcados con diferentes fluorocromos (Liu et al. 2012; Taylor et al. 2011). El software interno de Illumina CASAVA (Cock et al. 2010) traduce cada señal de fluorescencia en diferentes bases y les asigna un valor de calidad en PHRED score (Ewing et al. 1998).

### 2.1.3 Ensamblado de los transcriptomas

Un paso crucial para muchos estudios de RNAseq es el ensamblado de los reads, para reconstruir los transcriptos originales (Martin & Wang 2011). Actualmente existen dos estrategias básicas dependiendo de la existencia de una referencia apropiada (comúnmente un genoma) (Martin & Wang 2011). Los ensambladores que trabajan en base a una referencia primero mapean los *reads* a la referencia dada, determinando a que locus corresponde cada *read*, para luego ensamblar cada locus en paralelo (Martin

& Wang 2011). Este abordaje es eficiente computacionalmente y permite la reconstrucción de transcritos de baja abundancia, sin embargo depende de la disponibilidad de una referencia de alta calidad, por lo que es una estrategia inapropiada para el estudio de especies no modelo (Huang et al. 2016; Martin & Wang 2011). Alternativamente se ha desarrollado la aproximación *de novo*. Esta aproximación se basa en el solapamiento de los *reads* secuenciados (Martin & Wang 2011), lo que permite ser usado exitosamente en especies donde no se disponen de referencias adecuadas (Huang et al. 2016).

La aproximación *de novo* es en muchos casos la única opción viable, pero es computacionalmente más demandante, dificultad que ha motivado el desarrollo de herramientas para la reducción de la redundancia presente en los *reads* al mínimo necesario (Brown et al. 2012; Haas et al. 2013). Esta aproximación es además más vulnerable a errores en la secuenciación, es susceptible a generar *contigs* quiméricos entre transcritos similares (Martin & Wang 2011) y los ensamblados generados tienden a ser más fragmentarios (Huang et al. 2016).

Adicionalmente, el uso de *reads* cortos como los generados por los secuenciadores como los Illumina (ej. 100 pb en el caso de Illumina HiSeq 2000 (Taylor et al. 2011)) también impactan negativamente los ensamblados *de novo* (Paszkiwicz & Studholme 2010). Aunque esto puede ser compensado parcialmente con el uso de un protocolo de secuenciación pareada (*paired-end*), en la que se secuencian ambos extremos de cada fragmento secuenciado (Conesa et al. 2016). Como resultado no solo se dispone de la secuencia de los *reads* que conforman el par, sino que se conoce la distancia en pb que los separa (Fullwood et al. 2009).

En esta tesis se utilizó el ensamblador *de novo* Trinity que puede trabajar con esta clase de datos (Grabherr et al. 2011). Como otros ensambladores *de novo*, este utiliza grafos De Bruijn para la reconstrucción de los transcritos (Martin & Wang 2011).

## 2.2 Metodología

Los especímenes de *O. argentinensis* fueron colectados en 4 localidades de la costa Uruguay: Montevideo n=3, Punta Colorada n=1, Laguna de Rocha n=1 y La Paloma n=1 (Fig. 2). Los individuos PJ-6 y PJ-8 estaban expuestos a una salinidad de 8,8 g/L al momento de la captura, PJ-10 a 20,9 g/L, PC-5 a 16,9 g/L, LR-20 a 1,4 g/L y LP-9 a 32,3 g/L. Los ejemplares fueron sacrificados por dislocación cervical en el sitio de colecta y se tomaron muestras de tejido branquial que fueron almacenadas inmediatamente en nitrógeno líquido, hasta ser procesadas en el laboratorio. Las extracciones de ARN se realizaron con el kit comercial RNeasy de Quiagen, siguiendo las instrucciones del

fabricante, y la calidad de las mismas fue evaluada en geles de agarosa 0,4% en buffer TE 0,5 M, un fluorómetro Qubit 2.0 y posteriormente en un Agilent Technologies 2100 Bioanalyzer. La construcción de las librerías y la secuenciación fue realizada por la empresa MacroGen, en Corea del Sur. Se empleó el kit TruSeq Stranded mRNA LT Sample Prep para enriquecer las extracciones en ARNm por captura poli (A) y construir librerías. La secuenciación se realizó en la plataforma Illumina HiSeq 2000 (secuenciador con una capacidad máxima de hasta 6 billones de *reads* pareados por corrida), siguiendo un protocolo *paired-end*, con *reads* de 101 pb.

Tras la secuenciación, siempre es recomendable evaluar la calidad de los *reads* secuenciados y editarlos apropiadamente. Errores en la secuenciación, o la remoción incompleta de secuencias adaptadoras son problemas comunes en estas tecnologías (Conesa et al. 2016). Con este fin se utilizó la herramienta FastQC v0.10.1 (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) para visualizar la calidad de la secuenciación y detectar la presencia de secuencias adaptadoras, luego se utilizó Trimmomatic v0.35 (Bolger et al. 2014) para eliminar estas secuencias. Se descartaron los pares *reads* en los que uno de los miembros del par tuviera un largo menor a 60 pb después de su edición.



**Fig. 2:** Mapa de los sitios de colecta: Montevideo (n = 3), Punta Colorada (n = 1), Laguna de Rocha (n = 1) y La Paloma (n = 1).

El transcriptoma fue ensamblado *de novo* mediante el programa Trinity v2.1.1 con sus opciones por defecto (Grabherr et al. 2011). Además de un ensamblado individual para cada muestra, se siguió el ejemplo de varios trabajos en la literatura (ej. Giorello et al., 2014; Havird et al., 2016; Lam et al., 2014) y se combinaron los *reads* de todos los individuos en un ensamblado conjunto para obtener una reconstrucción de transcriptos más completa.

La calidad de los ensamblados se evaluó mediante el programa TrinityStats (Haas et al. 2013) de acuerdo a los valores de N50 y el largo promedio y mediana de los *reads* en base a todos los *contigs* generados. Los transcriptos obtenidos fueron anotados mediante

BLASTX (Altschul et al. 1990), usando como referencia las secuencias proteicas de *Danio rerio* obtenidas de la base de datos OMA-Browser (Versión Mayo 2016) (Altenhoff et al. 2015), solo se consideraron correspondencias significativas si presentaron un  $e^{-10}$ . Se escogió *D. rerio* por ser una de las especies de teleósteos mejor estudiadas (Ribas & Piferrer 2013) y porque las pruebas exploratorias con otras especies (ej. *O. nicolus*) evidenciaron una reducción en el número de genes anotados limitada usando BLASTX (datos no mostrados).

Para la anotación funcional se asumió que la información disponible para los genes de *D. rerio* es válida para sus correspondencias en *O. argentinensis*. Se consultaron las bases de datos PANTHER (versión 11.1) (Mi et al. 2016) y KEGG (Kanehisa et al. 2006).. Esta última fue consultada vía la herramienta de análisis de listas génicas g:profiler (Reimand et al. 2007) (versión 2016-12-16). Para limitar el número de transcriptos erróneamente identificados, y concentrarse únicamente en los genes de los que se obtuvo más información de cara a futuros estudios, solo se consideraron aquellos con una cobertura superior al 80% de sus respectivas correspondencias en *D. rerio*. Para ser conservador y eliminar el efecto de artefactos en los alineamientos, se definió cobertura como  $(N^{\circ} \text{ de correspondencias} + N^{\circ} \text{ de diferencias} - N^{\circ} \text{ de Gaps}) / \text{Largo de la secuencia de } D. rerio$ .

Se buscaron los términos GO de los genes anotados incluidos en los Términos PANTHER GO-Slim para "Función Molecular (GO:0003674)", "Procesos Biológicos (GO:0008150)" and "Componente Celular (GO:0005575)". Los Términos GO fueron creados por el Consorcio de Ontología Genética con el fin de brindar un lenguaje común para la anotación de las funciones genéticas entre diferentes organismos (Ashburner et al. 2000; Gaudet et al. 2011). Adicionalmente, se realizaron análisis de enriquecimiento de los Términos PANTHER GO-Slim, *Pathways* y Clases Proteicas registradas en esta base de datos, con el objetivo de determinar cuales se encuentran significativamente sobrerrepresentadas con respecto a la totalidad de su catálogo. Adicionalmente, también se consultaron las *Pathways* registradas en la base de datos KEGG. Estos análisis estadísticos se realizaron con las opciones por defecto de las respectivas herramientas de análisis de listas génicas.

Para complementar este análisis exploratorio de los transcriptomas, se realizó un análisis de enriquecimiento de las *Pathways* y Clases Proteicas registradas en PANTHER presentes en los ensamblados individuales, usando como referencia las listas de genes anotados para los transcriptomas branquiales de otros teleósteos sometidos a condiciones opuestas de salinidad. Para ello se descargaron datos de secuenciaciones de RNAseq para *O. mossambicus* (SRR941770 y SRR941772, agua dulce y salada respectivamente) y *Salvelinus alpinus* (SRR921670 y SRR921664, agua dulce y salada

respectivamente) (Lam et al. 2014; Norman et al. 2014). Al igual que en este estudio, estas muestras fueron enriquecidas por pesca poli (A) y secuenciadas con una plataforma Illumina Hiseq 2000. Los reads fueron procesados, ensamblados y anotados siguiendo el mismo procedimiento que con las muestras de pejerreyes, pero debido a limitaciones computacionales, las secuencias de *O. mossambicus* fueron normalizadas con el algoritmo TrinityNorm fijando una cobertura máxima de 50X (Haas et al. 2013).

### 2.3 Resultados

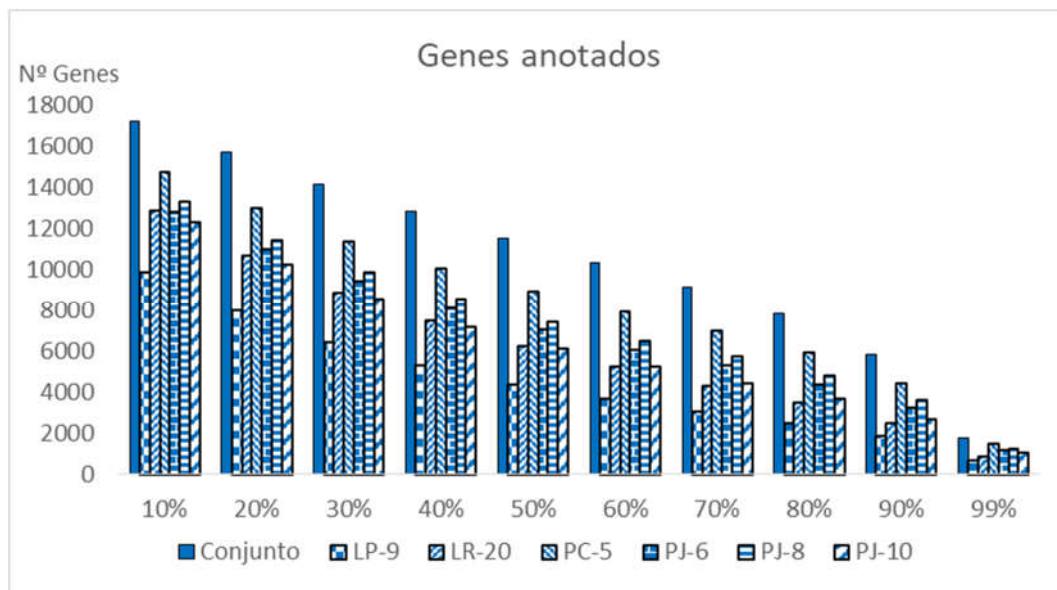
Las 6 librerías generadas en la secuenciación sumaron un total 27 GB de datos con ~300 Millones de *reads* de 101 pb de largo. El número de *contigs* promedio de los ensamblados individuales fue 86.840, con un largo promedio desde los 574 pb a los 744 pb (Tabla 1). El ensamblado conjunto generó un total de 226.995 *contigs*, con unas medidas de N50 y largo promedio de los *contigs* con valores intermedios.

**Tabla 1:** Medidas de calidad de las extracciones de ARN y los ensamblados individuales, el ensamblado Conjunto y los 4 archivos SRA descargados como referencia. Se reporta el RIN de las extracciones, el N° de *reads* pareados secuenciados, N° de *contigs* ensamblados, N50, Largo promedio y Mediana de los *contigs*.

Muestra	RIN	N° reads	N° contigs	N50	L. promedio	Mediana
LP-9	6,20	21.364.596	61.454	896	628	410
LR-20	5,00	24.441.564	110.841	890	622	389
PC-5	5,70	24.851.958	101.109	1.143	719	414
PJ-6	5,10	23.370.271	75.964	1.207	744	438
PJ-8	6,90	27.465.527	77.688	1.052	682	408
PJ-10	6,60	29.382.818	93.982	777	574	367
Conjunto	-	140.613.609	226.995	986	650	367
SRR921664	≥ 8,00	12.053.539	143.626	1.072	680	384
SRR921670	≥ 8,00	13.219.551	150.463	1.220	726	394
SRR941770	≥ 8,00	74.245.568	250.167	2.179	961	405
SRR941772	≥ 8,00	66.321.617	219.250	2.063	929	397

El ensamblado con un mayor número de genes anotados con una cobertura superior al 80% es el ensamblado conjunto, con 7.851 genes (Fig. 3), 854 de los cuales no estaban presentes en los ensamblados individuales a este nivel de cobertura. Debido a ello, y dado que no se realizó un filtrado de la lista de genes anotados con significado biológico (ej. análisis de expresión diferencial), solo se reporta la anotación funcional de este ensamblado. Comparativamente, se anotaron 9.077 y 8.654 genes para los ensamblados de *O. mossambicus* SRR941770 y SRR941772 respectivamente, y 5.616 y 6.219 genes para los ensamblados de *S. alpinus* SRR921664 y SRR921670 respectivamente.



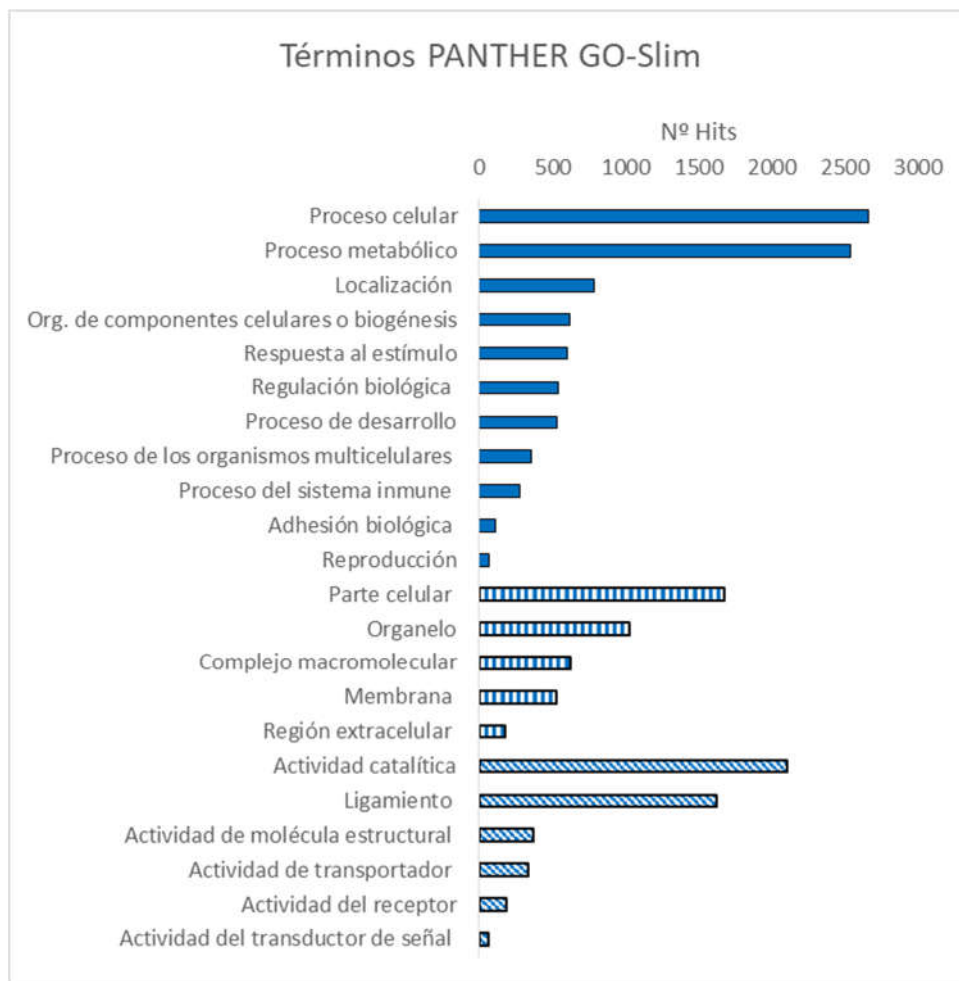


**Fig. 3:** Número de genes anotados mediante BLASTX en los ensamblados realizados (Conjunto, LP-9, LR-20, PC-5, PJ-6, PJ-8 y PJ-10) reconstruidos a niveles crecientes de cobertura. Los distintos ensamblados se muestran con columnas de patrones diferentes. Se escogió al ensamblado conjunto (columna lisa) con un porcentaje de cobertura mayor al 80% para la anotación funcional.

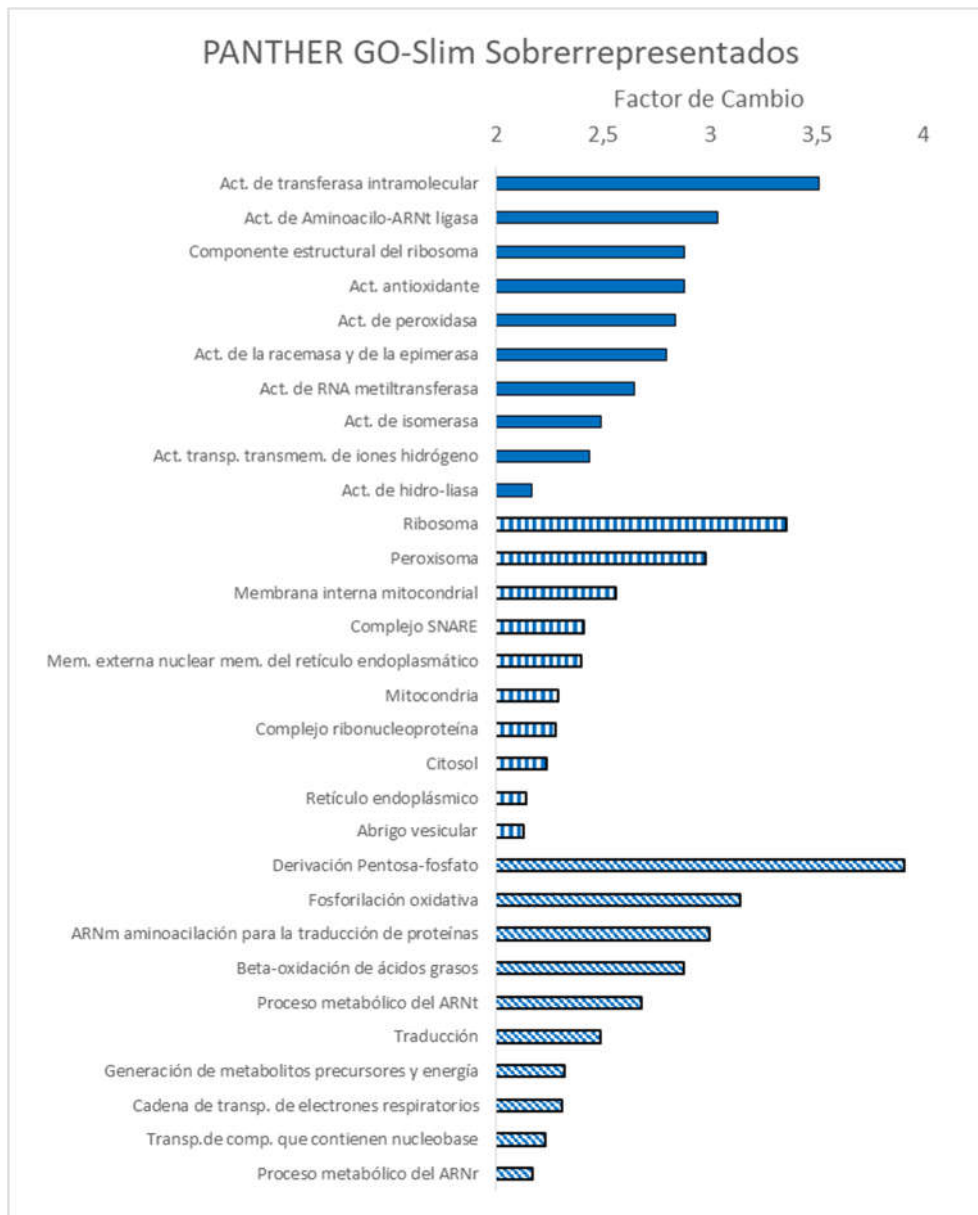
Mediante PANTHER se recuperó información para 6.379 genes anotados en el Ensamblado Conjunto, encontrándose 4.767 hits dentro de la categoría de Funciones Moleculares, 9.141 en Procesos Biológicos y 4.108 en Componentes Celulares (Fig. 4). Los términos GO con más hits dentro de Funciones Moleculares son: “Actividad catalítica” (GO: 0003824) y “Ligamiento” (GO: 0005488). Dentro de Procesos Biológicos los términos GO más representados son “Procesos Celulares” (GO: 0009987) y “Procesos metabólicos” (GO: 0071840). Por último, el principal término GO dentro de Componente celular es “Citoplasma” (GO: 0005737) con 1.023 hits (incluido en el término “Parte Celular” (GO: 0044464)). En total 146 Términos GO fueron encontrados significativamente sobre o subrepresentados (Fig. 5): 41 Función Molecular (Tabla A1), 23 Componente Celular (Tabla A2) y 82 Proceso Biológico (Tabla A3).

Adicionalmente, 5.449 genes fueron identificados como miembros de 26 clases proteicas principales (Tabla A4), 14 de ellas fueron encontradas significativamente sub o sobrerrepresentadas junto con otras 37 subclases (Fig. 6 Tabla A5). Con respecto a las *Pathways* registradas en PANTHER, 2.348 genes pertenecen al menos una de 151 de las *Pathways* (Tabla A6); 3 de ellas fueron encontradas significativamente sobrerrepresentadas con respecto a la base de datos: “*Pathway* del proteasoma de

ubiquitina (P00060)” con un factor de cambio de 2,58 (p-valor  $7,58E^{-06}$ ); “Enfermedad del Parkinson (P00049)” con un factor de cambio de 2,05 (p-valor  $2,08E^{-04}$ ) y “Pathway Ras (P04393)” con un factor de cambio de 1,78 (p-valor  $3,94E^{-02}$ ). Con respecto a las *Pathways* registradas en KEGG, 2.608 genes fueron identificados como miembros de al menos una de 161 *Pathways* diferentes (Tabla A7), de las cuales 21 de ellas fueron encontradas significativamente más representadas (Fig. 7 y Tabla A8).



**Fig. 4:** Términos PANTHER GO-Slim, con más de 50 hits, para Procesos Biológicos (Liso), Componente Celular (Rayado vertical) y Funciones Moleculares (Rayado diagonal).



**Fig. 5:** Los 10 primeros Términos PANTHER GO-Sim de acuerdo a su factor de cambio, significativamente sobrerrepresentados en el Ensamblado Conjunto, para Procesos Biológicos (Liso), Componente Celular (Rayado vertical) y Funciones Moleculares (Rayado diagonal).

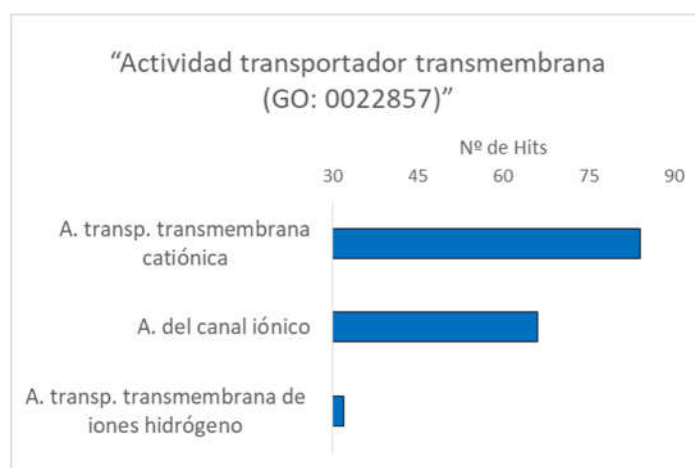


**Fig. 6:** Clases proteicas registradas en PANTHER encontradas significativamente sobrerrepresentadas en el Ensamblado Conjunto. Las Principales clases proteicas, de acuerdo a su nivel jerárquico en PANTHER, están marcadas con asteriscos.



**Fig. 7:** *Pathways* de la base de datos KEGG significativamente sobrerrepresentadas, ordenadas de acuerdo al porcentaje de representación encontrado.

Debido a su participación potencial en la osmorregulación, vale la pena destacar la detección de miembros de la Pathway “Angiotensina 2- señalamiento estimulado a través de proteínas G y beta arrestina (P05911)”, con 16 hits encontrados, debido a que la hormona angiotensina participa en la respuesta inmediata a los aumentos de la salinidad del medio (Takei et al. 2014). Así como 272 genes con el Término GO “Actividad transportador transmembrana (GO: 0022857)” (Fig. 8, Tabla A9).



**Fig. 8:** Términos GO pertenecientes a “Transportadores Transmembrana (GO: 0022857)”.

El análisis de enriquecimiento con respecto a los transcriptomas de *O. mossambicus* y *S. alpinus* no reveló *Pathways* consistentemente sub o sobrerrepresentadas con la salinidad al momento de colecta “Tabla A10”. La *Pathway* “Vía de señalización Wnt (P00057)” fue encontrada subrepresentada en los individuos LR-20, PJ-6 y PJ-10 con respecto a la lista de genes anotados para *O. mossambicus* en agua dulce (SRR941770), mientras que la *Pathway* “Enfermedad Parkinson (P00049)” se encontró sobrerrepresentada en LP-9 y LR-20 con respecto a *O. mossambicus* en agua salada (SRR941772). En ambos casos los individuos fueron colectados en condiciones de salinidad muy dispares. Adicionalmente, el individuo LR-20 presentó la *Pathway* “Coagulación Sanguínea (P00011)” se encuentra sobrerrepresentada con respecto a las 4 transcriptomas usados como referencia y “Regulación del Citoesqueleto por Rho GTPasa (P00016)” fue encontrada sobrerrepresentada para el individuo LP-9 con respecto a los dos transcriptomas de *O. mossambicus*.

Tampoco se identificaron Clases Proteicas registradas en PANTHER que variaran su nivel de representación conforme a la salinidad, ya que se obtuvieron resultados similares comparando con los transcriptomas para agua dulce y salada en ambas especies (Tabla A11).

## 2.4 Discusión

Los términos GO y clases Proteicas en PANTHER, así como KEGG *Pathways* encontradas significativamente sobrerrepresentadas representan principalmente funciones básicas de mantenimiento celular como Síntesis de Proteínas: “Ribosoma (GO:0005840)”, “Actividad de Aminoacilo-ARNt ligasa (GO:0004812)”, “Traducción (GO:0006412)”, “Proteína Ribosómica (PC00202)” y la *Pathway* “Ribosoma (KEGG:03010)”; Generación de energía: “Cadena de Transporte de electrones Respiratorio (GO:0022904)”, “Beta-oxidación de ácidos grasos (GO:0006635)”, “ATP sintasa (PC00002)”, “Glicolisis/Gluconeogénesis (KEGG:04141)”, “Generación de metabolitos precursores y energía (GO:0006091)”; y Funciones Regulatorias: “Modulador enzimático (PC00095)”, “Proteína G (PC00020)”. El análisis de enriquecimiento de *Pathways* registradas en PANTHER es concordante con esta observación. La *Pathway* “*Pathway* del proteasoma de ubiquitina (P00060)” participa en la degradación de Proteínas, los genes integrantes de la *Pathway* “Enfermedad del Parkinson (P00049)” participan en el transporte celular de vesículas, y “*Pathway* Ras (P04393)” participa en la proliferación, supervivencia y diferenciación celular.

Interpretar Términos GO, *Pathways* y Clases Proteicas subrepresentadas como resultado de procesos biológicos desarrollándose en los tejidos es peligroso en este estudio particular debido al filtrado de los genes con una cobertura inferior al 80% favorece la selección de genes de secuencia corta y/o de alta expresión. Este sesgo que seguramente juega un papel importante en la predominancia de genes involucrados en las funciones básicas para el mantenimiento celular ya mencionadas.

Asumiendo que los transcriptomas usados como referencia en este estudio son efectivamente representativos de una secuenciación típica del transcriptoma branquial de peces teleósteos, los datos recabados para este estudio son comparables aunque relativamente inferiores en calidad. Los conteos de genes; así como los valores de calidad de N50, largo promedio y mediana de los contigs ensamblados; son semejantes a los obtenidos a partir de las secuenciaciones para *S. alpinus*, aunque con el doble de reads utilizados. Las diferencias son más notorias con respecto a los transcriptomas de *O. mossambicus*, pero en este caso los datos fueron obtenidos mediante una secuenciación más profunda que la realizada en este estudio (más reads a partir de la misma muestra, aumentando el número promedio de reads conteniendo un fragmento dado de la secuencia original) y a partir de 4 individuos cada una (Lam et al. 2014). Esta pérdida de información fue causada probablemente por la calidad de las extracciones inferior a lo recomendados cuando se realiza un enriquecimiento por pesca poli (A) (RIN > 7) como en este estudio (Conesa et al. 2016). Como un punto positivo sin embargo cabe destacar

que las mediciones de calidad de los ensamblados son semejantes entre sí, independientemente del valor exacto de RIN, por lo que los ensamblados son comparables entre sí.

La falla en detectar *Pathways* y Clases Proteicas significativamente sub o sobrerrepresentadas asociadas con la salinidad puede deberse tanto a la falta de datos, causada tanto por la estrategia de filtrado de genes, los problemas de calidad ya mencionados y al poco poder de detección de los análisis realizados. Aquellas que si fueron encontradas en una proporción distinta a lo esperado por azar responden con toda seguridad a condiciones fisiológicas particulares de cada individuo al momento de la colecta. Por ejemplo la sobrerrepresentación de la *Pathway* “Coagulación Sanguinea (P00011)” en LR-20 fue probablemente causada por la existencia de lesiones en el tejido branquial poco antes o durante la colecta del ejemplar. Es interesante, sin embargo, haber observado sobrerrepresentación en la *Pathway* “Regulación Citoesqueleto por Rho GTPasa (P00016)” para LP-9 tanto en agua dulce y salada porque el mismo patrón fue reportado en *O. mossambicus* (Lam et al. 2014) y asociado con la regulación del pasaje paracelular de iones para compensar los cambios de salinidad (Lam et al. 2014). Los resultados no significativos en el resto de los ejemplares puede explicarse por el hecho de que, conforme a los resultados del trabajo original (Lam et al. 2014), esta vía se encuentra con un alto grado de representación en ambos transcriptomas de *O. mossambicus*. Sin embargo, considerando que solo se detectó en un individuo del que solo se anotaron exitosamente 2.509 genes, no es posible afirmar nada concreto para *O. argentinensis* sin un análisis más detallado.

A pesar de lo expuesto, hay motivo para pensar que en estos ensamblados contienen información biológica relevante para ampliar nuestro conocimiento de *O. argentinensis* mediante estudios más detallados. Los genes pertenecientes a la ya mencionada *Pathway* “Regulación del Citoesqueleto por Rho GTPasa (P00016)” son buenos candidatos a estar bajo expresión diferencial, así como aquellos pertenecientes a la *Pathway* “Angiotensina 2- señalamiento estimulado a través de proteínas G y beta arrestina (P05911)”. La angiotensina regula vasoconstricción, la secreción de aldosterona y movilidad y crecimiento de células del músculo liso (Shenoy & Lefkowitz 2005). Ante un aumento de la salinidad, la liberación de angiotensina estimula la vasoconstricción (aumentando la retención de agua dentro del organismo) e induce el comportamiento de beber (Takei et al. 2014). La liberación de angiotensina también estimula la secreción de cortisol, otra hormona que participa en la aclimatación a agua salada y dulce (Takei et al. 2014). Otros genes potencialmente involucrados en la osmorregulación son los pertenecientes a el término GO “Actividad transportador transmembrana” (GO: 0022857). Particularmente aquellos incluidos en los términos GO Actividad de canal de iones



(GO:0005216) y Actividad transportador de cationes (GO:0008324). Algunos ejemplos concretos son los genes CLCN-3, SLC9A2 y SLC12A2a (ver la sección 3.1.1 Osmorregulación, Tabla C3)

## 2.5 Conclusiones finales Capítulo 2

Se ensamblaron y anotaron exitosamente 6 transcriptomas branquiales de *O. argentinensis*. Aunque limitados por problemas técnicos, los ensamblados contienen información biológica relevante para futuros estudios. Particularmente, la presencia de genes conocidos por su importancia en la osmorregulación (ej. SLC12A2a), indican que los datos recabados serán de utilidad para estudiar la respuesta de la especie a los cambios de salinidad, contribuyendo a ampliar nuestro conocimiento sobre los mecanismos osmorregulatorios de los peces teleósteos.

Estos resultados también abren la puerta para abordar varias clases de preguntas biológicas. Por ejemplo un análisis comparativo entre las especies *O. argentinensis* y *O. bonaerensis*. Estas especies presentan grandes similitudes genéticas y morfológicas, aunque *O. bonaerensis* habita principalmente medios dulceacuícolas (Dyer 2006; Campanella et al. 2015; González-Castro et al. 2016; Setiamarga et al. 2008); por lo que representan un caso interesante de especiación reciente, en el que la selección natural a distintos ambientes podría estar jugando un papel importante (Butlin et al. 2012; Via 2009).

Otras posibles aplicaciones son los estudios filogenéticos y los análisis para detectar evidencia de selección. Datos transcriptómicos de RNAseq han sido utilizados para reconstruir las relaciones filogenéticas de los peces teleósteos (Zou et al. 2012), y en el Capítulo 3 de esta tesis se presenta un ejemplo de la utilización de estos datos para buscar evidencia de selección natural asociada a la osmorregulación en los teleósteos.

# Capítulo 3: Evidencia de selección en genes implicados en osmorregulación

## 3.1 Introducción

### 3.1.1 Mecanismos moleculares de la osmorregulación

Los principales órganos osmorreguladores son las branquias, el intestino y los riñones (Evans 2008; Hwang et al. 2011; Marshall & Grosell 2005; Takei et al. 2014). A continuación se detallan los movimientos de iones y agua en estos órganos en respuesta a los cambios osmóticos y los genes involucrados que servirán como genes candidatos para el análisis propuesto.

#### 3.1.1.1 Branquia

En el medio marino las principales proteínas involucradas en la secreción de  $\text{Cl}^-$  y  $\text{Na}^+$  son: la Na-K-ATPasa (NKA), Co-transportador Na-K-2Cl<sup>-</sup> 1 (SLC12A2), y el Regulador Transmembrana de la Conductancia de la Fibrosis Quística (CFTR).

La NKA cumple la función de generar el gradiente electroquímico que impulsa la secreción de iones, se localiza en la membrana basolateral de los ionocitos y consta de dos subunidades:  $\alpha$  y  $\beta$  (Hwang et al. 2011). En teleósteos existen varios genes parálogos que codifican para estas subunidades (ej. 9 parálogos de la subunidad  $\alpha$  en *D. rerio* (Kumai & Perry 2012)), y hay evidencia de que diferentes subunidades son expresadas en agua dulce y salada (Feng et al. 2002; Madsen et al. 2009; Tang et al. 2009).

Los co-transportadores SLC12A2, también localizados en la membrana basolateral de los ionocitos, representan la ruta de entrada del  $\text{Cl}^-$  corporal al ionocito. En varias especies de teleósteos como *A. anguilla*, *Oreochromis mossambicus* y *Oryzias dancena* este co-transportador es codificado por dos genes parálogos: SLC12A2a y SLC12A2b. De los cuales solo se ha reportado un aumento en la expresión asociado con la aclimatación a agua salada para SLC12A2a (Cutler & Cramb 2002; Hiroi et al. 2008; Kang et al. 2010).

La excreción del  $\text{Cl}^-$  al medio externo ocurre a través del transportador CFTR, localizado en la membrana apical de los ionocitos (Hwang et al. 2011; Takei et al. 2014). Varios trabajos han reportado un incremento en los niveles de expresión tras la aclimatación a agua salada en diversas especies (Tang & Lee 2007; Tse et al. 2006). En cuanto a la eliminación del  $\text{Na}^+$ , los iones son liberados al medio por vía paracelular a través de las uniones estrechas (Hwang et al. 2011). Se ha reportado un relajamiento de estas uniones como parte de la aclimatación al agua salada (Hwang 1987), así como cambios en los niveles de expresión de varias claudinas (Bagherie-Lachidan et al. 2009; Tipsmark et al. 2008).

Los mecanismos de captura de iones en los medios dulceacuícolas son menos comprendidos que los de secreción. Sin considerar aquellos no apoyados por evidencia molecular, la vía más aceptada de entrada del  $\text{Na}^+$  y el  $\text{Cl}^-$  es la mediada por un co-transportador de  $\text{Na}^+$  y  $\text{Cl}^-$  (SLC12A10), localizado en la membrana apical de algunos ionocitos (Hwang et al. 2011; Takei et al. 2014). Secundariamente, la actividad de intercambiadores  $\text{Na}^+/\text{H}^+$  (SLC9) representaría parte de la captura de iones de sodio; aunque la función principal de estos últimos sería la regulación del PH corporal (Hwang et al. 2011; Takei et al. 2014). Particularmente, en la mayoría de los casos parecen ser isoformas de SLC9A3 las involucradas en esta función (Hwang et al. 2011; Takei et al. 2014). Aunque en algunas especies como *Oncorhynchus mykiss* la proteína SLC9A2 también estaría involucrada (Ivanis et al. 2008).

Una vez en el interior de los ionocitos, el  $\text{Na}^+$  atraviesa la membrana basolateral vía la NKA o vía el co-transportador electrogénico  $\text{Na}^+/\text{HCO}_3^-$  (SLC4A4) (Hwang et al. 2011; Takei et al. 2014), aunque algunos estudios sugieren que en algunas especies, SLC4A4 es reemplazado por AE1b (SLC4A1b) (Lee et al. 2011).

### 3.1.1.2 Intestino

Aunque la absorción de  $\text{Na}^+$  de la dieta es seguramente importante (Marshall & Grosell 2005), son los mecanismos de absorción de agua en un medio hipertónico los que han sido estudiados en mayor profundidad (Grosell 2010).

Inicialmente, el agua es desalinizada parcialmente a nivel del esófago mediante la absorción del  $\text{Na}^+$  y el  $\text{Cl}^-$ , pero manteniendo una alta impermeabilidad del tejido al agua (Grosell 2010; Marshall & Grosell 2005). Aunque no se conocen en detalle los mecanismos moleculares para esta desalinización inicial, la evidencia apunta a la toma activa de iones por parte de miembros de la SLC9 (Grosell 2010). Recientemente, se ha identificado SLC9A2 como uno de los puntos de entrada del sodio en este tejido, al menos en la especie *Opsanus beta* (Esbaugh & Grosell 2014). Posteriormente, en el lumen intestinal, el  $\text{Na}^+$  es absorbido en la membrana apical a favor de su gradiente electroquímico mediante las proteínas SLC12A1 y SLC12A10 (Cutler & Cramb 2008; Grosell 2010; Watanabe et al. 2011). Adicionalmente, SLC9A2 y SLC9A3 podrían ser entradas alternativas para el  $\text{Na}^+$ , aunque la evidencia a favor de su participación es menor (Grosell 2011).

En adición a la actividad de SLC12A1 y SLC12A10, la entrada de  $\text{Cl}^-$  está asociada a la secreción  $\text{HCO}_3^-$  (Wilson et al. 1996), siendo el intercambiador de aniones SLC26A6 un candidato fuerte a cumplir esta función (Grosell 2010; Grosell 2011; Kurita et al. 2008). La secreción de  $\text{HCO}_3^-$  al lumen intestinal es crucial para la absorción del agua. Este es

convertido en  $\text{CO}_2$  por la Anhidrasa Carbónica ligada a membrana (CAIV), que luego es tritado en  $\text{CO}_3^-$ , facilitando la producción de precipitados de  $\text{CaCO}_3$ ; dando como resultado una reducción significativa de la osmolaridad del lumen intestinal (Grosell 2010). El  $\text{HCO}_3^-$  secretado entra a las células vía SLC4A4 localizado en la membrana basolateral, o es generado por hidratación de  $\text{CO}_2$  endógeno por parte de la Anhidrasa Carbónica Citosólica (CAc) (Grosell 2010).

### 3.1.1.3 Riñón

Considerablemente menos se sabe de los mecanismos moleculares involucrados en el procesamiento de la orina (Teranishi et al. 2013). Si bien los teleósteos no son capaces de producir orina hipertónica con respecto a su sangre, como sí hacen los mamíferos y aves, su osmolaridad está dada principalmente por la presencia de iones  $\text{Ca}_2^+$ ,  $\text{Mg}_2^+$ ,  $\text{SO}_4^{2-}$ ; el  $\text{Na}^+$  y el  $\text{Cl}^-$  son reabsorbidos de la orina primitiva para facilitar la reabsorción del agua (Marshall & Grosell 2005). El mismo mecanismo se utiliza para la generación de orina hipotónica en peces dulceacuícolas, pero manteniendo la impermeabilidad al agua de los túbulos distales y colector (Kato et al. 2011; Marshall & Grosell 2005). Esta tarea sería desempeñada por los transportadores SLC12A1, SLC12A3 y SLC9A3 (Kato et al. 2011; Teranishi et al. 2013).

### 3.1.1.4 Otros genes analizados

De los genes mencionados en la sección previa, se buscó evidencia de selección positiva en secuencias de CFTR, SLC9A2, SLC12A1, SLC12A2, SLC12A3, SLC26A6 (Ver la sección 3.2 Metodología). También se incluyeron otros 7 genes que no fueron mencionados antes porque su participación en la osmorregulación es indirecta y/o sugerida por una menor cantidad de evidencia. Estos son:

- 1) AQP-9b, -10b y -12: Las Acuaporinas son canales que permiten la difusión de agua a través de las membranas plasmáticas. Aquí se analizaron las Acuagliceroporinas AQP-9b y 10b, y la acuaporina no-ortodoxa AQP-12. Muy poco se sabe de la función de estas acuaporinas, pero han sido detectadas en branquia, intestino y riñón (Cerdà & Finn 2010).
- 2) CLCN-3: También denominado CLC-3, este canal de cloruro ha sido detectado en branquias *D. labrax* y *O. mossambicus* (Bossus et al. 2013; Miyazaki et al. 1999). En branquia, Bossus et al. (2013) sugieren que podría cumplir la función de prevenir cambios en el volumen celular, por reducciones rápidas en la osmolaridad, al permitir la salida de la célula del  $\text{Cl}^-$ .

- 3) SLC9A1: Previamente se mencionó la importancia de la secreción intestinal de  $\text{HCO}_3^-$  en la desalinización del agua marina ingerida, para su posterior absorción. Sin embargo no se mencionó que la síntesis endógena de  $\text{HCO}_3^-$  genera un exceso de  $\text{H}^+$ ; el cual debe ser eliminado para mantener el PH celular y permitir la acumulación del  $\text{HCO}_3^-$  (Grosell 2011). Basándose en que la secreción de  $\text{HCO}_3^-$  depende de la concentración serosa de  $\text{Na}^+$  y es inhibida por ouabaina (Grosell & Genz 2006), Grosell (2011) propone a SLC9A1, localizado en la membrana basolateral, como responsable de dicha función.
- 4) VDAC-1 y -2: A través de estas porinas ocurre el intercambio de ATP/ADP a través de la membrana mitocondrial (Schredelseker et al. 2014). Varios estudios han detectado aumentos de la expresión de estas proteínas ante cambios en la salinidad (Lam et al. 2014; Tine et al. 2008; Tine et al. 2011).

### 3.1.2 Análisis de selección basados en las tasas de sustitución

La comparación de las tasas de sustituciones sinónimas ( $\alpha$ ) y no sinónimas ( $\beta$ ) es una aproximación ampliamente utilizada para buscar evidencia de selección natural a nivel molecular (Kosakovsky et al. 2011; Murrell et al. 2012; Lu & Guindon 2013; Yang 2007). El objetivo básico es estimar  $\alpha$  y  $\beta$ , para poder determinar si el cociente  $\beta/\alpha$  ( $\omega$ ) es mayor, menor o igual a 1; indicativo de selección positiva, purificadora o neutralidad respectivamente. Con el correr de los años se han desarrollado varios métodos para detectar selección natural a partir de alineamientos de secuencias ortólogas, evaluando la evidencia a nivel de sitios, linajes o ambos (ej. Kosakovsky et al. 2011; Murrell et al. 2012; Yang 2007).

El programa CodeML, implementado en el paquete PAML, destaca por su versatilidad para evaluar distintos escenarios evolutivos (Yang 2007). Este estima las tasas de sustitución por máxima verosimilitud bajo una gran variedad de modelos evolutivos, los que permiten buscar evidencia de selección positiva actuando a nivel de linajes, sitios o ambos (Yang 2007).

Trabajando a nivel de linajes, CodeML permite construir modelos por ramas con diferentes asunciones sobre la presión selectiva en linajes pre-seleccionados (ramas *foreground*) (Yang 1998), las cuales son evaluadas en contraste al resto del conjunto de datos (ramas *background*). Las diferentes hipótesis modeladas pueden ser luego evaluadas mediante un Likelihood Ratios Test (LRT). Para buscar evidencia de selección a nivel de sitios, CodeML posee 6 modelos de evolución a nivel de codones (Yang 2000; Nielsen & Yang 1998; Yang et al. 2005):

- 1) M0 asume que todos los sitios están bajo el mismo  $\omega$ .
- 2) M1a divide a los sitios en dos categorías, una con  $\omega=1$  y otra con  $0 < \omega < 1$ . Su contraparte M2a incorpora una clase de sitios ( $\omega > 1$ , estimado a partir de los datos).
- 3) M3 asume una distribución general discreta del valor  $\omega$ .
- 4) M7 asume que el valor de  $\omega$  sigue una distribución beta entre 0 y 1. Su contraparte M8 incorpora una clase de sitios extra que permite la selección positiva con  $\omega > 1$ , estimado a partir de los datos.

Los modelos mencionados hasta ahora asumen la ausencia de variación en la presión selectiva entre codones o entre linajes, no ambos. Combinando ambos métodos se encuentran los modelos Ramas-Sitios, también implementados en CodeML (Yang & Nielsen 2002; Yang 2007). Estos modelos permiten a  $\omega$  variar tanto entre linajes como sitios, haciéndolo más apropiado para detectar genes bajo selección direccional cuando solo unos pocos codones se encuentran bajo selección positiva

Al igual que los modelos de Ramas, los modelos de Ramas-Sitios de CodeML requieren que se pre-seleccionen los linajes a ser estudiados en las categorías ya mencionadas. Los sitios por su lado se dividen en 4 categorías (Yang & Nielsen 2002): Sitios conservados (Categoría 0, con  $\omega \approx 0$ ); Sitios Neutrales o levemente restringidos (Categoría 1,  $\omega = 1$ ); Sitios potencialmente bajo selección positiva en las ramas foreground y purificadora en el background (Categoría 2a,  $\omega$  estimado); o sitios potencialmente bajo selección positiva en las ramas foreground y neutrales en el background (Categoría 2b,  $\omega$  estimado).

La principal limitante de los modelos por ramas y Ramas-Sitios implementados en CodeML es la necesidad de seleccionar las ramas foreground. Mientras que estos modelos son relativamente robustos a problemas como la saturación de sustituciones, manteniendo baja la tasa de falsos positivos (Gharib & Robinson-Rechavi 2013); las pruebas de simulación muestran que estos modelos son muy sensibles a errores en la selección de ramas (Kosakovsky et al. 2011; Lu & Guindon 2013). Un problema grave cuando no se dispone de conocimiento biológico con el que respaldar una elección de ramas particular (Kosakovsky et al. 2011; Lu & Guindon 2013). Se han propuesto varios métodos alternativos que no requieren la selección a priori de ramas (Anisimova & Yang 2007; Guindon et al. 2004; Kosakovsky et al. 2011; Murrell et al. 2012; Rodrigue et al. 2010), que resultan más apropiados para un análisis exploratorio. En esta tesis se utilizaron dos de estos métodos: Ramas-Sitios REL (BSR) (Kosakovsky et al. 2011) y Modelo de Evolución de Efectos Mixtos (MEME) (Murrell et al. 2012).

A diferencia del Modelo de Ramas-Sitios de CodeML, BSR no asume la existencia de únicamente 4 categorías de sitios sino que considera todos los patrones de evolución

posibles (Kosakovsky et al. 2011). Aunque el modelo de Ramas-Sitios de PAML es ligeramente más eficiente para detectar evidencia de selección positiva cuando sus asunciones son una buena aproximación a la historia evolutiva del grupo estudiado, BSR es más flexible y es más eficiente detectando selección episódica (Kosakovsky et al. 2011).

Por su parte, MEME es una modificación de los métodos utilizados en BSR para detectar selección a nivel de codones individuales; permitiendo variar la fuerza de selección simultáneamente entre sitios y linajes (Murrell et al. 2012). Al no necesitar la selección a priori de ramas, este método es más apropiado para estudios exploratorios y es capaz de detectar tanto eventos de selección episódica como constante en el tiempo (Murrell et al. 2012). En este método, los codones de una posición dada son divididos en dos categorías discretas, cada una con una serie de parámetros libres que determinan la fuerza de la selección (Kosakovsky et al. 2011): restringida  $\beta^-$  ( $\beta \leq \alpha$ ) y no restringida  $\beta^+$  (sin restricciones en los valores de  $\beta$ ).

## 3.2 Metodología

### 3.2.1 Filogenia utilizada

Para este estudio, la topología de los árboles utilizados para los análisis se construyó en base a lo reportado por varios autores. Principalmente se siguió el trabajo de Betancur-R et al. (2013) para determinar las relaciones filogenéticas de los principales grupos. Cuando fue necesario se recurrió a otros trabajos, concretamente para la resolución de las relaciones internas de los linajes Cypriniformes (Mayden et al. 2009), Gymnotiformes (Tagliacollo et al. 2016), Cichlidae (Sparks & Smith 2004) y Cyprinodontiformes (Pohl et al. 2015). Finalmente se consultó la base de datos Fishbase (<http://www.fishbase.org/search.php>) para conocer el ambiente de cada especie estudiada, distinguiendo entre especies dulceacuícolas, marinas y estuarinas, y también aquellas que pueden tolerar breves incursiones en medios estuarinos.

### 3.2.2 Obtención de datos

Primero se realizó una revisión bibliográfica sobre las bases moleculares de los mecanismos osmorregulatorios en teleósteos y se pre-seleccionaron genes disponibles para la especie modelo *D. rerio* en la base de datos Ensembl (Herrero et al. 2016) (43 genes sin considerar duplicaciones no compartidas por un amplio número de Actinoptergios). Luego se consultó la base de datos OMA browser (Versión Mayo 2016) (Altenhoff et al. 2015) para determinar cuáles de estos presentaban ortólogos uno a uno

en las especies *Astyanax mexicanus*, *Gadus morhua*, *Gasterosteus aculeatus*, *Lepisosteus oculatus*, *Oreochromis niloticus*, *Oryzias latipes*, *Takifugu rubripes*, *Tetraodon nigroviridis* y *Xiphophorus maculatus*. Se buscó seleccionar genes en los que se pudiera determinar una ortología uno a uno entre estas nueve especies y *D. rerio*. Sin embargo se tuvo tolerancia en caso que estuvieran ausentes en solo una especie, o estas fueran: Los Tetradontiformes (ej. *T. rubripes* y *T. nigroviridis*), porque el grupo ha experimentado un proceso de compactación del genoma (Noletto et al. 2009; Jaillon et al. 2004); Lepisosteiformes (ej. *L. oculatus*) porque el grupo divergió del resto de los peces actuales antes de la duplicación genómica completa específica de teleósteos (WGD) (Glasauer & Neuhaus 2014). En total 20 genes pasaron esta etapa de selección.

De la base de datos Ensembl se descargaron las secuencias codificantes (CDS), asociadas a las entradas de los genes identificados como ortólogos en OMA-Browser, a los que se sumaron las CDS de *Poecilia formosa* (ver más abajo), sumando 10 especies. Complementando este conjunto de datos, se incluyeron las secuencias obtenidas en el Capítulo 2 de esta tesis para *O. argentinensis*, y secuencias para otras 42 especies de actinopterigios registradas en el NCBI. De estas últimas, unas 15 especies están registradas en la base de datos Protein (<https://www.ncbi.nlm.nih.gov/protein>) y otras 27 en el Archivo de Secuencias de Reads (SRA) (<https://www.ncbi.nlm.nih.gov/sra>) (Tabla 2). Ambas bases de datos están a cargo del NCBI y fueron consultadas en Febrero del 2016.

Las especies de la base de datos Protein fueron elegidas por la disponibilidad de secuencias para los genes seleccionados. Se descargaron todas las secuencias proteicas para ellas y se determinó sus relaciones de ortología con *D. rerio* mediante un BLAST recíproco (Altschul et al. 1990; Fuchsman & Rocap 2006; Ward & Moreno-Hagelsieb 2014), una aproximación a la verdadera ortología comúnmente utilizada por su simplicidad (ej. Tsagkogeorga et al. 2015; Ward & Moreno-Hagelsieb 2014; Zou et al. 2012). Para esto se utilizaron como referencia las secuencias proteicas de *D. rerio* disponibles en OMA Browser y se seleccionaron las mejores correspondencias recíprocas en base a su Bit-score (Pearson 2013). El mismo procedimiento se realizó con las CDS descargadas para *P. formosa* de la base de datos Ensembl.

Finalmente, las especies de la base de datos SRA fueron seleccionadas buscando aumentar la cobertura de linajes reportados por Betancur-R et al. (2013), y dentro de los datos disponibles para cada una transcriptomas de los órganos: Branquia, Intestino y Riñón. Cuando fue posible se descargaron los 3 órganos (ej. *Pangasianodon hypophthalmus* y *Plecoglossus altivelis*). La calidad de estos datos descargados fue evaluada con el programa FastQC v0.10.1 y la edición se realizó con el programa Trimmomatic v0.35. Con el objetivo de minimizar las posibilidades de incluir errores de



**Tabla 2:** Lista de las especies utilizadas y la base de datos de procedencia.

Especie	Fuente	Especie	Fuente
<i>Acipenser schrenkii</i>	SRA	<i>Maylandia zebra</i>	Protein
<i>Anguilla japonica</i>	SRA	<i>Megalobrama amblycephala</i>	SRA
<i>Apteronotus albifrons</i>	SRA	<i>Moronoe saxatilis</i>	SRA
<i>Astyanax mexicanus</i>	ENSEMBL	<i>Neolamprologus brichardi</i>	Protein
<i>Austrofundulus limnaeus</i>	Protein	<i>Nibeia albifora</i>	SRA
<i>Brienomyrus brachyistius</i>	SRA	<i>Notothenia coriiceps</i>	Protein
<i>Clupea harengus</i>	Protein	<i>Odontesthes argentinensis</i>	Sec. Propia
<i>Coilia nasus</i>	SRA	<i>Oncorhynchus mykiss</i>	SRA
<i>Coregonus peled</i>	SRA	<i>Oreochromis niloticus</i>	ENSEMBL
<i>Cynoglossus semilaevis</i>	Protein	<i>Oryzias latipes</i>	ENSEMBL
<i>Cyprinodon variegatus</i>	Protein	<i>Pangasianodon hypophthalmus</i>	SRA
<i>Danio rerio</i>	ENSEMBL	<i>Platichthys stellatus</i>	SRA
<i>Diplodus puntazo</i>	SRA	<i>Plecoglossus altivelis</i>	SRA
<i>Electrophorus electricus</i>	SRA	<i>Poecilia formosa</i>	ENSEMBL
<i>Elopichtys bambusa</i>	SRA	<i>Pundamilia nyererei</i>	Protein
<i>Esox lucius</i>	Protein	<i>Salivenus alpinus</i>	SRA
<i>Fundulus heteroclitus</i>	Protein	<i>Salmo salar</i>	Protein
<i>Gadus morhua</i>	ENSEMBL	<i>Schizothorax richardsonii</i>	SRA
<i>Gasterosteus aculeatus</i>	ENSEMBL	<i>Sebastiscus marmoratus</i>	SRA
<i>Haplochromis burtoni</i>	Protein	<i>Siniperca chuatsi</i>	SRA
<i>Hippocampus hippocampus</i>	SRA	<i>Stegastes partitus</i>	Protein
<i>Ictalurus furcatus</i>	SRA	<i>Sternopygus macrurus</i>	SRA
<i>Labeo rohita</i>	SRA	<i>Takifugu rubripes</i>	ENSEMBL
<i>Larimichthys crocea</i>	Protein	<i>Tetraodon nigroviridis</i>	ENSEMBL
<i>Lepisosteus oculatus</i>	ENSEMBL	<i>Tripterygion delaisi</i>	SRA
<i>Leuciscus burdigalensis</i>	SRA	<i>Xiphophorus maculatus</i>	ENSEMBL
<i>Lutjanus erythropterus</i>	SRA		

secuenciación en el ensamblado final, se optó por un enfoque más conservador que en el Capítulo 2 para la edición de los *reads*. En adición a las secuencias adaptadoras se eliminaron: las primeras pares de bases en las que FastQC identificó sesgos en la composición de bases en el extremo 5', y las bases del extremo 3' si una porción importante de estos presentaba una calidad  $Q < 10$  y los *reads* de baja calidad cuando una porción significativa de estos presentó una calidad promedio  $Q=2$ . Los comandos utilizados en cada caso se detallan en el Anexo B.

Excepciones a este plan general fueron las secuencias de *Tripterygion delaisi* (SRR768809) y *Sebastiscus marmoratus* (SRR491336) por ser secuenciaciones *single-end*; en estos sets de datos se eliminaron las bases al extremo 5' con calidad  $Q < 7$  y se conservaron los *reads* con un largo superior a 30 pb. Otra excepción fue *Schizothorax richardsonii* (SRR1552917) debido a un error en la versión del archivo SRA descargada. Parte de la secuencia de un par de los *reads* se encuentra incluida en el otro *read*. La región afectada fue eliminada. Por último, debido a limitaciones de Hardware, los datos de las especies *Sternopygus macrurus* (SRR1299088), *Electrophorus electricus* (SRR1299086), *Platichthys stellatus* (SRR1023744), *Ictalurus furcatus* (SRR070302) fueron normalizados. En estos casos se eliminaron los *reads* de baja calidad promedio ( $Q < 5$ ) y se normalizaron con el programa TrinityNorm (Haas et al. 2013)

Los transcritos fueron anotados mediante un BLAST recíproco como el ya descrito. Las CDS fueron identificadas mediante el algoritmo TransDeoder (Haas et al. 2013) para identificar la región codificante de los transcritos reconstruidos, seguido de curación manual. Se descartaron aquellos con una cobertura inferior al 75% de la secuencia de referencia correspondiente de *D. rerio*.

### 3.2.3 Alineamiento de secuencias y curación manual

En este punto se hizo una nueva reducción de la lista de genes. Se seleccionaron de acuerdo a su importancia funcional y a la disponibilidad de datos.

Las secuencias para los genes seleccionados fueron alineadas con el algoritmo MAFFT (Kato & Standley 2013) en sus opciones por defecto, implementado en el servidor de EMBL-EBI. Los alineamientos resultantes fueron curadas manualmente en el programa MEGA6 (Tamura et al. 2013). Cuando fue necesario, se dividió el alineamiento en bloques separados regiones del alineamiento ricas en múltiples indels (inserciones y/o deleciones) taxón específicos. La curación manual se realizó por etapas:

- 1) Se identificaron regiones de alta divergencia, que afectaran una o pocas especies. Como la mayoría de las secuencias con este problema corresponden a datos

provenientes de la base de datos Ensembl, para genes identificados como ortólogos uno a uno en OMA Browser, se asumió que estos representaban diferencias en el uso de exones o errores en su delimitación. Estas regiones fueron eliminadas y se re-alinearon las secuencias.

- 2) Se eliminaron las inserciones compartidas por menos de 10 especies. Debido a que el algoritmo MAFFT no tiene en cuenta la secuencia proteica, los indels en el alineamiento no necesariamente empiezan en la posición 1 de un codón. En estos casos primero se intentó desplazar el gap 1 o 2 pares de bases (pb) para hacerlo coincidir con el marco de lectura, siempre y cuando la secuencia respaldara el cambio en una única dirección. En caso contrario también fue eliminado el codón afectado por la inserción de todo el alineamiento.
- 3) Se reajustaron las inserciones presentes en más de 10 especies, pero sin eliminar los codones interrumpidos cuando la secuencia no respaldó un desplazamiento sobre el otro.
- 4) Se editaron los gaps mayores a 10 codones. Estos gaps raramente coincidían con el marco de lectura y presentaban codones incompletos en ambos extremos, ocasionalmente con secuencias de pocos pares de bases (pb) mal alineadas en el centro. Estas bases también fueron eliminadas.

#### 3.2.4 Análisis de selección

Para realizar un análisis exploratorio de los linajes bajo selección positiva se utilizó el algoritmo BSR (Kosakovsky et al. 2011), implementado en el programa HyPhy y disponible en el servidor DataMonkey (Delport et al. 2010). Los linajes encontrados bajo selección positiva por este fueron usados como ramas foreground para los modelos de CodeML, implementado en PAML4 (Yang 2007), vía la interfaz gráfica pamIX (Xu & Yang 2013). En concreto, se evaluaron modelos de selección por ramas (NSsites=0, model=2) y por Ramas-Sitios (NSsites=2, model=2).

Se evaluaron 3 modelos por ramas: con un  $\omega$  para las ramas foreground (MA) y con un  $\omega$  independiente por rama foreground (MB), ambos evaluados entre sí mediante un LRT y con un modelo nulo (MCon) (NS=0, m=2, fix\_omega=1, omega=1). Por su lado, los modelos de Ramas-Sitios considerados analizaron cada linaje sugerido por BSR de forma individual, y fueron comparados con sus respectivos modelos nulos (fix\_omega=1, omega=1), siguiendo las sugerencias de la documentación de PAML4. Todos estos LRT se realizaron con un umbral de significancia de 3,84 (p-valor  $\leq 0,05$ ).

A nivel de codones, se corrieron 6 de los modelos de sitios (M0, M1a, M2a, M3, M7, M8) disponibles en el algoritmo CodeML, y se realizaron los LRT también, sugeridos en la documentación de PAML4 para evaluar variaciones en el  $\omega$  entre sitios con un umbral de significancia de 3,84 ( $p\text{-valor} \leq 0,05$ ): M0 vs M3, M1a vs M2a, M7 vs M8. La primera comparación de modelos analiza la existencia de variación en el  $\omega$  entre sitios, mientras que el resto analizan la presencia de selección natural con modelos de creciente complejidad. También se realizó el LRT M8a vs M8, con un umbral de significancia de 2,71 ( $p\text{-valor} \leq 0,05$ ). El modelo M8a es una modificación del M8 para servir como modelo nulo ( $NSsites = 8$ ,  $fix\_omega = 1$ ,  $omega = 1$ ).

Para identificar codones individuales bajo selección positiva se utilizó algoritmo MEME (Murrell et al. 2012), implementado en el programa HyPhy y disponible en el servidor DataMonkey, así como los modelos de Sitios y Ramas-Sitios ya mencionados, utilizando Bayes Empirical Bayes (BEB) (Yang et al. 2005) para identificar los codones candidatos. Con el objetivo de filtrar falsos positivos se siguió el ejemplo de otros estudios similares (ej. Forni et al. 2015; Mozzi et al. 2016; Stager et al. 2014) y solo se consideraron los codones identificados bajo selección positiva que hayan sido detectados conjuntamente por MEME y alguno de los modelos de CodeML, o mostraran un q-valor significativo en MEME. Como los modelos de CodeML utilizados tienen restricciones más severas que MEME; el modelo de Sitios asume que todos los linajes están bajo la misma presión selectiva y el de Ramas-Sitios que solo el linaje seleccionado como foreground está bajo selección positiva; en estas comparaciones se incluyeron sitios con un p-valor no significativo en los modelos de CodeML. También se consideraron los codones con un q-valor significativo en MEME.

Para evaluar la posible importancia funcional de los codones detectados bajo selección, se recurrió a la base de datos PFAM (Finn et al. 2014) para anotar los dominios proteicos presentes en los genes estudiados y se consultó la base de datos Uniprot (Bateman et al. 2015) para verificar si alguno de los sitios identificados está asociado un sitio con una función conocida en *Homo sapiens* y/o *Mus musculus*. En concreto, se consultaron las isoformas canónicas de los genes CFTR (P13569-1), SLC12A1 (Q13621-1), SLC12A2 (P55011-1), SLC26A6 (Q9BXS9-1) y VDAC-2 (P45880-3) de *H. sapiens*, y SLC9A1 (Q61165-1) de *M. musculus*. La homología entre sitios de estas secuencias canónicas y los actinopterigios estudiados fue determinada alineando las secuencias proteicas mediante Muscle (Edgar 2004) en el paquete MEGA6 (Tamura et al. 2013).

Por último, se utilizó el programa Cn3D v.3.4 (Wang et al. 2000) para visualizar la localización en la estructura terciaria de los sitios encontrados bajo selección positiva para los genes de los que se dispone modelos tridimensionales de su estructura. En concreto

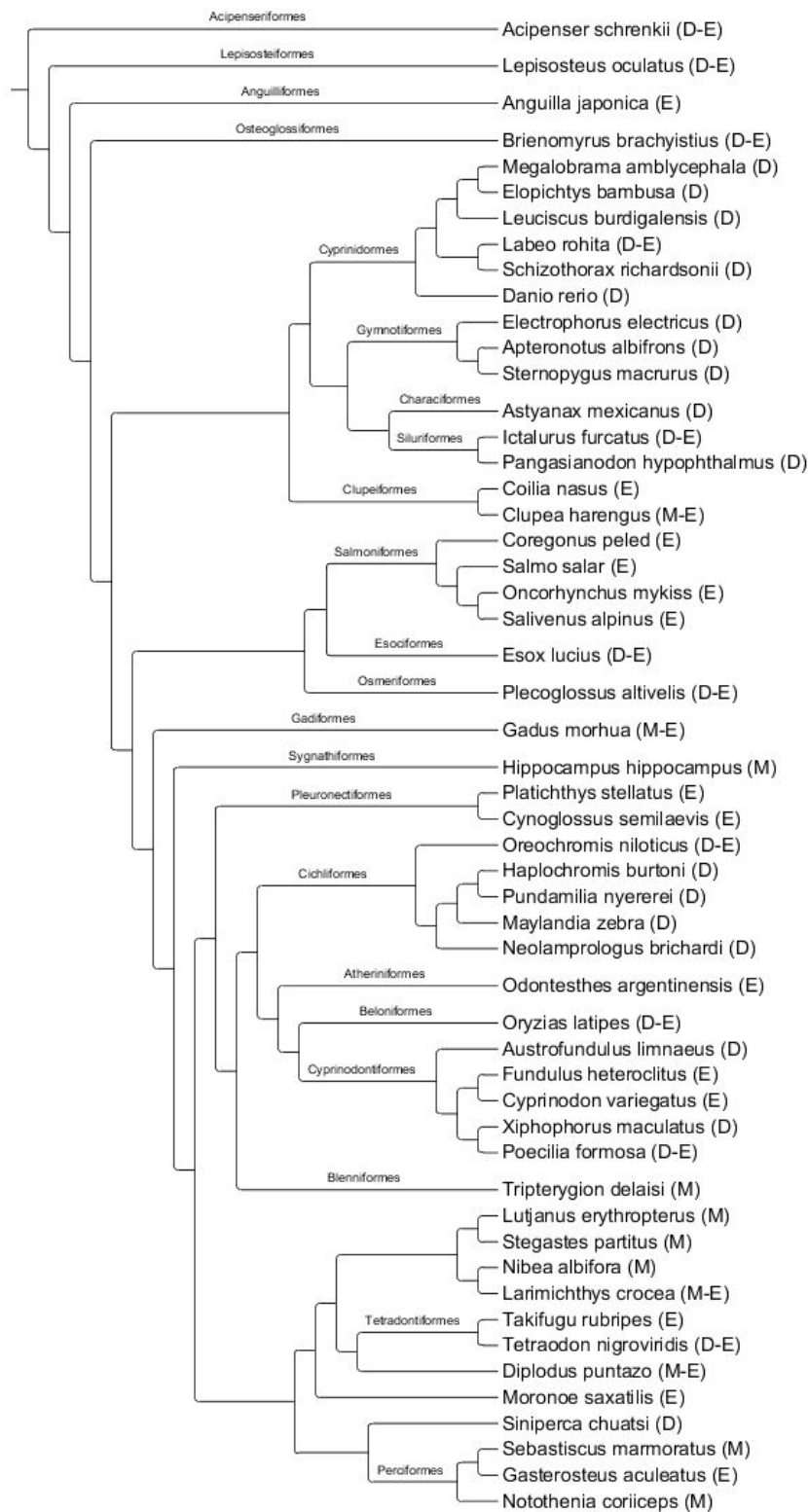
los genes CFTR (Zhang & Chen 2016) (ID: 5UAR) y VDAC-2 (Schredelseker et al. 2014) (ID: 4BUM).

### 3.3 Resultados

En la Figura 9 se detalla la topología de la filogenia utilizada, construida en base a la revisión bibliográfica, junto con las condiciones de salinidad preferidas por cada especie según la base de datos Fishbase. Las 53 especies analizadas comprenden 22 de los 66 órdenes de teleósteos definidos en el trabajo de Betancur-R (2013). Siendo los más representados los órdenes Cichliformes (5 especies), Cypriniformes (6 especies), Cyprinodontiformes (5 especies) y Salmoniformes (4 especies).

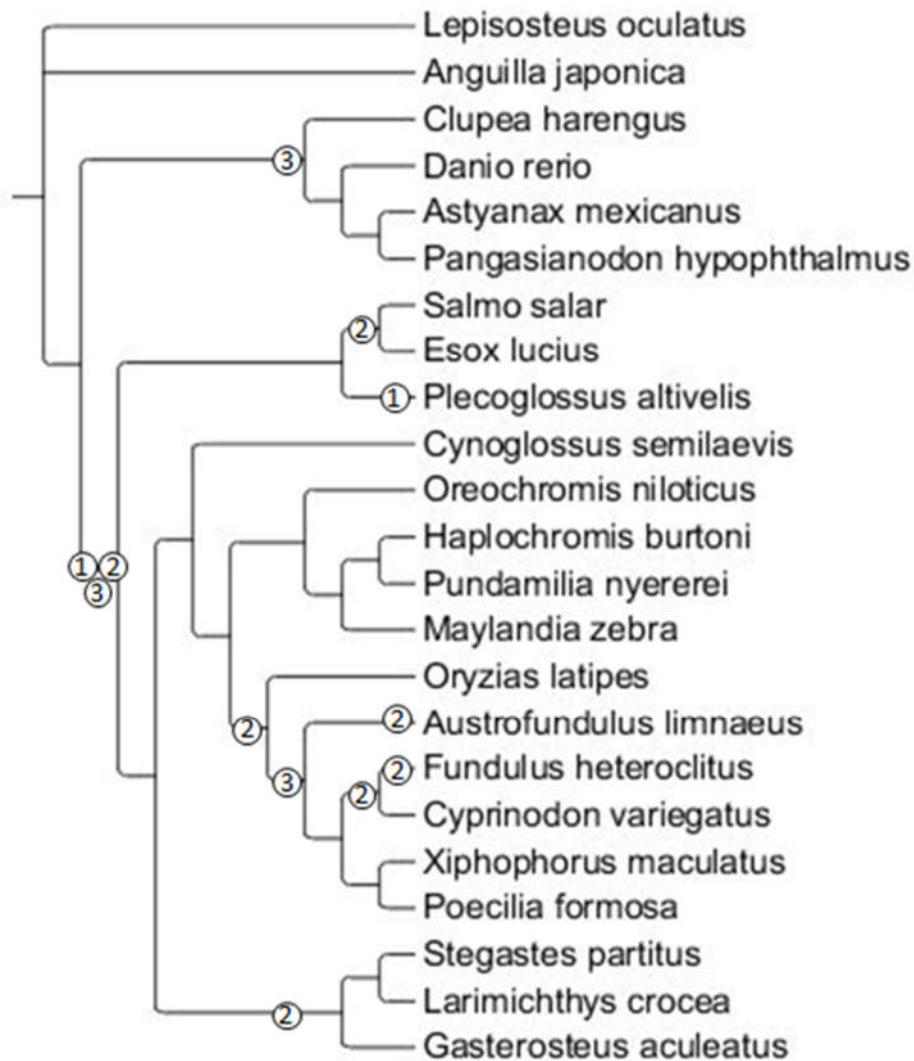
Por simplicidad, se dividieron los genes en dos grupos de acuerdo a la disponibilidad de secuencias: Grupo A con 23 especies (genes AQP10b, CLCN-3, SLC9A1, SLC9A2 y SLC12A2a) (Fig. 10) y Grupo B con 19 especies (genes SLC12A3, AQP-12 y AQP-9b) (Fig. 11) respectivamente. También se analizaron los genes VDAC-1 (41 especies) (Fig. 12) y VDAC-2 (52 especies) (Fig. 13) por la abundancia de secuencias obtenidas, y los genes CFTR (26 especies) (Fig. 14) SLC12A1 (32 especies) (Fig. 15) y SLC26A6 (31 especies) (Fig. 16) por su importancia funcional. La disponibilidad de secuencias para los genes analizados se detalla en el Anexo C. Debido a falta de datos, solo se analizaron 13 genes de los 20 seleccionados.

Los resultados para el gen CLCN-3 no fueron confiables por motivos que se detallan más adelante. También llama la atención que las secuencias de *T. rubripes* y *T. nigroviridis* para el gen VDAC-2 sean idénticas. Por ello no se incluyó la secuencia de *T. nigroviridis* para este gen.

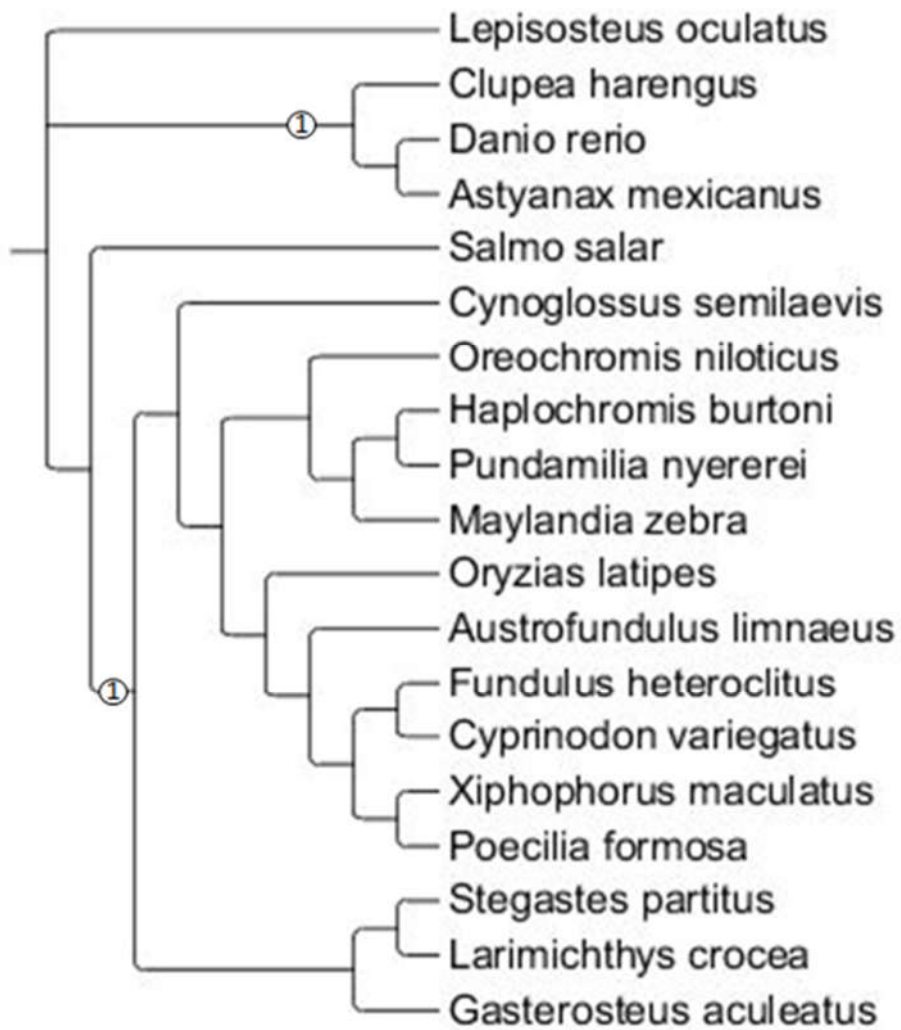


**Fig. 9:** Filogenia utilizada, principalmente basada en el trabajo de Betancur-R et al. (2013). Se marcan los órdenes incluidos en este estudio, y el tipo de medio preferido por

cada especie de acuerdo a Fishbase: Dulceacuícola (D), Estuarino (E) o Marino (M). Se señalaron con un “-E” las especies pueden ocupar temporalmente medios con condiciones intermedias de salinidad.

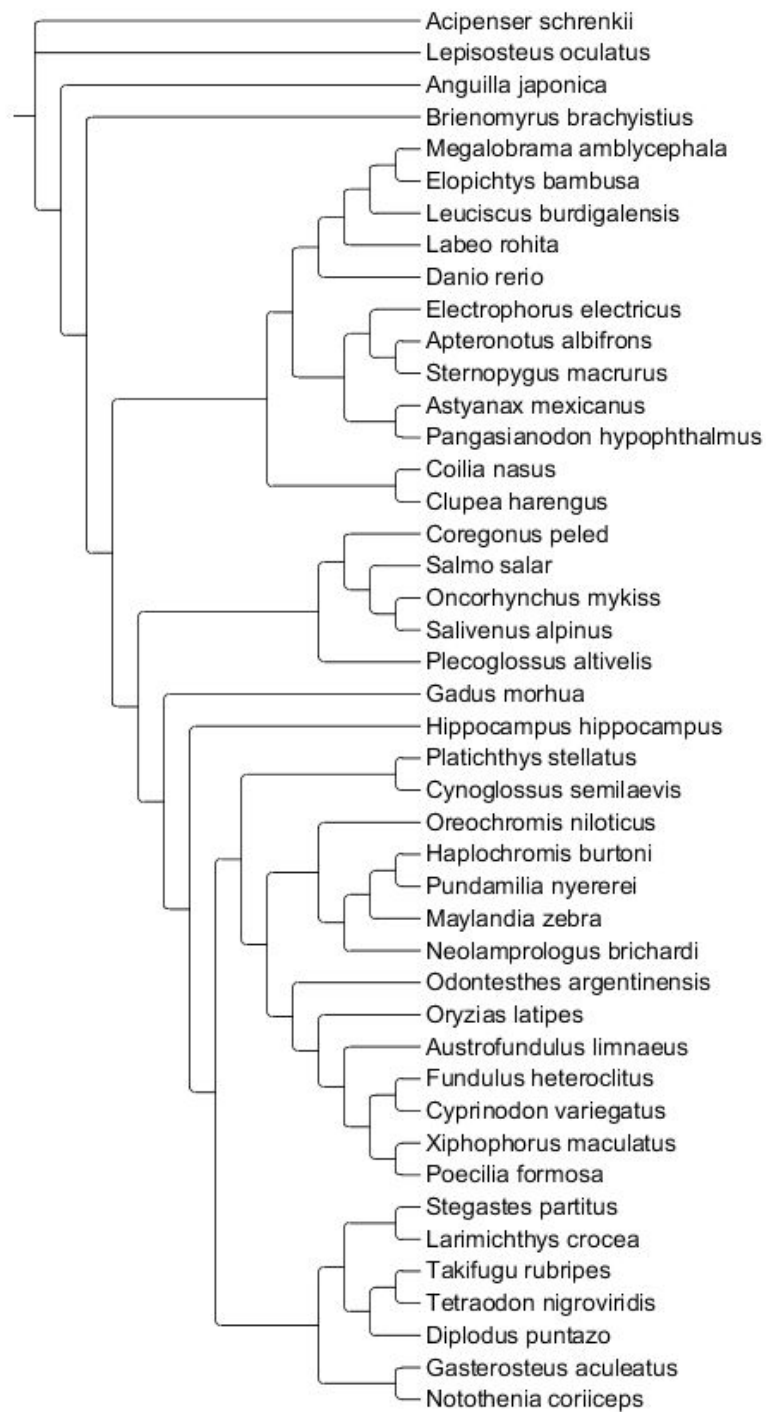


**Fig. 10:** Árbol utilizado en los análisis de selección de MEME, BSR y CodeML para el Grupo A (Genes AQP10b, CLCN-3, SLC9A1, SLC9A2 y SLC12A2a. Los círculos numerados indican las ramas foreground en los modelos de evaluados en CodeML para los genes: 1) CLCN-3, 2) SLC9A2 y 3) SLC12A2a.

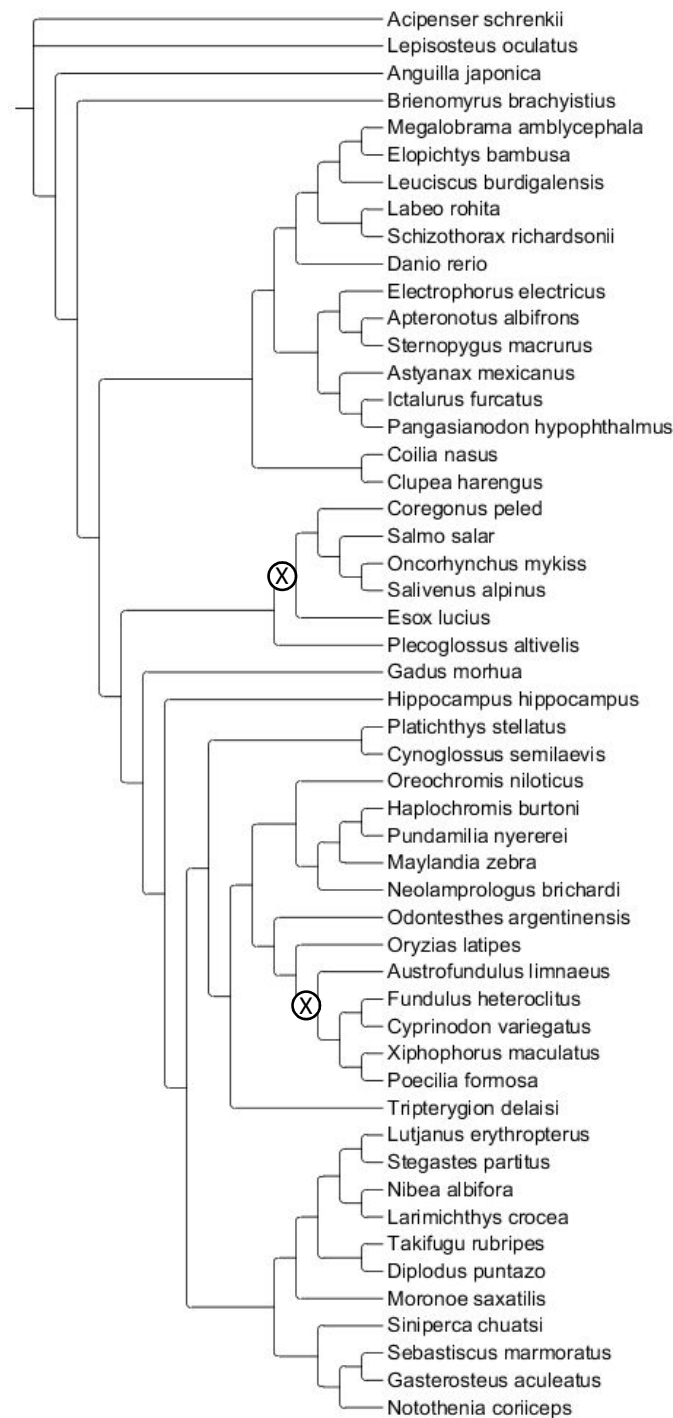


**Fig. 11:** Árbol utilizado en los análisis de selección de MEME, BSR y CodeML para el Grupo B (Genes SLC12A3, AQP-12 y AQP-9b). Los círculos numerados indican las ramas foreground en los modelos de evaluados en CodeML para el gen 1) SLC12A3

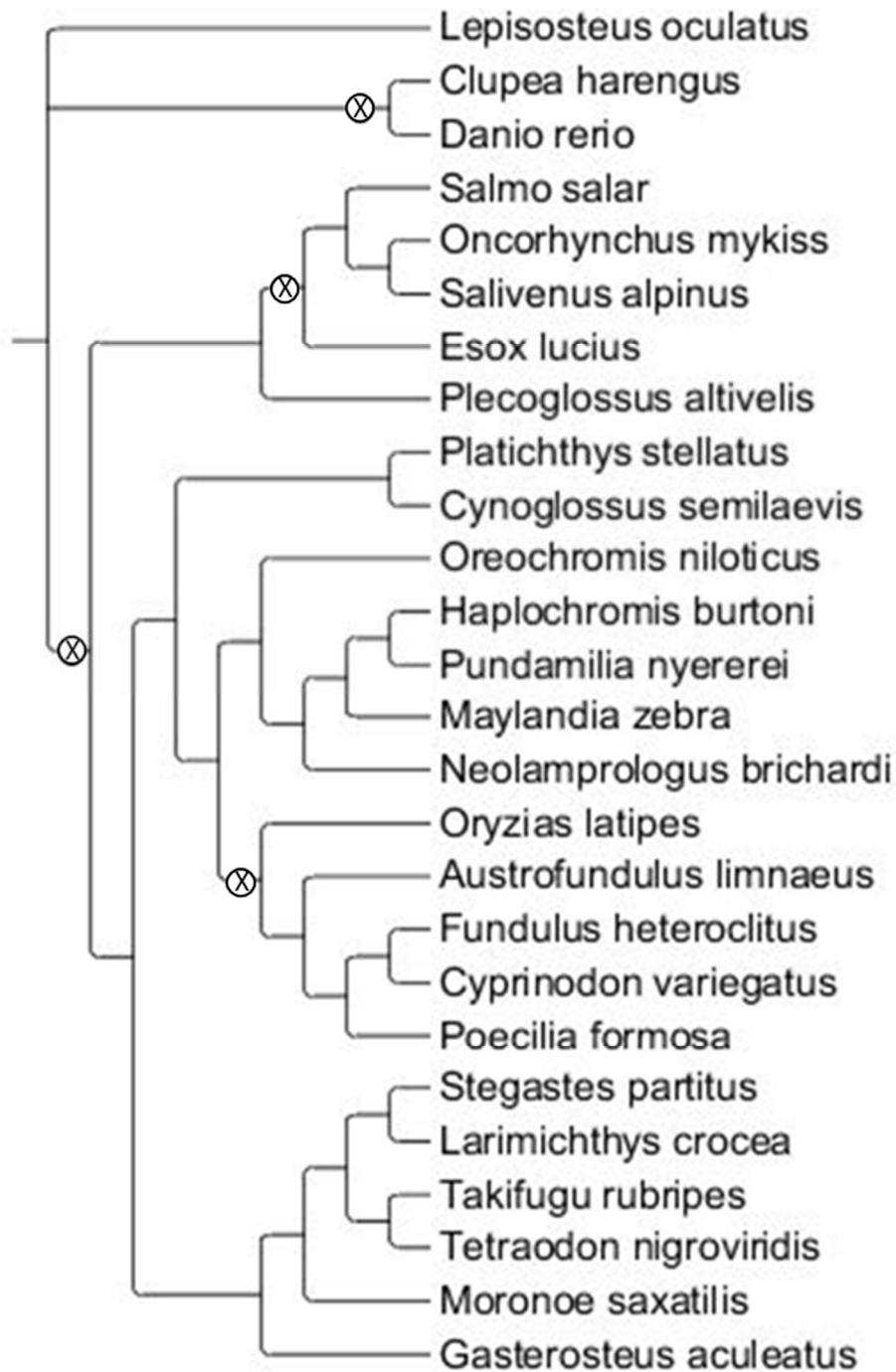




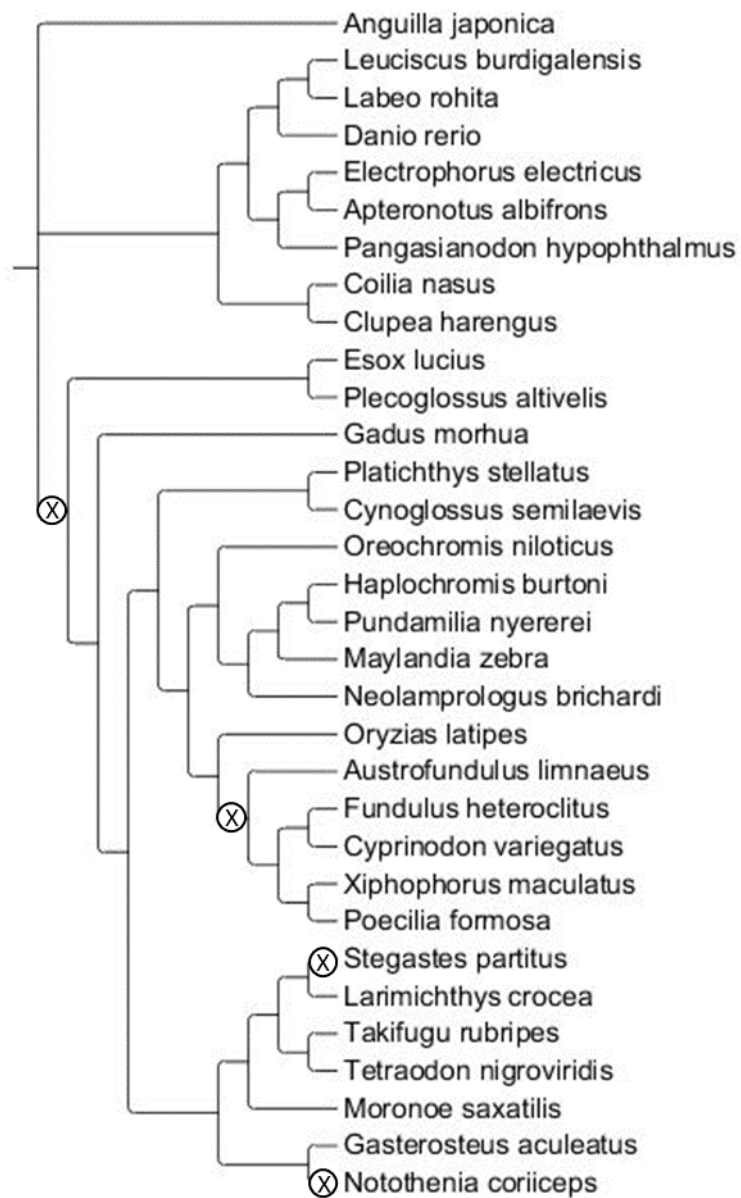
**Fig. 12:** Árbol utilizado en los análisis de selección de MEME, BSR y CodeML para el gen VDAC-1.



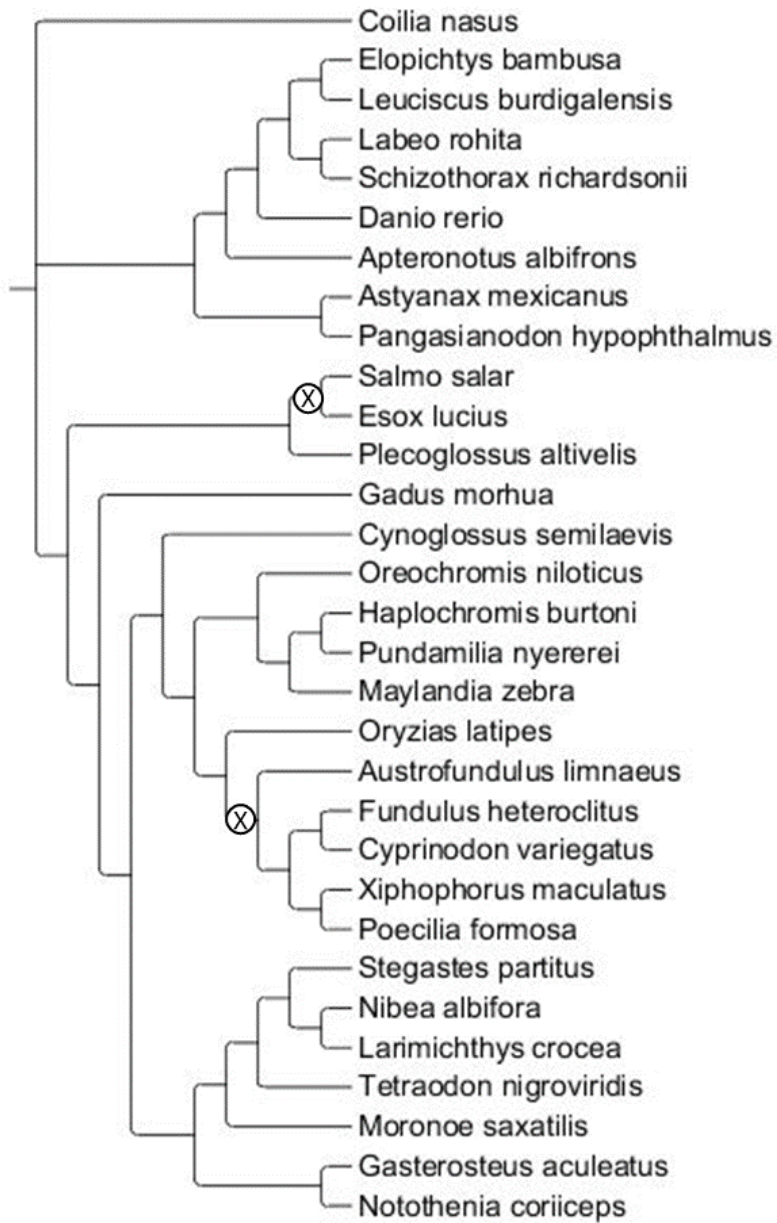
**Fig. 13:** Árbol utilizado en los análisis de selección de MEME, BSR y CodeML para el gen VDAC-2. Los círculos indican las ramas foreground en los modelos de evaluados en CodeML.



**Fig. 14:** Árbol utilizado en los análisis de selección de MEME, BSR y CodeML para el gen CFTR. Los círculos indican las ramas foreground en los modelos de evaluados en CodeML.



**Fig. 15:** Árbol utilizado en los análisis de selección de MEME, BSR y CodeML para el gen SLC12A1. Los círculos indican las ramas foreground en los modelos de evaluados en CodeML.



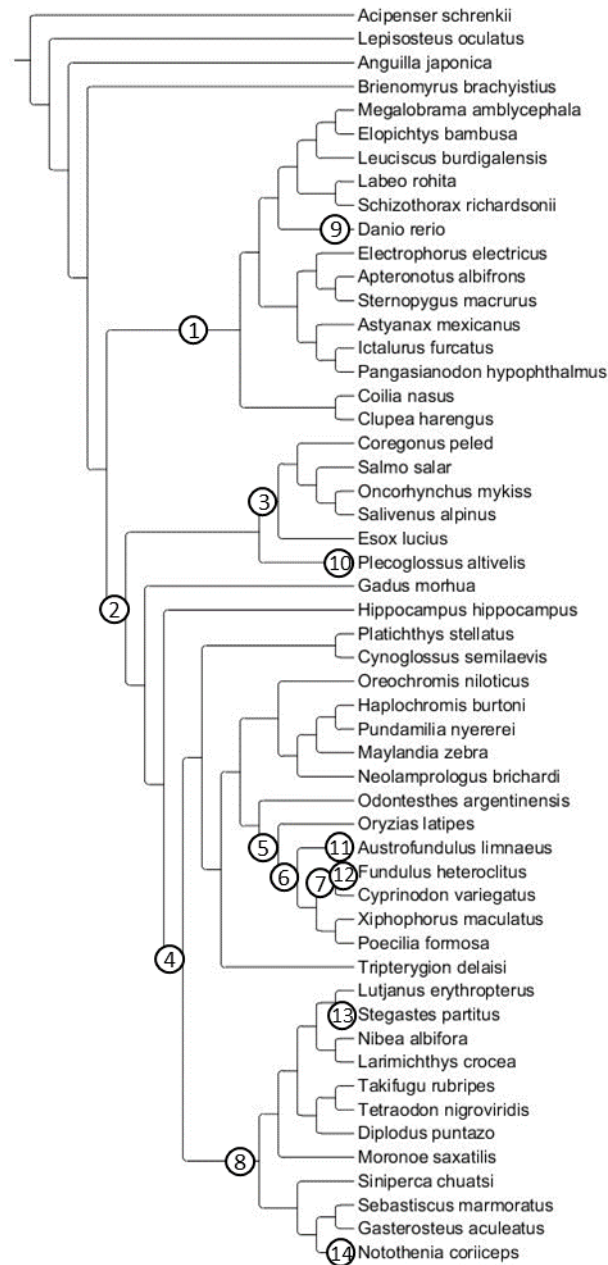
**Fig. 16:** Árbol utilizado en los análisis de selección de MEME, BSR y CodeML para el gen SLC26A6. Los círculos indican las ramas foreground de los modelos de evaluados en CodeML.

### 3.3.1 Selección a nivel de linajes

Los resultados de BSR se detallan en la Figura 17 (Tabla 3). La mayoría de los genes coincidieron en linajes: Atherinomorphae (CFTR y SLC9A1) Cyprinodontiformes (SLC12A1, SLC12A2a y VDAC-2,) Esociformes + Salmoniformes, (CFTR, SLC9A1 y VDAC-2), Euteleosteomorpha (CFTR, CLCN-3, SLC9A1, SLC12A1, SLC12A2a y SLC26A6) y Otomorpha (CFTR, SLC12A2a y SLC12A3). Consultar el Anexo D para los árboles resultantes de cada gen.

No se puede determinar con seguridad la posición en el árbol del Nodo 10 de SLC12A3 y el Nodo 40 de SLC9A1 debido a que la falta de datos genera grandes discrepancias sobre la identidad del linaje bajo selección. Por ejemplo, no puede asegurarse si el definido aquí como el linaje Nodo 10 del gen SLC12A3 corresponde a los grupos definidos por Betancur-R (2013) Neoteleostei, Acanthomorpha o Euacanthomorpha (entre otros) al no disponerse de secuencias para los órdenes Gadiformes y/o Syngnathiformes (especies *G. morhua* y *Hippocampus hippocampus*). En esta tesis se referirá a estos grupos por el número de nodo del árbol utilizado en su análisis.

El logaritmo de verosimilitud para los modelos de Ramas de CodeML realizados se reporta en la Tabla 4. Los LRT MA vs MCon y MB vs MCon arrojaron resultados significativos para todos los genes analizados, mientras que el LRT MB vs MA fue significativo para los genes CFTR, SLC9A1 y SLC12A2 (Tabla 4). Los modelos Ramas-Sitios, por su parte, arrojaron resultados marginalmente significativos para el linaje Otomorpha del gen SLC12A3 y Esociformes + Salmoniformes de VDAC-2; y resultados no significativos para los linajes Esociformes + Salmoniformes, Atherinomorphae, Fundulidae + Cyprinodontidae y 40 para el gen SLC9A1 (Tabla 5), y el linaje Cyprinodontiformes para SLC12A1.



**Fig. 17:** Linajes encontrados bajo selección por BSR. Los círculos numerados corresponden a los grupos: 1) Otomorpha (CFTR, SLC12A2a y SLC12A3), 2) Euteleosteomorpha (CFTR, CLCN-3, SLC9A1\*, SLC12A1, SLC12A2a y SLC26A6), 3) Escociformes + Salmoniformes (CFTR, SLC9A1 y VDAC2), 4) Nodo 10 (SLC12A3), 5) Atherinomorphae (CFTR y SLC9A1\*), 6) Cyprinodontiformes (SLC12A1\*, SLC12A2a y VDAC-2), 7) Fundulidae + Cyprinodontidae (SLC9A1\*), 8) Nodo 40 (SLC9A1\*), 9) *D. rerio* (SLC26A6), 10) *P. altivelis* (CLCN-3\*), 11) *A. limnaeus* (SLC9A1), 12) *F. heteroclitus* (SLC9A1), 13) *S. partitus* (SLC12A1), 14) *N. coriiceps* (SLC12A1). Los genes marcados con asteriscos indican que la señal de selección en ese gen no fue recuperada para ese linaje en el modelo de Ramas-Sitios correspondiente.

**Tabla 3:** Resultados de BSR. Se reporta el p-valor corregido para múltiples pruebas por el método de Holm-Bonferroni (Cor. p-valor), el valor de  $\omega$  para los sitios bajo selección positiva  $\omega+$  y la proporción bajo selección (Pr [ $\omega=\omega+$ ]). Abreviaciones: linaje ancestral Esociformes y Salmoniformes (Esoci + Salmo) Fundulidae y Cyprinodontidae (Fundu + Cyprinodon)

<b>Gen/Linaje</b>	<b>Cor. p-value</b>	<b><math>\omega+</math></b>	<b>Pr [<math>\omega=\omega+</math>]</b>
<b><u>CFTR</u></b>			
Atherinomorphae	0,049	29,09	0,01
Esociformes y Salmoniformes	0,020	23,38	0,01
Euteleosteomorpha	0,004	229,51	0,02
Otomorpha	0,001	141,33	0,03
<b><u>CLCN-3</u></b>			
Euteleosteomorpha	0,029	52,07	0,01
<i>P. altivelis</i>	<0,001	28,87	0,11
<b><u>SLC9A1</u></b>			
<i>A. limnaeus</i>	0,041	513,10	0,02
Euteleosteomorpha	0,047	10000	0,01
Esociformes y Salmoniformes	0,008	74,44	0,02
Atherinomorphae	0,006	25,27	0,03
<i>F. heteroclitus</i>	0,018	1465,75	0,04
Fundulidae y Cyprinodontidae	0,049	10000	0,03
Nodo 40	0,039	410,88	0,02
<b><u>SLC12A1</u></b>			
Cyprinodontiformes	0,049	10,45	0,03
Euteleosteomorpha	<0,0001	10000	0,02
<i>N. coriiceps</i>	0,009	10,56	0,03
<i>S. partitus</i>	0,007	3112,23	0,02
<b><u>SLC12A2</u></b>			
Cyprinodontiformes	0,002	30,66	0,01
Euteleosteomorpha	0,011	137,37	0,01
Otomorpha	<0,0001	253,12	0,03
<b><u>SLC12A3</u></b>			
Nodo10	0,049	29,93	0,05
Otomorpha	0,013	959,59	0,03
<b><u>SLC26A6</u></b>			
<i>D. rerio</i>	0,024	9,78	0,03
Euteleosteomorpha	<0,0001	137,20	0,05
<b><u>VDAC-2</u></b>			
Cyprinodontiformes	0,042	10000	0,02
Esociformes y Salmoniformes	0,032	12,63	0,03



**Tabla 4:** Resultados de los LRT para los modelos de Ramas y sus respectivos grados de libertad (d.f.). Los LRT realizados son MA vs MCon, MB vs MCon y MB vs MA. Umbral de significancia de 3,84 ( $p\text{-valor} \leq 0,05$ )

<b>Gen</b>	<b>MA vs MCon</b>	<b>d.f.</b>	<b>p-valor</b>	<b>MB vs MCon</b>	<b>d.f.</b>	<b>p-valor</b>	<b>MB vs MA</b>	<b>d.f.</b>	<b>p-valor</b>
CFTR	71,58	1	<0,05	128,59	4	<0,05	57,01	3	<0,05
CLCN-3	187,95	1	<0,05	188,34	2	<0,05	0,38	1	>0,05
SLC9A1	259,70	1	<0,05	287,67	7	<0,05	27,97	6	<0,05
SLC12A1	278,13	1	<0,05	282,42	4	<0,05	4,29	3	>0,05
SLC12A2a	11,07	1	<0,05	22,02	3	<0,05	10,95	2	<0,05
SLC12A3	113,75	1	<0,05	114,29	2	<0,05	0,54	1	1,00
SLC26A6	95,66	1	<0,05	95,94	2	<0,05	0,28	1	1,00
VDAC-2	13,59	1	<0,05	15,88	2	<0,05	2,29	1	>0,05

**Tabla 5:** Resultados de los LRT para los modelos de Ramas-Sitios, un grado de libertad. Se reporta el logaritmo de la verosimilitud para los modelos Control (LnV Con) y alternativo (LnV A), el LRT comparando ambos modelos y su p-valor. Umbral de significancia de 3,84 (p-valor  $\leq$  0,05).

<b>Gen/Linaje</b>	<b>LnV Con</b>	<b>LnV A</b>	<b>LRT</b>	<b>p-valor</b>
<b><u>CFTR</u></b>				
Atherinomorphae	-41303,28	-41299,58	7,38	<0,05
Esociformes y Salmoniformes	-41320,88	-41302,41	21,25	<0,05
Euteleosteomorpha	-41313,17	-41310,26	21,52	<0,05
Otomorpha	-41295,99	-41291,67	8,64	<0,05
<b><u>CLCN-3</u></b>				
Euteleosteomorpha	-16139,38	-16135,80	7,16	<0,05
<i>P. altivelis</i>	-16096,80	-16096,80	0	1
<b><u>SLC9A1</u></b>				
<i>A. limnaeus</i>	-15447,80	-14326,65	2242	<0,05
Atherinomorphae	-15447,80	-15447,80	0	1,00
Esociformes y Salmoniformes	-15440,37	-15430,81	19,11	<0,05
Euteleosteomorpha	-15447,70	-15447,70	0	1,00
<i>F. heteroclitus</i>	-15446,08	-14321,64	2248	<0,05
Fundulidae y Cyprinodontidae	-15441,11	-15437,32	7,59	<0,05
Nodo 40	-15447,80	-15447,81	0,03	1,00
<b><u>SLC12A1</u></b>				
Cyprinodontiformes	-39131,82	-39131,82	0	1
Euteleosteomorpha	-39127,18	-39116,45	21,47	<0,05
<i>N. coriiceps</i>	-39122,72	-39110,39	24,68	<0,05
<i>S. partitus</i>	-39113,45	-39119,58	12,27	<0,05
<b><u>SLC12A2</u></b>				
Cyprinodontiformes	-27792,93	-27789,46	6,94	<0,05
Euteleosteomorpha	-27792,91	-27790,15	5,51	<0,05
Otomorpha	-27766,64	-27755,12	23,04	<0,05
<b><u>SLC12A3</u></b>				
Nodo 10	-20535,48	-20533,13	4,70	<0,05
Otomorpha	-20542,68	-20541,13	3,09	$\approx$ 0,05
<b><u>SLC26A6</u></b>				
<i>D. rerio</i>	-26616,56	-26611,50	10,10	<0,05
Euteleosteomorpha	-26616,41	-26609,89	13,03	<0,05
<b><u>VDAC-2</u></b>				
Cyprinodontiformes	-12738,22	-12731,80	4,70	<0,05
Esociformes y Salmoniformes	-12739,81	-12738,09	3,44	$\approx$ 0,05

### 3.3.2 Selección a nivel de sitios

El logaritmo de la verosimilitud de los modelos de Sitios se reportan en la Tabla E2. El LRT M0 vs M3 fue significativo para todos los genes, contrario al LRT M1a vs M2a que solo dio un resultado significativo para SLC12A2 (Tabla 6). Por su lado los LRT M7 vs M8 y M8a vs M8 arrojaron ambos resultados significativos para los genes CLCN-3, SLC9A1, SLC9A2, SLC12A1, SLC12A2, VDAC-1 y, con un p-valor marginal (p-valor  $\approx$ 0,05) del LRT M8a vs M8, SLC12A3 (Tabla 6). El LRT M7 vs M8 arrojó resultados significativos para los

**Tabla 6:** Resultados de los LRT M0 vs M3, M1a vs M2a, M7 vs M8 con un umbral de significancia de 3,84 ( $p\text{-valor} \leq 0,05$ ), y el LRT M8a vs M8 con un umbral de 2,71 ( $p\text{-valor} \leq 0,05$ ). Los grados de libertad son 4, 2, 2 y 1 respectivamente.

<b>Gen</b>	<b>M0 vs M3</b>	<b>p-valor</b>	<b>M1a vs M2a</b>	<b>p-valor</b>	<b>M7 vs M8</b>	<b>p-valor</b>	<b>M8a vs M8</b>	<b>p-valor</b>
AQP-9b	374,84	<0,05	<0,01	1,00	3,95	<0,05	1,30	>0,05
AQP-10b	498,51	<0,05	<0,01	1,00	-6,54	X	-9,99	X
AQP-12	442,78	<0,05	<0,01	1,00	4,54	<0,05	0,59	>0,05
CFTR	1490,80	<0,05	<0,01	1,00	6,65	<0,05	<0,01	1,00
CLCN-3	646,19	<0,05	<0,01	1,00	17,58	<0,05	20,41	<0,05
SLC9A1	1757,37	<0,05	<0,01	1,00	23,55	<0,05	18,06	<0,05
SLC9A2	719,67	<0,05	<0,01	1,00	7,84	<0,05	3,36	<0,05
SLC12A1	3221,29	<0,05	<0,01	1,00	46,23	<0,05	15,81	<0,05
SLC12A2	1981,72	<0,05	13,65	<0,05	97,65	<0,05	14,68	<0,05
SLC12A3	1642,63	<0,05	<0,01	1,00	19,43	<0,05	2,40	≈0,05
SLC26A6	2003,25	<0,05	<0,01	1,00	28,00	<0,05	1,55	>0,05
VDAC-1	936,27	<0,05	<0,01	1,00	24,01	<0,05	9,78	<0,05
VDAC-2	442,78	<0,05	<0,01	1,00	<0,01	1,00	13,20	<0,05

genes AQP-9b, AQP-12, CFTR y SLC26A6 y el LRT M8a vs M8 para el gen VDAC-2. Por ello solo se consideraron los codones propuestos bajo selección positiva por el modelo de sitios M8 (Tabla 7).

**Tabla 7:** Codones identificados bajo el modelo de sitios M8 para los genes AQP-9b, SLC9A1, SLC12A1, SLC12A2, SLC12A3 y SLC26A6 por BEB. Se reporta la estimación de  $\omega$ , su varianza y la probabilidad del codón de estar bajo selección positiva ( $\Pr(\omega > 1)$ ). Los codones identificados por otros métodos se resaltan en negrita.

Gen	Codón	$\omega$	Varianza	$\Pr(\omega > 1)$
AQP-9b	134	2,25	1,76	0,54
SLC9A1	<b>137</b>	1,00	0,58	0,57
	388	2,00	2,00	0,88
SLC12A1	84	1,00	0,26	0,93
	355	2,00	0,24	0,96
	362	1,00	0,26	0,93
	<b>363</b>	2,00	0,24	0,96
	<b>698</b>	1,00	0,37	0,65
	760	1,42	0,32	0,85
SLC12A2	19	1,87	0,91	0,63
	279	2,57	0,54	1,00
	345	2,57	0,54	1,00
	432	2,27	0,80	0,84
	442	1,62	0,90	0,50
	<b>461</b>	2,54	0,59	0,97
SLC12A3	258	1,45	0,59	0,83
	263	1,23	0,42	0,59
	620	1,36	0,38	0,76
SLC26A6	<b>35</b>	1,97	1,00	1,00
	162	1,69	1,00	0,78
	164	1,44	1,00	0,62
	229	1,95	1,00	0,98
	<b>314</b>	1,80	1,00	0,84

A continuación se presentan gen a gen los codones detectados bajo selección positiva por MEME y los modelos de CodeML de Sitios (modelo M8) y Ramas-Sitios con un LRT significativo. El  $\omega$  de las clases de sitios de los modelos de Ramas-Sitios se reporta en la Tabla E3. Para los codones detectados por estos métodos se incluyen los resultados de PFAM y de la búsqueda en Uniprot.

#### 1) AQP-9:

El análisis del MEME identificó 2 sitios candidatos de estar bajo selección direccional, ninguno con un q-valor significativo (Tabla F1 y Figs. G1-G2), mientras que el modelo por Sitios de CodeML detectó un único sitio candidato de estar bajo selección (Tabla 7). No se presentaron coincidencias entre los modelos.

#### 2) AQP-10b:

El análisis del MEME identificó 3 sitios candidatos de estar bajo selección direccional, ninguno con un q-valor significativo (Tabla F2 y Figs. G3-G5). Los LRT M7 vs M8 y M8a vs M8 arrojaron valores negativos muy lejanos a 0 (Tabla 6), indicando problemas en la convergencia de los modelos. Aunque el problema parece solucionarse parcialmente cambiando el modelo de frecuencias de codones por uno más simple (ej. F3X4) se optó por descartar este análisis.

#### 3) AQP-12:

El análisis del MEME identificó 1 sitio candidato de estar bajo selección direccional, ninguno con un q-valor significativo (Tabla F3; Fig. G6), mientras que el modelo de Sitios de CodeML no identificó sitios bajo selección.

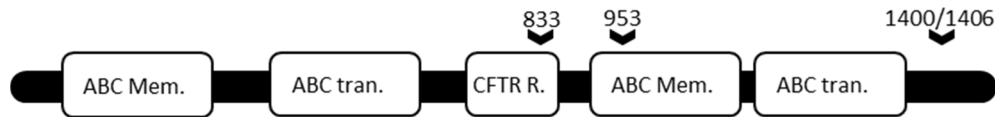
#### 4) CFTR:

El análisis del MEME identificó 24 sitios candidatos de estar bajo selección direccional, ninguno con un q-valor significativo (Tabla F4 Figs. G7-G28). Por su lado los modelos de sitios para los linajes grupo ancestral Esociformes + Salmoniformes (Tabla E4) Euteleostomorpha (Tabla E5), Atherinomorphae (Tabla E6), identificaron 6 (2 con p-valor significativo), 18 (7 con un p-valor significativo) y 33 (3 significativos) respectivamente. El modelo por Sitios de CodeML, por su lado, no identificó codones candidatos de estar bajo selección.

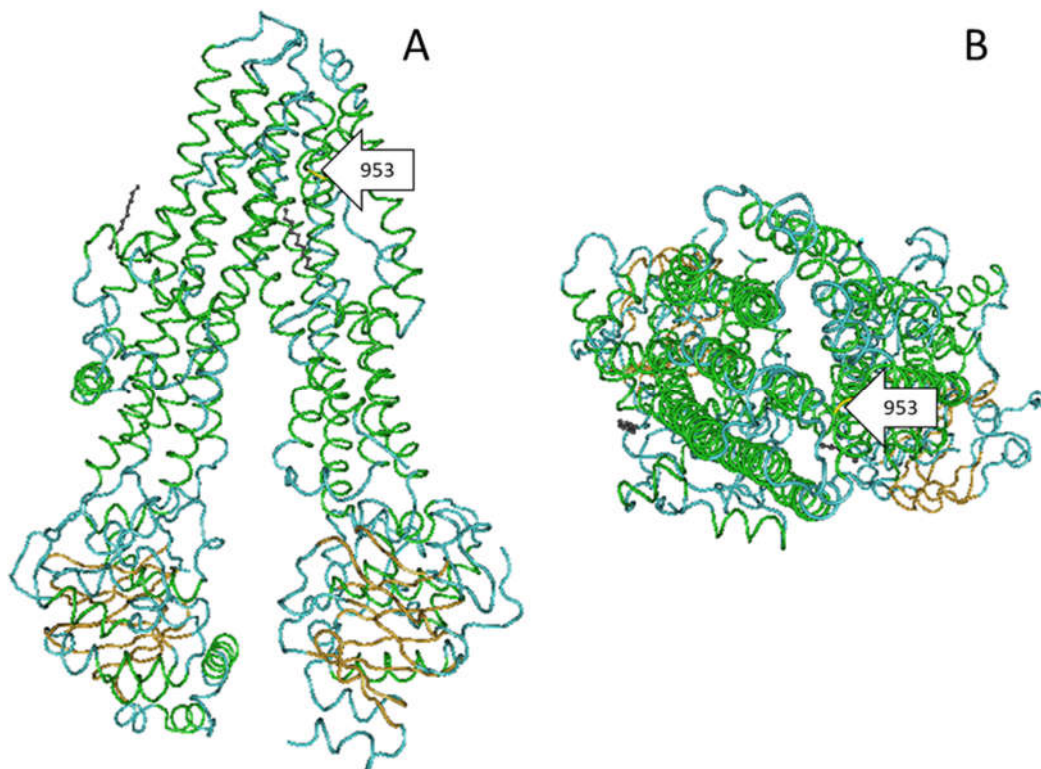
El modelo de Ramas-Sitios para el linaje Otomorpha no fue considerado a la hora de buscar sitios candidatos de estar bajo selección debido a anomalías en los codones identificados. Este modelo identificó un total de 124 sitios bajo selección (datos no mostrados), 9 sitios compartidos con los otros modelos de Ramas-Sitios (1 con Esociformes + Salmoniformes, 4 con Euteleostomorpha y otros 4 con Atherinomorphae). Si bien los codones identificados por más de un modelo de Ramas-Sitios podrían representar selección sobre los mismos sitios, el alto número total de sitios candidatos de estar bajo selección positiva sugiere que se trata de un artefacto.

El codón 1406 fue identificado tanto por MEME como el modelo de Ramas-Sitios Esociformes + Salmoniformes, así como los codones 833, 953 y 1400 con el Modelo de Ramas-Sitios para el linaje Atherinomorphae. De acuerdo con PFAM, el codón 833 se encuentra en un dominio "Regulador de la Conductancia de la Fibrosis Quística" y 953 en una región "Transporte Transmembrana ABC". El codón 833 no tiene un homólogo en la isoforma canónica de CFTR consultada (P13569-1), mientras que 953 se encuentra en una hélice transmembrana (denominada N°9) y los codones 1400 y 1406 en un dominio

topológico citoplasmático. El codón 1400 (homólogo al sitio 1456 en la isoforma P13569-1) es una Serina fosforilada en humanos (McClure et al. 2012). La posición de estos codones en el gen se muestra en la Figura 18. Solo el codón 953 está incluido en el modelo de la estructura terciaria de la proteína propuesto por Zhang & Chen (2016) (Fig. 19).



**Fig. 18:** Posición relativa de los codones identificados por al menos dos métodos de detección de evidencia de selección positiva en el gen CFTR. También se representan los Dominios identificados por PFAM: dos Regiones Transportador ABC transmembrana (ABC Mem.), dos Transportadores ABC (ABC tran.) y Regulador Transmembrana de la Conductasa de la Fibrosis Quística (CFTR-R.).

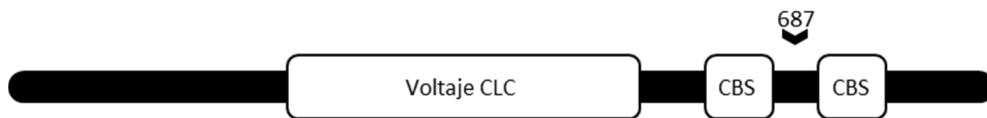


**Fig. 19:** Estructura tridimensional del canal CFTR (ID: 4BUM) A) visto lateralmente, la membrana mitocondrial interior debajo de la imagen, y B) Vista apical. Regiones plegadas formando  $\alpha$  hélices se muestran en verde, láminas  $\beta$  en dorado y loops en celeste. El codón 953 se muestra en amarillo y señalado con una flecha.

#### 5) CLCN-3:

El análisis del MEME identificó 13 codones candidatos de estar bajo selección direccional, de los cuales el sitio 687 presentó un q-valor significativo (Tabla F5 y Figs. G29-G39), mientras que el Modelo Ramas-Sitios para el linaje Euteleostomorpha (Tabla E7) identificó 6 (2 con un p-valor significativo), y el modelo por Sitios de CodeML no identificó ninguno.

El sitio 687 fue identificado tanto por MEME como por Ramas-Sitios. Sin embargo no se observan cambios no sinónimos en esta posición. Aparentemente, la señal detectada es un artefacto causado por una violación de las asunciones del modelo de sustitución de codones utilizados por ambos análisis (Kosakovsky et al. 2011; Murrell et al. 2012; Nielsen & Yang 1998). Las especies *L. oculatus*, *D. rerio*, *A. japonica*, *A. mexicanus* y *P. hypophthalmus* tienen el codón AGC en esta posición mientras que el resto TCC, ambos codificantes para Serina. La posición de este codón en el gen se muestra en la Figura 20.

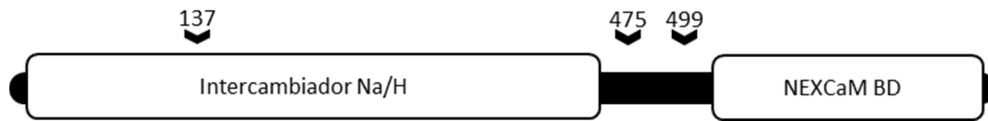


**Fig. 20:** Posición relativa del codón 687 los codones identificados por al menos dos métodos de detección de evidencia de selección positiva en el gen CLCN-3. También se representan los Dominios identificados por PFAM: Canal de Cloro abierto por Voltaje (Voltage CLC).

#### 6) SLC9A1:

El análisis del MEME identificó 12 codones candidatos de estar bajo selección direccional, ninguno con un q-valor significativo (Tabla F6 y Figs. G40-G51), mientras que los Modelos Ramas-Sitios para los linajes el grupo ancestral Esociformes + Salmoniformes (Tabla E8) y *F. heteroclitus* (Tabla E9) identificaron 7 (dos con p-valores significativos) y 2 (uno significativo) respectivamente. El Modelo por Sitios de CodeML por su parte identificó dos con p-valores no significativos (Tabla 7).

El codón 499 fue identificado por MEME y por el Modelo Ramas-Sitios para el linaje con el grupo Esociformes + Salmoniformes, así como el 475 aunque con un p-valor no significativo. El Modelo de Sitios y MEME identificaron ambos el sitio 137. De acuerdo con PFAM, el codón 137 estaría comprendido dentro una región de "Intercambiador Sodio/hidrógeno". No hay datos para este gen en Uniprot para humanos, pero el sitio homólogo en *M. musculus* el codón 137 pertenece a un dominio topológico extracelular y los codones 475 y 499 al mismo dominio topológico citoplasmático. La posición de estos codones en el gen se muestra en la Figura 21.



**Fig. 21:** Posición relativa de los codones identificados por al menos dos métodos de detección de evidencia de selección positiva en el gen SLC9A1. También se representan los Dominios identificados por PFAM: Familia intercambiador  $\text{Na}^+/\text{H}^+$  (Intercambiador Na/H) y Región reguladora del intercambiador de  $\text{Na}^+/\text{H}^+$  NHE, unión a calmodulina (NEXCaM BD)

Los modelos de Ramas-Sitios para el linaje ancestral de *A. limnaeus* y el del linaje ancestral de Fundulidae y Cyprinodontidae presentaron resultados anómalos. El primero reportó un  $\omega$  para las categorías de sitios bajo selección (2a y 2b) igual a 1 (Tabla E3), con una probabilidad de que un sitio perteneciera a una de ellas de 0. En el segundo por su parte, el método BEB identificó que la gran mayoría los codones como candidatos de estar bajo selección, con una  $\text{Pr}(\omega > 1) \approx 0,57$  (datos no mostrados). Curiosamente, esto no afectó la inferencia de codones bajo selección mediante Naive Empirical Bayes (NEB), el método para calcular las probabilidades a posteriori de los sitios candidatos bajo selección en las primeras versiones de PAML (Nielsen & Yang 1998; Yang 2000), pero que luego fue abandonado en favor de BEB (Yang et al. 2005). No está claro el porqué de estos resultados; presuntamente es debido a problemas de convergencia del modelo utilizado; por los sitios identificados por ambos modelos no fueron considerados.

#### 7) SLC9A2:

El análisis del MEME identificó 10 codones candidatos de estar bajo selección direccional, ninguno con un q-valor significativo (Tabla F7 y Figs. G52-G61). El modelo de Sitios de CodeML no identificó ningún sitio bajo selección.

#### 8) SLC12A1:

El análisis del MEME identificó 30 codones candidatos de estar bajo selección direccional, ninguno con un q-valor significativo (Tabla F8 y Figs. G62-G91), mientras que los modelos Ramas-Sitios para los linajes Euteleostomorpha (Tabla E10), *N. coriiceps* (Tabla E11) y *S. partitus* (Tabla E12) identificaron 11 codones (2 con p-valores significativos), 17 (6 con p-valores significativos) y 7 (6 con p-valores significativos). Por su lado el Modelo por Sitios de CodeML identificó 3 con p-valores no significativos (Tabla 7).



El codón 787 fue encontrado bajo selección positiva por MEME, y los modelos de Ramas-Sitios para los linajes Euteleosteomorpha y *S. partitus*. Otros codones identificados por MEME y uno de los modelos de CodeML son: el codón 131 fue identificados tanto con el Modelo Ramas-Sitios Euteleosteomorpha, los codones 343, 381 y 840 con el Modelo Ramas-Sitios *N. coriiceps*, el codón 391 con Modelo Ramas-Sitios *S. partitus*, y los codones 363 y 698 con el Modelo de Sitios de CodeML. De acuerdo con PFAM, los codones 131, 343, 363, 381 y 391 se encuentran en una región “Permeasa a Aminoácidos”, mientras que los codones 698, 787 y 840 en una región Transportadora de Solutos 12”. La búsqueda en Uniprot indica que el codón 363 (homólogo al sitio 454 en la isoforma Q13621-1) se encuentra a 2 codones de distancia de un posible sitio de glicosilación ligada a nitrógeno (Codón 456). Adicionalmente, Uniprot indica que los codones 131 y 343 se encuentran en hélices transmembrana diferentes y que los codones 787 y 840 en un dominio topológico citoplasmático. La posición de estos codones en el gen se muestra en la Figura 22.



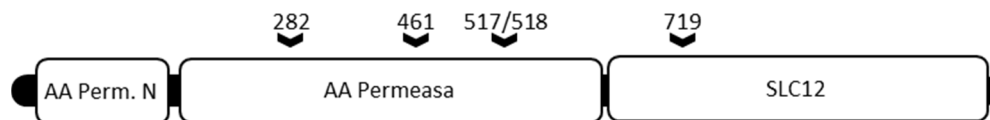
**Fig. 22:** Posición relativa de los codones identificados por al menos dos métodos de detección de evidencia de selección positiva en el gen SLC12A1. También se representan los Dominios identificados por PFAM: Permeasa AA N terminal (AA Perm. N), Permeasa a Aminoácidos (AA Permeasa) y Carriers de solutos, familia 12 (SLC12).

#### 9) SLC12A2a:

El análisis del MEME identificó 17 codones candidatos de estar bajo selección direccional, ninguno con un q-valor significativo (Tabla F9 y Figs. G92-G108), mientras que los Modelos Ramas-Sitios para el linaje Cyprinodontiformes (Tabla E13) Euteleosteomorpha (Tabla E14) y Otomorpha (Tabla E15) identificaron 11 (1 con p-valor significativo), 10 (1 con p-valor significativo) y 36 (8 con p-valor significativo) codones candidatos de estar bajo selección respectivamente. Por su parte, el modelo por Sitios de CodeML identificó 6 sitios (3 con un p-valor significativo) (Tabla 7).

Los codones 282, 518 y 719 fueron identificados por MEME y el modelo de Ramas-Sitios del linaje Otomorpha y el 517 con Cyprinodontiformes. Mientras que el codón 461 fue identificado por MEME y el Modelo de sitios. De los 36 sitios identificados por el Modelo de Ramas-Sitios para Otomorpha, 3 fueron identificados por el Modelo Ramas-Sitios Euteleosteomorpha y 1 para el Cyprinodontiformes; aunque en ningún caso los sitios fueron significativos simultáneamente en ambos modelos. De acuerdo con PFAM,

los codones 282, 461, 517 y 518 se encuentran en una región “Permeasa a Aminoácidos” mientras que el 719 se encuentra en una región “Transportadora de Solutos 12”. Uniprot indica que el codón 282 se encuentra en una hélice transmembrana, el 461 en un dominio topológico extracelular, y los codones 517, 518 en uno citoplasmático. Por último el codón 719 en dominios topológicos citoplasmáticos diferente. La posición de estos codones en el gen se muestra en la Figura 23.



**Fig. 23:** Posición relativa de los codones identificados por al menos dos métodos de detección de evidencia de selección positiva en el gen SLC12A2a. También se representan los Dominios identificados por PFAM: Permeasa AA N terminal (AA Perm. N), Permeasa a Aminoácidos (AA Permeasa) y Carriers de solutos, familia 12 (SLC12).

#### 10) SLC12A3:

El análisis del MEME identificó 9 codones candidatos de estar bajo selección direccional, ninguno con un q-valor significativo (Tabla F10 y Figs. G109-G117), mientras que modelo de Ramas-Sitios para el linaje Nodo 10 (Tabla E16) y Otomorpha (Tabla E17) identificaron 7 (dos con p-valores significativos) y 3 (ninguno con p-valor significativo) respectivamente. El modelo por Sitios de CodeML identificó 3 con p-valores no significativos (Tabla 7). No se detectaron codones identificados por estos 3 métodos.

#### 11) SLC26A6:

El análisis del MEME identificó 26 codones candidatos de estar bajo selección direccional, los codones 35, 308 y 314 con un q-valor significativo (Tabla F11 y Figs. G118-G143), mientras que los modelos de Ramas-Sitios para los linajes *D. rerio* (Tabla E18) y Euteleosteomorpha (Tabla E18) identificaron 6 (1 con p-valor significativo) y 5 codones (1 con p-valor significativo) respectivamente. Mientras que el Modelo por Sitios de CodeML identificó otros 5 (2 con p-valor significativo) (Tabla 7).

Los codones 35 y 314 fueron identificados tanto por MEME como por el Modelo de Sitios, y el codón 278 por el modelo de Ramas-Sitios para *D. rerio*. De acuerdo con PFAM, los codones 278, 308 y 314 se encuentran en una región “Permeable a Sulfato”. Uniprot indica que el codón 35 se encuentra en un dominio topológico citoplasmático, el 278 en uno extracelular y los codones 308 y 314 juntos en otro dominio topológico citoplasmático. La posición de estos codones en el gen se muestra en la Figura 24.



**Fig. 24:** Posición relativa de los codones identificados por al menos dos métodos de detección de evidencia de selección positiva en el gen SLC26A6, y el codón 308 debido a que presentó un q-valor significativo en MEME. También se representan los Dominios identificados por PFAM: Permeasa Sulfato y Transportador Sulfato y Anfagonista del factor AntiSigma (STAS).

#### 12) VDAC-1:

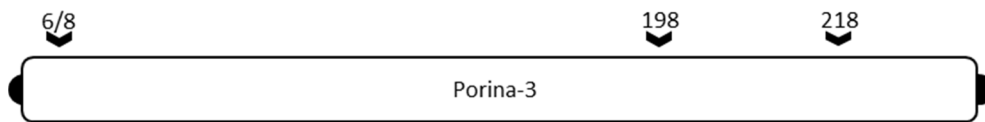
El análisis del MEME detectó un único sitio candidato bajo selección positiva con un q-valor no significativo (Tabla F12 y Fig. G144). No se detectaron codones bajo selección en el modelo de Sitios de CodeML.

#### 13) VDAC-2:

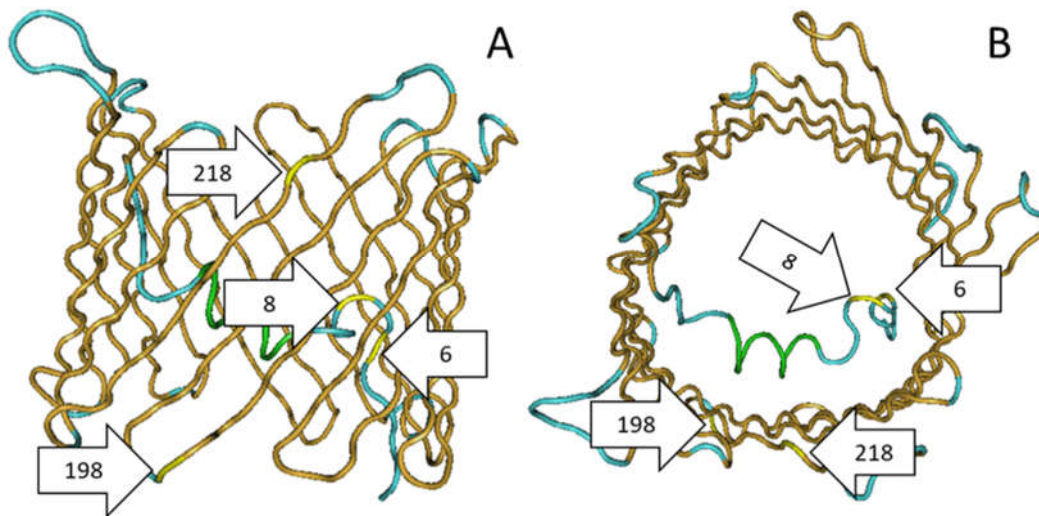
El análisis del MEME identificó 8 codones candidatos de estar bajo selección direccional, de las cuales las posiciones 6, 198 y 218 presentaron un q-valor significativo (Tabla F13 y Figs. G145-G152), mientras que los Modelos Ramas-Sitios de los linajes Esociformes + Salmoniformes (Tabla E20) y Cyprinodontiformes (Tabla E21) 3 codones candidatos de estar bajo selección (1 con p-valor significativo) y 4 (2 con p-valores significativo) sitios candidatos de estar bajo selección, respectivamente. El modelo de Sitios de CodeML no detectó sitios candidatos de estar bajo selección positiva.

El codón 198 fue reconocido conjuntamente por MEME y el modelo de Ramas-Sitios para el linaje el grupo Esociformes + Salmoniformes, así como el codón 8, aunque este último con un p-valor no significativo. El Codón 218 fue encontrado bajo selección tanto por MEME como por ambos modelos de Ramas-Sitios. PFAM indica que todos se encuentran dentro de una región "Porina-3" (Fig. 25), y la búsqueda de Uniprot indica que los sitios homólogos en humanos para los codones 198 y 218 se ubican en diferentes láminas beta transmembrana. Estas corresponden a las láminas  $\beta_{13}$  y  $\beta_{15}$  en el modelo estructural desarrollado por Schredelseker et al. (2014), ubicadas en la proximidad del extremo N-terminal (Fig. 26).

En la Tabla 8 se resumen los codones encontrados bajo selección positiva para cada gen, su sitio homólogo en *D. rerio* y humano, y el dominio proteico en el que se encuentran de acuerdo a PFAM.



**Fig. 25:** Posición relativa de los codones identificados por al menos dos métodos de detección de evidencia de selección positiva en el gen VDAC-2, y el codón 6 debido a que presentó un q-valor significativo en MEME. También se representa el dominio identificado por PFAM: Porina-3



**Fig. 26:** Estructura tridimensional del canal VDAC-2 (ID: 4BUM) A) visto lateralmente, la membrana mitocondrial interior debajo de la imagen, y B) Vista desde el Citoplasma celular. Regiones plegadas formando  $\alpha$  hélices se muestran en verde, láminas  $\beta$  en dorado y loops en celeste. Los codones identificados bajo selección se muestran en amarillo y señalados con flechas.

### 3.4 Discusión

Los análisis realizados encontraron consistentemente evidencia de selección positiva para 8 de los 13 genes estudiados en cuatro linajes de teleósteos: Cyprinodontiformes, Euteleostomorpha, Otomorpha, y el linaje ancestral de Salmoniformes y Esociformes. Con excepción de VDAC-2, todos participan en el movimiento de solutos a través de epitelios (ver la sección 3.1.1 Osmorregulación). No fue el caso con las acuaporinas incluidas en este análisis. Si bien los genes estudiados participan en varias funciones que no necesariamente están ligadas con la osmorregulación. Por ejemplo SLC12A2 en *D. rerio* participa en la regulación de la endolinfa de la vesícula ótica y la vejiga natatoria (Abbas & Whitfield 2009), y VDAC-2 en la regulación del ritmo cardíaco (Shimizu et al.

2015). El hecho de que sean varios los genes bajo selección positiva en los mismos linajes sugiere que es efectivamente su papel en la osmorregulación el que se encontró bajo selección, aunque no es necesariamente concluyente.

**Tabla 8:** Resumen de los sitios encontrados bajo selección positiva por al menos 2 métodos para los genes CFTR, SLC9A1, SLC12A1, SLC12A2, SLC26A6, y VDAC-2, sus sitios homólogos *D. rerio* sin editar, los sitios homólogos las secuencias de Uniprot consultadas, si el p-valor en el modelo de CodeML que lo detectó fue significativo, y las regiones proteicas donde se encuentran de acuerdo a PFAM. Aquí se utilizan las abreviaciones de los nombres de cada región utilizada en PFA: “Regulador de la Conductancia de la Fibrosis Quística” (CFTR-R), “Región Transporte Transmembrana ABC” (ABC membrane), “Intercambiador Sodio/hidrógeno” (Na H Exchanger), Permeasa a Aminoácidos (AA Permeasa), “Carriers de solutos, familia 12” (SLC 12), “Permeable a Sulfato” (Sulfato transp) y Porina-3

Gen	Codón	<i>D. rerio</i>	Uniprot	CodeML sign.	PFAM
CFTR	833	896	X	Si	CFTR-R
	953	1014	1004	Si	ABC membrane
	1400	1459	1455	No	-
	1406	1468	1461	Si	-
SLC9A1	137	106	224	No	Na H Exchanger
	475	442	562	No	-
	499	466	586	Si	-
SLC12A1	131	26	220	Si	AA Permeasa
	343	238	432	Si	AA Permeasa
	363	260	454	Si	AA Permeasa
	381	276	473	No	AA Permeasa
	391	289	483	No	AA Permeasa
	698	586	790	No	SLC 12
	787	X	887	Si	SLC 12
	840	X	940	Si	SLC 12
SLC12A2	282	55	399	Si	AA Permeasa
	461	232	580	Si	AA Permeasa
	517	295	638	Si	AA Permeasa
	518	296	639	Si	AA Permeasa
	719	497	839	Si	SLC 12
SLC26A6	35	39	53	Si	-
	278	284	289	No	Sulfato transp
	308	314	319	X	Sulfato transp
	314	320	325	No	Sulfato transp
VDAC-2	6	6	17	X	Porina-3
	8	8	19	No	Porina-3
	198	198	209	Si	Porina-3
	218	218	229	Si	Porina-3

### 3.4.1 Sitios bajo selección

Incluso considerando los codones con p-valores no significativos en los modelos de CodeML, una forma de relajar las fuertes restricciones impuestas por los modelos implementados en este programa, muy pocos codones fueron identificados simultáneamente bajo selección positiva por alguno de los modelos utilizados de CodeML y MEME. Este bajo número de coincidencias es similar a lo ya observado en otros trabajos (ej. Stager et al. 2014) y sugiere que este procedimiento es conservador. Las diferencias en los resultados de estos métodos pueden atribuirse a las importantes diferencias en sus asunciones, detalladas en la sección “3.2.3 Análisis de selección” de esta tesis.

Desafortunadamente, la información disponible sobre la estructura terciaria y los cambios conformacionales asociados con la actividad de la mayoría de los genes es limitada o simplemente no disponible. Por ejemplo, una búsqueda en la base de datos “Protein Data Bank” mediante BLASTP identificó modelos 3D de proteínas adecuados para el gen VDAC-2, CFTR sin incluir el dominio CFTR-R y el C' terminal, y los dominios conservados de CLCN-3 con respecto a CLCN-5 (datos no mostrados). Por ello la discusión de los impactos funcionales se limitará a los genes y sitios para los que se dispone de un mínimo de información: CFTR, CLCN-3, SLC12A1 y VDAC-2.

#### 3.4.1.1 Sitios bajo selección positiva en el gen CFTR

Considerando que el codón 833 se encuentra en el dominio CFTR-R, el codón 1400 es homólogo a un sitio de fosforilación en humanos (McClure et al. 2012) y 1406 se encuentra en la cercanía de este, es esperable que estas sustituciones afectaran la regulación del gen. De hecho, la regulación de CFTR depende esencialmente de eventos de fosforilación, principalmente en el dominio “Regulador de la Conductancia de la Fibrosis Quística” (Bozoky et al. 2013). Desafortunadamente, la estructura del dominio regulador ha demostrado ser particularmente difícil de determinar debido a su flexibilidad (Bozoky et al. 2013; Zhang & Chen 2016). La señal de selección en el codón 953 por otro lado, al localizarse en una  $\alpha$  hélice transmembrana que forma parte del canal (Fig. 19), es más probable que afecte directamente el funcionamiento del canal.

Si bien se requieren análisis más exhaustivos para analizar su importancia funcional, es interesante observar el número de sustituciones ocurridas en el codón 1400. Este sitio es homólogo a una Serina (McClure et al. 2012), igual que en la mayoría de nuestro conjunto de datos, pero que es sustituida por Asparagina en los cíclidos estudiados (*Haplochromis burtoni*, *Maylandia zebra*, *Neolamprologus brichardi*, *Pundamilia nyererei*, y *Oreochromis niloticus*); Isoleucina en las especies *Fundulus heteroclitus*, *Poecilia*

*formosa*, *Cyprinodon variegatus* y *Takifugu rubripes*; Treonina en las especies *Austrofundulus limnaeus* y *Oryzias latipes*; y Valina en *Tetraodon nigroviridis*.

#### 3.4.1.2 Sitios bajo selección positiva en el gen CLCN-3

La señal de selección en el codón 687 del gen CLCN-3 sin una sustitución del aminoácido Serina llama la atención. Como los métodos utilizados asumen que la probabilidad de que ocurra un cambio de un codón con más de dos sustituciones nucleotídicas es 0 (Kosakovsky et al. 2011; Murrell et al. 2012; Nielsen & Yang 1998), una explicación es que al no existir sustituciones sinónimas que permitan pasar de uno de los codones a otro, estos métodos lo detectan bajo selección positiva. Bajo este escenario, en algún momento ocurrió una sustitución no sinónima en esta posición que luego fue revertida a Serina (ej. AGC → ACC → TCC). Otra posibilidad es la ocurrencia de un único evento de mutación que afectó la primer y segunda posición del codón, similar a lo ya reportado en varios tipos de cáncer (Chang et al. 2016; Rubin & Green 2009).

#### 3.4.1.3 Sitios bajo selección positiva en el gen SLC12A1

Sin conocer la estructura terciaria del canal SLC12A1 es muy poco lo que se puede hipotetizar sobre el impacto funcional de los sitios encontrados bajo selección positiva para el gen SLC12A1. Dicho eso, la búsqueda en Uniprot indica que sitio homólogo al codón 363 en humanos se encuentra en las cercanías a un lugar predicho de glicosilación ligada a nitrógeno. Por lo que puede que las sustituciones en este sitio afecten la regulación del gen.

#### 3.4.1.4 Sitios bajo selección positiva en el gen VDAC-2

La búsqueda en Uniprot no arrojó resultados concretos para los codones bajo selección positiva de VDAC-2. Sin embargo estos se ubican en la cercanía del extremo N-terminal de la proteína, el cual es importante para el funcionamiento del canal (Mannella 1998) (Fig. 26). Esta localización sugiere que estas sustituciones impactan en la regulación del transporte de metabolitos en la mitocondria, pero también deben investigarse en más detalles su posible importancia en otras funciones del gen como por ejemplo su papel en la apoptosis (Shoshan-Barmatz et al. 2010).

### 3.4.2 Linajes bajo selección

Es significativo haber encontrado varios genes bajo selección positiva en los mismos linajes en lugar de estar dispersos en la filogenia, especialmente considerando que los

modelos de Ramas-Sitios de CodeML pierden poder de detección cuando aumenta la saturación de los sitios, pero manteniendo los falsos positivos bajo control (Gharib & Robinson-Rechavi 2013).

Varios trabajos con dataciones moleculares sitúan los últimos ancestros comunes de los linajes Euteleosteomorpha y Otomorpha en algún momento entre finales del Pérmico y comienzos del Triásico medio (Betancur et al. 2013; Broughton et al. 2013; Nakatani et al. 2011; Near et al. 2012); mientras que el linaje ancestral de Salmoniformes y Esociformes, Atherinomorpha y Cyprinodontiformes en distintos momentos del Cretácico (Betancur et al. 2013; Broughton et al. 2013; Campbell et al. 2013; Nakatani et al. 2011; Near et al. 2012).

Debido a limitaciones en el muestreo de especies, no pude establecerse con certeza a que grupos corresponde la señal de selección positiva encontrada para el linaje Nodo 10 del gen SLC12A3, por lo que no van a ser tenidos en cuenta en esta discusión. Tampoco lo será el linaje ancestral de Cyprinodon + Fundulus para el gen SLC9A1 por los resultados anómalos del modelo de Ramas-Sitios.

#### 3.4.2.1 Señal de selección positiva en los linajes Euteleosteomorpha y Otomorpha

Es muy interesante que los grupos que comprenden a la gran mayoría de los teleósteos actuales (Nelson 2006) hayan divergido junto con la mayor extinción masiva de la historia de la Tierra, la extinción del Pérmico-Triásico. Más aún, aquí se presenta evidencia de que genes importantes en la osmorregulación estuvieron bajo selección positiva durante este mismo período.

Esta extinción eliminó el 90% de las especies marinas (Knoll et al. 2007). Aunque no libre de discusión, el detonante más aceptado de la extinción es la formación de las Escaleras Siberianas (*Siberian Traps*), una provincia ígnea que cubre un área de ( $\sim 7 \times 10^6$  km<sup>2</sup>) y un volumen de ( $\sim 4 \times 10^6$  km<sup>3</sup>) (Ivanov et al. 2013; Chen & Benton 2012). Las formas de vida de este período debieron afrontar anoxia, hipercapnia, acidificación de los océanos, envenenamiento por ácido sulfhídrico y altas temperaturas (Knoll et al. 2007; Shen et al. 2016; Song et al. 2014).

La recuperación de la diversidad durante el Triásico fue anormalmente larga, extendiéndose hasta 10 millones de años dependiendo del taxón (Chen & Benton 2012), debido al menos en parte, a la persistencia de condiciones adversas hasta el Triásico medio (Pietsch & Bottjer 2014). Esta larga recuperación es la que permite proponer los cambios ambientales de este período como los causantes de las señales de selección encontradas, a pesar de los amplios márgenes de error de las dataciones de la diversificación de estos grupos. Por ejemplo la datación del grupo Euteleostei



(Euteleostomorpha en Betancur-R (2013)) obtenida por Near et al (2012) tiene un intervalo de confianza del 95% entre los 220,2 a 258,5 millones de años atrás.

Hay dos formas, no necesariamente excluyentes, en las que los cambios ambientales desencadenados al final del Pérmico pudieron traducirse en presiones selectivas sobre los mecanismos osmorregulatorios:

a) Adaptación al incremento de la acidez

Los mecanismos moleculares implicados en la osmorregulación están muy relacionados con la regulación ácido-base (Hwang et al. 2011; Marshall & Grosell 2005), especialmente en agua salada, donde el gradiente de Na<sup>+</sup> con el medio es lo que impulsa la secreción de H<sup>+</sup> (Hwang et al. 2011). Por ello en agua salada la regulación del PH corporal y el balance de la osmolaridad son antagónicos, un aumento notorio en la acidez oceánica podría haberse traducido en un incremento en la carga de trabajo sobre los mecanismos de eliminación del exceso de Na<sup>+</sup> y Cl<sup>-</sup>. Un gran punto a favor de esta hipótesis corresponde al hecho que los primeros fósiles encontrados de Teleostei en el Triásico Medio son especies marinas, potencialmente eurihalinas (Furrer 1995; Ticino & Bürgin 1995; Tintori 1990).

De estar en lo cierto; debería de encontrarse evidencia de selección positiva en los genes directamente implicados en la regulación ácido-base. De momento solo se ha analizado el gen SLC9A2 de los genes que participan en la regulación del PH corporal y los resultados sugieren que no estuvo bajo selección positiva durante la diversificación del grupo.

b) Disponibilidad de refugios

Durante este período, las especies que fueran incapaces de tolerar las condiciones adversas, debieron desplazarse a ambientes con condiciones más propicias. Los refugios propuestos a la fecha son ambientes particularmente bien oxigenados (Beatty et al. 2008; Zonneveld et al. 2010), para escapar de la anoxia y la hipercapnia (Knoll et al. 2007). En su mayoría ambientes de aguas poco profundas susceptibles a aportes de agua dulce desde el continente (Beatty et al. 2008; Zonneveld et al. 2010), incluyendo ambientes estuarinos (Gall & Grauvogel-Stamm 2005; Haig et al. 2015).

Aunque estos refugios han sido descritos principalmente para grupos de invertebrados; debido a que su registro fósil tiende a ser más consistente que el de vertebrados (Romano et al. 2016), se ha reportado un desplazamiento a ambientes dulceacuícolas de varios Palaeonisciformes (generos: *Acrolepis*, *Atherstonia*, *Palaeoniscum* y posiblemente *Pteronisculus* (Romano et al. 2016)). Concordante con esta idea, los fósiles de Teleostei

del Triásico medio provienen de ambientes marinos influenciados por agua dulce (Furrer 1995; Ticino & Bürgin 1995; Tintori 1990).

Por lo expuesto, es plausible que los ancestros de los linajes Euteleosteomorpha y Otomorpha fueran empujados a esta clase de ambientes. Este desplazamiento forzado a nuevos hábitats puede explicar por sí mismo la evidencia de selección encontrada en estos linajes, al margen de la acidificación oceánica. Sin embargo, para probarlo se necesitan fósiles provenientes del Triásico temprano para descartar refugios que no requirieran adaptarse a cambios de salinidad, como por ejemplo regiones oxigenadas de mares someros con poca influencia de agua dulce o mares profundos (Algeo et al. 2011; Guinot et al. 2013; Winguth & Winguth 2012).

#### 3.4.2.2 Señal de selección positiva en el linaje ancestral de Esociformes y Salmoniformes

Curiosamente, la señal de selección positiva detectada en el linaje ancestral de Esociformes y Salmoniformes es previa al establecimiento de su estilo de vida anádromo en salmoniformes y la duplicación genómica específica de salmoniformes (Davidson 2013). Algo ya observado en la subunidad  $\alpha 1a$  de la NKA (subunidad expresada en agua dulce en *S. salar*) (Dalziel et al. 2014). Varios autores datan la divergencia los órdenes Esociformes y Salmoniformes en el Cretácico temprano/comienzos del tardío (Betancur-R. et al. 2013; Campbell et al. 2013; Near et al. 2012), y su distribución actual, y el registro fósil, los sitúan en Laurasia (Campbell et al. 2013) durante la formación del Atlántico Norte (Seton et al. 2012).

El clima del Cretácico se caracterizó por temperaturas cálidas y un alto nivel del mar (Holz 2015), llevando a la formación de varios mares epicontinentales alrededor del mundo (ej. la “Western interior seaway” que cubrió la región central de la actual Norteamérica durante 100 millones de años (Miall et al. 2008)). Estas transgresiones marinas, en conjunto con la abertura del atlántico norte (Seton et al. 2012), seguramente redujeron significativamente la disponibilidad de ambientes estrictamente dulceacuícolas. La evidencia de selección aquí encontrada quizás represente una huella molecular de la adaptación a vivir en estos nuevos ambientes más expuestos al agua salada. Desafortunadamente, no se dispone de evidencia en el registro fósil que respalde esta hipótesis ya que muy pocos fósiles de este grupo han sido encontrados provenientes de este período (ej. Wilson et al. 1992).

#### 3.4.2.3 Señal de selección positiva en los linajes Atherinomorpha y Cyprinodontiformes

Debido a la ausencia de secuencias de *O. argentinensis* para los genes SLC9A1 y CFTR (Figs. 10 y 14 respectivamente), la asignación de la señal del grupo a

Atherinomorpha y su ubicación posterior a la divergencia de los Atheriniformes (Fig. 17) es relativamente arbitraria. Sin embargo dada la cercanía filogenética de estos grupos y que los rangos temporales estimados para la divergencia de estos grupos son similares (Betancur-R. et al. 2013; Broughton et al. 2013; Campanella et al. 2015; Near et al. 2012) la interpretación de los resultados no se ve afectada.

Con respecto a los Cyprinodontiformes, en este estudio solo se incluyeron especies nativas de Norteamérica y Sudamérica. Pero la inclusión de la especie *Austrofundulus limnaeus* permite acotar la señal de selección encontrada previa a la divergencia entre Aplocheiloidei y Cyprinodontoidei, los dos subórdenes de Cyprinodontiformes (Pohl et al. 2015). El grupo en si presenta una distribución principalmente Gondwánica, lo que ha generado varias discusiones sobre la edad del grupo y los mecanismos de dispersión: la fragmentación de Gondwana y/o eventos de dispersión a través los océanos en formación por parte de especies eurihalinas en el grupo (Costa 2013; Pohl et al. 2015).

A lo largo de su evolución, los Atherinomorpha (incluyendo Cyprinodontiformes) han invadido repetidamente ambientes dulceacuícolas, estuarinos y marinos (Campanella et al. 2015; Nelson 2006) pero ninguna está asociada a las señales de selección encontradas. Las dataciones moleculares sitúan al ancestro común de los Cyprinodontiformes y Atherinomorpha en el Cretácico tardío (Betancur-R. et al. 2013; Broughton et al. 2013; Nakatani et al. 2011; Near et al. 2012; Near et al. 2013), por lo que los mismos cambios en el nivel del mar que afectaron al ancestro de Esociformes y Salmoniformes podrían ser responsables de las señales de selección encontradas en este grupo.

#### 3.4.2.4 Señal de selección en linajes terminales

Interpretar las señales de selección positiva en linajes terminales encontradas en los genes SLC9A1 (*A. limnaeus* y *F. heteroclitus*), SLC12A1 (*N. coriiceps* y *S. partitus*) y SLC26A6 (*D. rerio*) es más difícil y relativamente más abierto a interpretaciones. Ninguno de estas ramas fue encontrada bajo selección por más de un gen, por lo que no se puede asegurar que efectivamente sea su rol en la osmorregulación el causante de estar sometido bajo selección positiva en estas especies. Otro problema que afecta la interpretación de estas señales de selección es la posibilidad de que representen falsos positivos causados por errores en la secuenciación y/o alineamiento de las secuencias, u otro tipo de artefacto. Los resultados anómalos del LRT de los modelos de Ramas-Sitios para *A. limnaeus* sugieren que podría ser un ejemplo de esto.

### 3.4.3 Observaciones sobre la metodología utilizada.

El estudio aquí realizado presenta varias limitaciones que es necesario considerar a la hora de evaluar los resultados obtenidos. En su mayoría son problemáticas de la aplicación de los métodos para la detección de selección basados en las tasas de sustitución en general, debido a violaciones de las asunciones de los modelos utilizados. A continuación se discutirá las principales limitaciones que afectan este estudio y las medidas que se tomaron, o pueden realizarse en futuros estudios, para reducir su efecto.

#### a) Sesgo en la representación de especies

El sesgo más notorio en los datos es la sobre-representación de especies dulceacuícolas y/o estuarinas. Mientras que las especies de teleósteos que habitan, o entran en contacto frecuentemente con el agua dulce, representan cerca del 46% de las especies descritas (Nelson 2006), en este estudio son el 80% (44 especies) las que pertenecen a esta categoría.

Esto responde a la disponibilidad de datos en las bases de datos públicas, las que parecen reflejar los grupos que han recibido mayor interés por parte de los biólogos y los problemas logísticos en la obtención y secuenciación de las muestras de especies marinas. Por ejemplo, de las 11 especies de teleósteos disponibles en la base de datos Ensembl solo tres son marinas (*G. aculeatus*, *G. morhua* y *T. rubripes*) y todas son tolerantes a cierto grado de contacto con el agua dulce. A esto se suman sesgos en la procedencia filogenética de las especies: Cuatro pertenecen al grupo Ovalenteridae (*O. latipes*, *O. niloticus*, *P. formosa*, *X. maculatus*) y dos al orden Tetraodontiformes (*T. nigroviridis* y *T. rubripes*).

Este problema podrá ser solventando en futuros estudios a medida que se hagan públicos más datos genómicos y/o transcriptómicos de especies de teleósteos. Proyectos como PhyloFish (Pasquier et al. 2016) y Fish-T1K (Sun et al. 2016) prometen una solución en el futuro cercano.

#### b) Filogenia utilizada

Los análisis aquí realizados son susceptibles al uso de una filogenia inadecuada. Afortunadamente, las relaciones de parentesco entre las especies incluidas en el análisis son recuperadas frecuentemente en varias filogenias (Mirande 2016; Near et al. 2012) para las especies incluidas en este estudio. Excepciones a esto son las relaciones internas de Atherinomorpha, Otophysi y Protacanthopterygi. Ambos grupos con un largo historial de filogenias conflictivas (Nakatani et al. 2011; Nelson 2006).

### 1) Discrepancias en el linaje Atherinomorpha

En la actualidad no existe un consenso sobre las relaciones internas dentro de Atherinomorpha (Betancur-R. et al. 2013; Mirande 2016; Near et al. 2012; Setiamarga et al. 2008): Atheriniformes, Beloniformes y Cyprinodontiformes. Al solo disponerse de secuencias de *O. argentinensis* para este grupo, solo los análisis para los genes VDAC-1 y VDAC-2 estarían afectados por la discrepancias en las reconstrucciones filogenéticas de este grupo.

### 2) Discrepancias en la filogenia del linaje Protacanthopterygi

Los órdenes pertenecientes a Protacanthopterygi son motivo de constante revisión (Nelson 2006). En lo referente a este estudio, se afecta la posición del grupo Osmeriformes (*Plecoglossus altivelis*). En una versión actualizada de la filogenia utilizada en este estudio, Betancur-R, ([http://www.deepfin.org/Classification\\_v3.htm](http://www.deepfin.org/Classification_v3.htm)), sitúa a este linaje (junto con Stomiatiformes) por fuera de Protacanthopterygi, como grupo hermano de los Neoteleostei; pero con un Bootstrap <75%. Mirande (2016) también encontró dificultades para determinar las relaciones filogenéticas del grupo. Aunque su hipótesis final fue consistente, para este grupo, con la topología de Betancur-R (2013).

### 3) Discrepancias en la filogenia del linaje Otophysi

Distintos autores han llegado a conclusiones contradictorias sobre las relaciones de parentesco entre los órdenes Gymnotiformes, Characiformes y Siluriformes. Afectando la posición de las especies *Electrophorus electricus*, *Apteronotus albifrons* y *Sternopygus macrurus* (Gymnotiformes), *Astnax mexicanus* (Characiformes), *Ictalurus furcatus* y *Pangasianodon hypophthalmus* (Siluriformes).

Además del agrupamiento aquí utilizado (Gymnotiformes (Characiformes + Siluriformes)), varios autores han propuesto una amplia gama de propuestas alternativas, algunas incluso implican la parafilia del grupo Characiformes (ver (Nakatani et al. 2011) para una breve revisión histórica). Como en este estudio solo se incluyó una única especie de Characiformes (*A. mexicanus*): el árbol utilizado es consistente con lo propuesto por varios autores, incluyendo varias en las que el grupo es parafilético (ej. Mirande 2016; Near et al. 2012). Por lo que las relaciones internas de este grupo no deberían ser un problema en este estudio en particular.

Mientras que para este estudio en particular la elección de la filogenia parecería haber sido la adecuada, teniendo en cuenta la baja cobertura de especies en grupos conflictivos, la filogenia a seguir va a ser una elección más crítica a medida que se amplíe el número

de especies estudiado. Principalmente al abordar grupos como Protacanthopterygi o Eupercanomorpha (Betancur-R. et al. 2013; Mirande 2016).

c) Errores en el alineamiento

La principal fuente de falsos positivos en un análisis de selección provienen de errores en el alineamiento de las secuencias (Fletcher & Yang 2010), y se ha reportado que el uso de distintos algoritmos de alineación de secuencias puede cambiar significativamente los resultados (Blackburne & Whelan 2013; Fletcher & Yang 2010; Markova-Raina & Petrov 2011). Para lidiar con este problema se recurrió al alineador MAFFT; utilizado en múltiples estudios filogenéticos recientes (Alda et al. 2013; Betancur-R. et al. 2013; Campanella et al. 2015; Campbell et al. 2013; Tagliacollo et al. 2016), combinado con edición manual, la cual fue posible gracias al limitado conjunto de datos utilizado.

Sería interesante en futuros análisis el uso de alternativas como PRANK (Löytynoja & Goldman 2008). Este algoritmo de alineamiento supera el desempeño de MAFFT en lo referente a calidad de alineamiento, en algunas regiones problemáticas, gracias su manejo fino de los indels (Fletcher & Yang 2010). Si bien es poco probable que los resultados obtenidos sean alterados por el uso de este algoritmo, en la edición manual se eliminaron las regiones conflictivas donde PRANK tiene un mejor desempeño (Löytynoja & Goldman 2008), el uso de PRANK probablemente se traduciría en un mayor aprovechamiento de los datos. Aunque es discutible si realmente es seguro analizar regiones cuyo alineamiento es tan dependiente del algoritmo utilizado, cuando un enfoque más conservador es posible.

d) Presencia de recombinación

Dependiendo de los datos analizados, la existencia de eventos de recombinación en las secuencias analizadas puede ser una importante fuente de falsos positivos en esta clases de análisis si no es adecuadamente controlada (Anisimova et al. 2003; Pond et al. 2006). Hay dos posibles mecanismos por los que la recombinación puede afectar la inferencia de selección (Anisimova et al. 2003): alterando el largo de las ramas de la filogenia utilizada y haciendo que la topología de la filogenia utilizada sea inválida. En el caso de este estudio particular la escala temporal entre la mayoría de los nodos (la mayoría datados en el Cretácico (Near et al. 2012)), hace que la recombinación tenga un efecto limitado. Por este motivo es difícil que uno de estos eventos invalide la filogenia utilizada o afecte de forma desmedida el largo de las ramas.

### 3.5 Conclusiones finales Capítulo 3

Las señales de selección positiva encontradas en este estudio parecen reflejar la compleja historia de los peces teleósteos y su compleja relación con el mundo que habitaron. La extinción del Pérmico-Triásico forzó a los organismos a retirarse a ambientes mejor oxigenados evitando la anoxia, entre los que se cuentan regiones estuarinas, mientras que la formación de los extensos mares epicontinentales del cretácico redujo significativamente la disponibilidad de ambientes puramente dulceacuícolas. Quizás más importante, ambos eventos se extendieron y persistieron en el tiempo durante varios millones de años, reforzando su impacto en la evolución de los organismos.

El cambio climático más dramático, por un amplio margen, fue la extinción a finales del Pérmico y las condiciones adversas que le siguieron del Triásico temprano. Cabe preguntarse qué otras funciones biológicas fueron sometidas bajo selección durante este período. Genes implicados en la tolerancia a la anoxia y la acidez son firmes candidatos a haber estado bajo selección.

## Conclusiones finales

En esta tesis de maestría se presentaron dos modestos aportes al conocimiento de cómo los peces actinopterigios enfrentan la diferencia de osmolaridad del medio que habitan con sus cuerpos.

El primer aporte es la secuenciación y ensamblado de varios transcriptomas branquiales de *O. argentinensis*, presentados en el Capítulo 2. La obtención de estos datos representan los primeros pasos para hacer de esta especie un nuevo modelo para estudiar la osmorregulación de los peces Actinopterigios. Aunque relativamente cercano filogenéticamente a otras especies modelo como *O. mossambicus* y *F. heteroclitus* (Betancur-R. et al. 2013; Mirande 2016), *O. argentinensis* es interesante porque los Atheriniformes han invadido medios dulceacuícolas y marinos repetidamente durante el cenozoico (Campanella et al. 2015).

El otro aporte aquí presentado es la identificación de evidencia de selección natural en varios linajes de Actinopterigios, presentados en el Capítulo 3, comprendiendo a la gran mayoría de peces existentes en la actualidad (Nelson 2006). Sorprendentemente, los linajes encontrados bajo selección no parecen estar asociados a invasiones duraderas a un medio acuático con régimen salino distinto al de origen, pero si lo hacen con varios eventos en la historia geológica que arrastraron muchas especies a medios eurihalinos.

## Perspectivas futuras

Los transcriptomas de *O. argentinensis* generados ofrecen la materia prima necesaria para seguir expandiendo nuestro conocimiento sobre la biología de la especie y los Atheriniformes en general. Un estudio comparativo entre *O. argentinensis* y *O. bonaerensis*, dos especies cercanas filogenéticamente pero que ocupan ambientes con salinidades diferentes (Dyer 2006; Campanella et al. 2015; Setiamarga et al. 2008), sería particularmente valioso para entender los mecanismos osmorregulatorios de los Atheriniformes.

Por su lado, los resultados de la búsqueda de evidencia de selección en genes involucrados en la osmorregulación de Actinopterigios son una ventana a los procesos que moldearon la evolución del grupo. La mayoría de las señales de selección encontradas a nivel de linajes coinciden con la extinción del Pérmico-Triásico. Sería valioso extender este análisis genes involucrados en la respuesta a la acidez y a la anoxia en busca de señales de selección análogas a las encontradas, condiciones más importante para la supervivencia de los seres vivos durante este periodo (Knoll et al. 2007; Shen et al. 2016; Song et al. 2014).



Otra extensión interesante de este estudio sería incluir otros grupos marinos que fueron afectados por esta extinción. Idealmente estos deben ser grupos en los que sus filogenéticas internas sean comprendidas con suficiente detalle, que se conozca la identidad de los genes implicados en la osmorregulación y se disponga de un registro fósil abundante e ininterrumpido durante este periodo. Un buen grupo candidato son los Decápodos. Su registro fósil es más completo, incluye su presencia en varios de los mares poco profundos que hicieron de refugios durante este período (Haig et al. 2015; Zonneveld et al. 2010) y se dispone de información básica sobre los mecanismos moleculares de su respuesta osmorregulatoria (ej. Havird et al. 2016), aunque desafortunadamente aún no hay un consenso las relaciones filogenéticas internas (Ahyong & Meally 2004; Bybee et al. 2011; Porter et al. 2005; Tsang et al. 2008).

## **Anexos:**

Anexo A: Resultados de la anotación funcional del ensamblado conjunto de *O. argentinensis*: Términos PANTHER GO-Slim encontrados significativamente sub- o sobrerrepresentados dentro de Función Molecular (Tabla A1), Componente Celular (Tabla A2) y Proceso Biológico (Tabla A3). Categorías Proteicas de PANTHER presentes en los ensamblados de *O. argentinensis* (Tabla A4). Clases Proteicas de PANTHER significativamente sobre o subrepresentadas (Tabla A5). Pathways de PANTHER presentes en los ensamblados de *O. argentinensis* (Tabla A6). Pathways de KEGG presentes en los ensamblados de *O. argentinensis* (Tabla A7). Pathways de KEGG significativamente sobre o subrepresentadas (Tabla A8). Términos GO influidos en "Transportadores Transmembrana (GO: 0022857)" (Tabla A9). Pathways encontradas sub- o sobrerrepresentadas en al menos uno de los ejemplares analizados de *O. argentinensis* con respecto a los transcriptomas de *O. mossambicus* y *S. alpinus* (Tabla A10). Clases Proteicas encontradas sub- o sobrerrepresentadas en al menos uno de los ejemplares analizados de *O. argentinensis* con respecto a los transcriptomas de *O. mossambicus* y *S. alpinus* (Tabla A10). Lista Completo de los genes encontrados en la base de datos PANTHER (Tabla A12).

Anexo B: Comandos usados de Trimmomatic v0.35 para editar los datos descargados de la base de datos SRA (Tabla B1).

Anexo C: Disponibilidad de secuencias de cada gen para las especies utilizadas: Ensembl (Tabla C1), Protein (Tabla C2) y SRA (Tabla C3).

Anexo D: Árboles resultantes del análisis de BSR para cada gen: AQP-9b (Fig. D1); AQP-10b (Fig. D2); AQP-12 (Fig. D3); CFTR (Fig. D4); CLCN-3 (Fig. D5); SLC9A1 (Fig. D6); SLC9A2 (Fig. D7); SLC12A1 (Fig. D8); SLC12A2 (Fig. D9); SLC12A3 (Fig. D10); SLC26A6 (Fig. D11); VDAC-1 (Fig. D12); VDAC-2 (Fig. D13).

Anexo E: Datos suplementarios de los resultados obtenidos mediante modelos de CodeML: Logaritmo del valor de Verosimilitud de los modelos de Ramas de CodeML (Tabla E1). Logaritmo del valor de Verosimilitud de los modelos de Sitios de CodeML (Tabla E2). Estimación de W de los modelos de Ramas-Sitios (Tabla E3). Sitios identificados como candidatos para los genes CFTR por los modelos de Ramas y Sitios de CodeML para los linajes Esociformes + Salmoniformes (Tabla E4), Euteleostomorpha (Tabla E5), Atherinomorphae (Tabla E6). Sitios identificados como candidatos para los genes CLCN-3 por los modelos de Ramas y Sitios de CodeML para el linaje Euteleostomorpha (Tabla E7). Sitios identificados como candidatos para los genes SLC9A1 por los modelos de Ramas y Sitios de CodeML para los linajes Esociformes + Salmoniformes (Tabla

E8) y *F. heteroclitus* (Tabla E9). Sitios identificados como candidatos para los genes SLC12A1 por los modelos de Ramas y Sitios de CodeML para los linajes Euteleostomorpha (Tabla E10), *N. coriiceps* (Tabla E11), *S. partitus* (Tabla E12). Sitios identificados como candidatos para los genes SLC12A2a por los modelos de Ramas y Sitios de CodeML para los linajes Cyprinodontiformes (Tabla E13), Euteleostomorpha (Tabla E14), Otomorpha (Tabla E15). Sitios identificados como candidatos para los genes SLC12A3 por los modelos de Ramas y Sitios de CodeML para los linajes Notozosteoidei (Tabla E16), Otomorpha (Tabla E17). Sitios identificados como candidatos para los genes SLC26A6 por los modelos de Ramas y Sitios de CodeML para los linajes *D. rerio* (Tabla E18) y Euteleostomorpha (Tabla E19). Sitios identificados como candidatos para los genes VDAC-2 por los modelos de Ramas y Sitios de CodeML para los linajes Esociformes + Salmoniformes (Tabla E20), Cyprinodontiformes (Tabla E21)

Anexo F: Sitios identificados como candidatos de estar bajo selección por MEME: AQP-9 (Tabla F1), AQP-10b (Tabla F2), AQP-12 (Tabla F3), CFTR (Tabla F4), CLCN-3 (Tabla F5), SLC9A1 (Tabla F6), SLC9A2 (Tabla F7), SLC12A1 (Tabla F8), SLC12A2a (Tabla F9), SLC12A3 (Tabla F10), SLC26A6 (Tabla F11), VDAC-1 (Tabla F12) y VDAC-2 (Tabla F13)

Anexo G: árboles resultantes del análisis de MEME para cada gen: AQP-9b (Fig. G1-G2); AQP-10b (Fig. G3-G5); AQP-12 (Fig. G6); CFTR (Fig. G7-G28); CLCN-3 (Fig. G29-G39); SLC9A1 (Fig. G40-G51); SLC9A2 (Fig. G52-G61); SLC12A1 (Fig. G62-G91); SLC12A2a (Fig. G92-G108); SLC12A3 (Fig. G109-G117); SLC26A6 (Fig. G118-G142); VDAC-1 (Fig. G144); VDAC-2 (Fig. G145-G152).

## Bibliografía

- Abbas, L. & Whitfield, T.T., 2009. Nkcc1 (Slc12a2) is required for the regulation of endolymph volume in the otic vesicle and swim bladder volume in the zebrafish larva. *Development*, 136(6), pp.2837–2848.
- Ahyong, S.T. & Meally, D.O., 2004. Phylogeny of the decapoda reptantia: resolution using three molecular loci and morphology. *The Raffles Bulletin of Zoology*, 52(2), pp.673–693.
- Alda, F. et al., 2013. Phylogeny and biogeography of the *Poecilia sphenops* species complex (Actinopterygii, Poeciliidae) in Central America. *Molecular Phylogenetics and Evolution*, 66(3), pp.1011–1026. Available at: <http://dx.doi.org/10.1016/j.ympev.2012.12.012>.
- Algeo, T.J. et al., 2011. Spatial variation in sediment fluxes, redox conditions, and productivity in the Permian-Triassic Panthalassic Ocean. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 308, pp.65–83.
- Altenhoff, A.M. et al., 2015. The OMA orthology database in 2015: Function predictions, better plant support, synteny view and other improvements. *Nucleic Acids Research*, 43(D1), pp.D240–D249.
- Altschul, S.F. et al., 1990. Basic local alignment search tool. *Journal of molecular biology*, 215(3), pp.403–10. Available at: <http://www.sciencedirect.com/science/article/pii/S0022283605803602>.
- Anisimova, M., Nielsen, R. & Yang, Z., 2003. Effect of recombination on the accuracy of the likelihood method for detecting positive selection at amino acid sites. *Genetics*, 164(3), pp.1229–1236.
- Anisimova, M. & Yang, Z., 2007. Multiple Hypothesis Testing to Detect Lineages under Positive Selection that Affects Only a Few Sites. *Molecular Biology and Evolution*,

- 24(5), pp.1219–1228.
- Ashburner, M. et al., 2000. Gene ontology: Tool for the unification of biology. *Nature Genetics*, 25(1), pp.25–29.
- Bagherie-Lachidan, M., Wright, S.I. & Kelly, S.P., 2009. Claudin-8 and -27 tight junction proteins in puffer fish *Tetraodon nigroviridis* acclimated to freshwater and seawater. *Journal of Comparative Physiology B: Biochemical, Systemic, and Environmental Physiology*, 179(4), pp.419–431.
- Bahassi, E.M. & Stambrook, P.J., 2014. Next-generation sequencing technologies: Breaking the sound barrier of human genetics. *Mutagenesis*, 29(5), pp.303–310.
- Barros, S.E., Regidor, H. & Iwaszkiw, J., 2004. Biología pesquera del Pejerrey *Odontesthes bonariensis* (Cuvier y Valenciennes, 1835) en el subtrópico de Argentina. *Revista AquaTIC*, 20, pp.32–37. Available at: [http://revistaaquatic.com/aquatic/pdf/20\\_04.pdf](http://revistaaquatic.com/aquatic/pdf/20_04.pdf).
- Bateman, A. et al., 2015. UniProt: A hub for protein information. *Nucleic Acids Research*, 43(D1), pp.D204–D212.
- Beatty, T.W., Zonneveld, J.P. & Henderson, C.M., 2008. Anomalously diverse Early Triassic ichnofossil assemblages in northwest Pangea: A case for a shallow-marine habitable zone. *Geology*, 36(10), pp.771–774.
- Beheregaray, L.B. & Levy, J.A., 2000. Population Genetics of the Silverside *Odontesthes argentinensis* (Teleostei, Atherinopsidae): Evidence for Speciation in an Estuary of Southern Brazil. *Copeia*, 2, pp.441–447.
- Beheregaray, L.B. & Sunnucks, P., 2001. Fine-scale genetic structure, estuarine colonization and incipient speciation in the marine silverside fish *Odontesthes argentinensis*. *Molecular Ecology*, 10(12), pp.2849–2866.
- Bemvenuti, M. a., 2006. Silversides in South Brazil: morphological and ecological aspects. *Biocell*, 30(1), pp.111–118.
- Betancur-R., R. et al., 2013. The Tree of Life and a New Classification of Bony Fishes. *PLOS Currents Tree of Life*, pp.1–51.
- Betancur, R. et al., 2013. Addressing gene tree discordance and non-stationarity to resolve a multi-locus phylogeny of the flatfishes (Teleostei: Pleuronectiformes). *Systematic Biology*, 62(5), pp.763–785.
- Bhatt, S., Holmes, E.C. & Pybus, O.G., 2011. The genomic rate of molecular adaptation of the human influenza A virus. *Molecular Biology and Evolution*, 28(9), pp.2443–2451.
- Blackburne, B.P. & Whelan, S., 2013. Class of multiple sequence alignment algorithm affects genomic analysis. *Molecular Biology and Evolution*, 30(3), pp.642–653.
- Bolger, A.M. et al., 2014. Trimmomatic: A flexible trimmer for Illumina Sequence Data. *Bioinformatics.*, 30, pp.1–7.
- Bossus, M. et al., 2013. The CIC-3 chloride channel and osmoregulation in the European Sea Bass, *Dicentrarchus labrax*. *Journal of Comparative Physiology B*, 183(5), pp.641–662. Available at: <http://dx.doi.org/10.1007/s00360-012-0737-9>  
<http://www.springerlink.com>  
<http://archimer.ifremer.fr>.
- Bozoky, Z. et al., 2013. Structural changes of CFTR R region upon phosphorylation: A plastic platform for intramolecular and intermolecular interactions. *FEBS Journal*, 280(18), pp.4407–4416.

- Broughton, R.E. et al., 2013. Multi - locus phylogenetic analysis reveals the pattern and tempo of bony fish evolution. *PLOS Currents Tree of Life*, pp.1–36.
- Brown, C.T. et al., 2012. A Reference-Free Algorithm for Computational Normalization of Shotgun Sequencing Data. *arXiv*, pp.1–18. Available at: <http://arxiv.org/abs/1203.4802>.
- Butlin, R. et al., 2012. What do we need to know about speciation? *Trends in Ecology and Evolution*, 27(1), pp.27–39.
- Bybee, S.M. et al., 2011. Directed next generation sequencing for phylogenetics: An example using Decapoda (Crustacea). *Zoologischer Anzeiger*, 250(4), pp.497–506. Available at: <http://dx.doi.org/10.1016/j.jcz.2011.05.010>.
- Campanella, D. et al., 2015. Multi-locus fossil-calibrated phylogeny of Atheriniformes (Teleostei, Ovalentaria). *Molecular Phylogenetics and Evolution*, 86, pp.8–23. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S1055790315000561>.
- Campbell, M.A. et al., 2013. Pike and salmon as sister taxa: Detailed intraclade resolution and divergence time estimation of Esociformes+Salmoniformes based on whole mitochondrial genome sequences. *Gene*, 530(1), pp.57–65. Available at: <http://dx.doi.org/10.1016/j.gene.2013.07.068>.
- Cao, J. & Shi, F., 2012. Evolution of the RALF gene family in plants: Gene duplication and selection patterns. *Evolutionary Bioinformatics*, 2012(8), pp.271–292.
- Cerdà, J. & Finn, R.N., 2010. Piscine aquaporins: An overview of recent advances. *Journal of Experimental Zoology Part A: Ecological Genetics and Physiology*, 313 A(10), pp.623–650.
- Chang, M.T. et al., 2016. Identifying recurrent mutations in cancer reveals widespread lineage diversity and mutational specificity. *Nature biotechnology*, 34(2), pp.155–163. Available at: <http://dx.doi.org/10.1038/nbt.3391>  
<http://www.ncbi.nlm.nih.gov/pubmed/26619011>.
- Chen, Z.-Q. & Benton, M.J., 2012. The timing and pattern of biotic recovery following the end-Permian mass extinction. *Nature Geoscience*, 5(6), pp.375–383. Available at: <http://dx.doi.org/10.1038/ngeo1475>.
- Cock, P.J.A. et al., 2010. The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic Acids Research*, 38(6), pp.1767–1771.
- Collins, F.S., Morgan, M. & Patrinos, A., 2003. The Human Genome Project: lessons from large-scale biology. *Science*, 300(5617), pp.286–90. Available at: <http://www.sciencemag.org/cgi/doi/10.1126/science.1084564>  
<http://www.ncbi.nlm.nih.gov/pubmed/12690187>.
- Conesa, A. et al., 2016. A survey of best practices for RNA-seq data analysis. *Genome Biology*, 17(1). Available at: <http://genomebiology.biomedcentral.com/articles/10.1186/s13059-016-0881-8>.
- Costa, W.J.E.M., 2013. Historical biogeography of aplocheiloid killifishes (Teleostei: Cyprinodontiformes). *Vertebrate Zoology*, 63(2), pp.139–154.
- Cristobo, S.S., 2014. *Hábitos tróficos de cuatro especies de peces en los estuarios Pando, Solus Chico y Solis Grande (Dpto. Canelones, Uruguay)*.
- Cutler, C.P. & Cramb, G., 2008. Differential expression of absorptive cation-chloride-

- cotransporters in the intestinal and renal tissues of the European eel (*Anguilla anguilla*). *Comparative Biochemistry and Physiology - B Biochemistry and Molecular Biology*, 149(1), pp.63–73.
- Cutler, C.P. & Cramb, G., 2002. Two isoforms of the Na<sup>+</sup>/K<sup>+</sup>/2Cl<sup>-</sup> cotransporter are expressed in the European eel (*Anguilla anguilla*). *Biochimica et Biophysica Acta - Biomembranes*, 1566(1–2), pp.92–103.
- Dalziel, A.C. et al., 2014. Origins and functional diversification of salinity-responsive Na<sup>+</sup>, K<sup>+</sup> ATPase  $\alpha$ 1 paralogs in salmonids. *Molecular Ecology*, 23(14), pp.3483–3503.
- Davidson, W.S., 2013. Understanding salmonid biology from the Atlantic salmon genome. *Genome*, 56, pp.548–550. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/24237334>.
- Delport, W. et al., 2010. Datamonkey 2010: A suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics*, 26(19), pp.2455–2457.
- Demogines, A. et al., 2010. Ancient and recent adaptive evolution of primate non-homologous end joining genes. *PLoS Genetics*, 6(10), pp.1–12.
- Dyer, B.S.H., 2006. Systematic revision of the South American silversides (Teleostei, Atheriniformes). *Biocell*, 30(1), pp.69–88.
- Edgar, R.C., 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32(5), pp.1792–1797.
- Eklom, R. & Galindo, J., 2011. Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity*, 107(1), pp.1–15. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3186121&tool=pmcentrez&rendertype=abstract%5Cnhttp://www.ncbi.nlm.nih.gov/pubmed/21139633>.
- Esbaugh, A.J. & Grosell, M., 2014. Esophageal desalination is mediated by Na<sup>+</sup>, H<sup>+</sup> exchanger-2 in the gulf toadfish (*Opsanus beta*). *Comparative Biochemistry and Physiology - A Molecular and Integrative Physiology*, 171, pp.57–63. Available at: <http://dx.doi.org/10.1016/j.cbpa.2014.02.012>.
- Evans, D.H., 2008. Teleost fish osmoregulation: what have we learned since August Krogh, Homer Smith, and Ancel Keys. *American journal of physiology. Regulatory, integrative and comparative physiology*, 295, pp.R704–R713.
- Ewing, B. et al., 1998. Base-Calling of Automated Sequencer Traces Using. *Genome Research*, 8, pp.186–194.
- Feng, S.H. et al., 2002. Gene expression of Na<sup>+</sup>-K<sup>+</sup>-ATPase  $\alpha$ 1 and  $\alpha$ 3 subunits in gills of the teleost *Oreochromis mossambicus*, adapted to different environmental salinities. *Marine Biotechnology*, 4(4), pp.379–391.
- Finn, R.D. et al., 2014. Pfam: The protein families database. *Nucleic Acids Research*, 42(D1), pp.D222–D230.
- Fletcher, W. & Yang, Z., 2010. The effect of insertions, deletions, and alignment errors on the branch-site test of positive selection. *Molecular Biology and Evolution*, 27(10), pp.2257–2267.
- Forni, D. et al., 2015. Diverse selective regimes shape genetic diversity at ADAR genes and at their coding targets. *RNA biology*, 12(2), pp.149–61. Available at: [http://www.tandfonline.com.libproxy.albany.edu/doi/abs/10.1080/15476286.2015.1017215?url\\_ver=Z39.88-](http://www.tandfonline.com.libproxy.albany.edu/doi/abs/10.1080/15476286.2015.1017215?url_ver=Z39.88-)

2003&rfr\_id=ori:rid:crossref.org&rfr\_dat=cr\_pub%3Dpubmed.

- Freeman, S. & Herron, J.C., 2014. *Evolutionary Analysis*, Available at: <http://www.amazon.com/dp/B001KQUG9S>.
- Fuchsman, C.A. & Rocap, G., 2006. Whole-Genome Reciprocal BLAST Analysis Reveals that Planctomycetes Do Not Share an Unusually Large Number of Genes with Eukarya and Archaea. , 72(10), pp.6841–6844.
- Fullwood, M.J. et al., 2009. Next-generation DNA sequencing of paired-end tags (PET) for transcriptome and genome analyses. *Genome Research*, 19(4), pp.521–532.
- Furrer, H., 1995. The Prosanto Formation, a marine Middle Triassic fossil-Lagerstätte near Davos (Canton Graubünden, Eastern Swiss Alps). *Eclogae Geologicae Helvetiae*, 88(3), pp.681–683.
- Gall, J.C. & Grauvogel-Stamm, L., 2005. The early Middle Triassic “Grès à Voltzia” Formation of eastern France: A model of environmental refugium. *Comptes Rendus - Palevol*, 4(6–7), pp.569–584.
- Gaudet, P. et al., 2011. Phylogenetic-based propagation of functional annotations within the Gene Ontology consortium. *Briefings in Bioinformatics*, 12(5), pp.449–462.
- Gharib, W.H. & Robinson-Rechavi, M., 2013. The branch-site test of positive selection is surprisingly robust but lacks power under synonymous substitution saturation and variation in GC. *Molecular Biology and Evolution*, 30(7), pp.1675–1686.
- Giorello, F.M. et al., 2014. Characterization of the kidney transcriptome of the South American olive mouse *Abrothrix olivacea*. *BMC genomics*, 15(1). Available at: <http://www.scopus.com/inward/record.url?eid=2-s2.0-84903899079&partnerID=tZOtx3y1>.
- Glasauer, S.M.K. & Neuhauss, S.C.F., 2014. Whole-genome duplication in teleost fishes and its evolutionary consequences. *Molecular Genetics and Genomics*, 289(6), pp.1045–1060.
- González-Castro, M. et al., 2016. Surfing among species, populations and morphotypes: Inferring boundaries between two species of new world silversides (Atherinopsidae). *Comptes Rendus - Biologies*, 339, pp.10–23.
- González Castro, M. et al., 2009. Fish composition in a south-western Atlantic temperate coastal lagoon: spatial–temporal variation and relationships with environmental variables. *Journal of the Marine Biological Association of the United Kingdom*, 89(3), pp.593–604. Available at: [http://www.journals.cambridge.org/abstract\\_S0025315409003002](http://www.journals.cambridge.org/abstract_S0025315409003002).
- Grabherr, M.G. et al., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature biotechnology*, 29(7), pp.644–652. Available at: <http://dx.doi.org/10.1038/nbt.1883>.
- Grosell, M., 2011. Intestinal anion exchange in marine teleosts is involved in osmoregulation and contributes to the oceanic inorganic carbon cycle. *Acta physiologica*, 202(3), pp.421–434.
- Grosell, M., 2010. The role of the gastrointestinal tract in salt and water balance. In *Fish Physiology*. Elsevier Inc., pp. 135–164. Available at: [http://dx.doi.org/10.1016/S1546-5098\(10\)03004-9](http://dx.doi.org/10.1016/S1546-5098(10)03004-9).
- Grosell, M. & Genz, J., 2006. Ouabain-sensitive bicarbonate secretion and acid absorption by the marine teleost fish intestine play a role in osmoregulation. *The*

- American Journal of Physiology*, 291, pp.R1145–R1156.
- Guidon, S. et al., 2004. Modeling the site-specific variation of selection patterns patterns Along Lineages. *Proceedings of the National Academy of Sciences of the USA*, 101(35), pp.12957–12962.
- Guinot, G. et al., 2013. Cretaceous stem chondrichthyans survived the end-Permian mass extinction. *Nature communications*, 4. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/24169620>.
- Haas, B.J. et al., 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols*, 8(8), pp.1494–1512. Available at: <http://dx.doi.org/10.1038/nprot.2013.084>.
- Haas, B.J. & Zody, M.C., 2010. Advancing RNA-Seq analysis. *Nature Biotechnology*, 28(5), pp.421–423. Available at: <http://www.nature.com/doi/10.1038/nbt0510-421>.
- Haig, D.W. et al., 2015. Early Triassic (early Olenekian) life in the interior of East Gondwana: Mixed marine-terrestrial biota from the Kockatea Shale, Western Australia. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 417, pp.511–533. Available at: <http://dx.doi.org/10.1016/j.palaeo.2014.10.015>.
- Havird, J.C. et al., 2016. Salinity-induced changes in gene expression from anterior and posterior gills of *Callinectes sapidus* (Crustacea: Portunidae) with implications for crustacean ecological genomics. *Comparative biochemistry and physiology. Part D, Genomics & proteomics*, 19, pp.34–44. Available at: <http://dx.doi.org/10.1016/j.cbd.2016.06.002>.
- Herrero, J. et al., 2016. Ensembl comparative genomics resources. *Database*, 2016.
- Hiroi, J. et al., 2008. Evidence for an apical Na–Cl cotransporter involved in ion uptake in a teleost fish. *Journal of Experimental Biology*, 211(16), pp.2584–2599. Available at: <http://jeb.biologists.org/content/211/16/2584>  
<http://jeb.biologists.org/content/211/16/2584.full.pdf>  
<http://jeb.biologists.org/content/211/16/2584.short>  
<http://www.ncbi.nlm.nih.gov/pubmed/18689412>.
- Holz, M., 2015. Mesozoic paleogeography and paleoclimates - A discussion of the diverse greenhouse and hothouse conditions of an alien world. *Journal of South American Earth Sciences*, 61, pp.91–107. Available at: <http://dx.doi.org/10.1016/j.jsames.2015.01.001>.
- Huang, X., Chen, X.-G. & Armbruster, P.A., 2016. Comparative performance of transcriptome assembly methods for non-model organisms. *BMC Genomics*, 17(1). Available at: <http://bmcbgenomics.biomedcentral.com/articles/10.1186/s12864-016-2923-8>  
<http://www.ncbi.nlm.nih.gov/pubmed/27464550>  
<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4964045>.
- Hull, P., 2015. Life in the aftermath of mass extinctions. *Current Biology*, 25(19), pp.R941–R952. Available at: <http://dx.doi.org/10.1016/j.cub.2015.08.053>.
- Hwang, P.-P. et al., 2011. Ion regulation in fish gills: recent progress in the cellular and molecular mechanisms. *American journal of physiology. Regulatory, integrative and comparative physiology*, 301(1), pp.R28–R47. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/21451143>.
- Hwang, P.P., 1987. Tolerance and ultrastructural responses of branchial chloride cells to salinity changes in the euryhaline teleost *Oreochromis mossambicus*. *Marine*

- Biology*, 94(4), pp.643–649.
- Ivanis, G., Esbaugh, A. & Perry, S.F., 2008. Branchial expression and localization of SLC9A2 and SLC9A3 sodium/hydrogen exchangers and their possible role in acid-base regulation in freshwater rainbow trout (*Oncorhynchus mykiss*). *The Journal of experimental biology*, 211(Pt 15), pp.2467–2477.
- Ivanov, A. V. et al., 2013. Siberian Traps large igneous province: Evidence for two flood basalt pulses around the Permo-Triassic boundary and in the Middle Triassic, and contemporaneous granitic magmatism. *Earth-Science Reviews*, 122, pp.58–76. Available at: <http://dx.doi.org/10.1016/j.earscirev.2013.04.001>.
- Jaillon, O. et al., 2004. Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. *Nature*, 431, pp.946–957.
- Kanehisa, M. et al., 2006. From genomics to chemical genomics: new developments in KEGG. *Nucleic acids research*, 34(Database issue), pp.D354–D357. Available at: [http://nar.oxfordjournals.org/cgi/content/abstract/34/suppl\\_1/D354](http://nar.oxfordjournals.org/cgi/content/abstract/34/suppl_1/D354).
- Kang, C.K. et al., 2010. Salinity-dependent expression of a Na<sup>+</sup>, K<sup>+</sup>, 2Cl<sup>-</sup> cotransporter in gills of the brackish medaka *Oryzias dancena*: A molecular correlate for hyposmoregulatory endurance. *Comparative Biochemistry and Physiology - A Molecular and Integrative Physiology*, 157(Part A), pp.7–18. Available at: <http://dx.doi.org/10.1016/j.cbpa.2010.05.013>.
- Kato, A. et al., 2011. Differential expression of Na<sup>+</sup>-Cl<sup>-</sup> cotransporter and Na<sup>+</sup>-K<sup>+</sup>-Cl<sup>-</sup> cotransporter 2 in the distal nephrons of euryhaline and seawater pufferfishes. *American journal of physiology. Regulatory, integrative and comparative physiology*, 300(2), pp.R284–R297. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/21084680>.
- Katoh, K. & Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, 30(4), pp.772–780.
- Knoll, A.H. et al., 2007. Paleophysiology and end-Permian mass extinction. *Earth and Planetary Science Letters*, 256(3–4), pp.295–313.
- Kosakovsky, S. et al., 2011. A random effects branch-site model for detecting episodic diversifying selection. *Molecular Biology and Evolution*, 28(11), pp.3033–3043.
- Kumai, Y. & Perry, S.F., 2012. Mechanisms and regulation of Na<sup>+</sup> uptake by freshwater fish. *Respiratory Physiology and Neurobiology*, 184(3), pp.249–256. Available at: <http://dx.doi.org/10.1016/j.resp.2012.06.009>.
- Kurita, Y. et al., 2008. Identification of intestinal bicarbonate transporters involved in formation of carbonate precipitates to stimulate water absorption in marine teleost fish. *The American Journal of Physiology, Integrative and Comparative Physiology*, pp.1402–1412.
- Lam, S.H. et al., 2014. Differential transcriptomic analyses revealed genes and signaling pathways involved in iono-osmoregulation and cellular remodeling in the gills of euryhaline Mozambique tilapia, *Oreochromis mossambicus*. *BMC genomics*, 15. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4213501&tool=pmcentrez&rendertype=abstract>.
- Lee, Y.-C. et al., 2011. Anion exchanger 1b, but not sodium-bicarbonate cotransporter 1b, plays a role in transport functions of zebrafish H<sup>+</sup>-ATPase-rich cells. *American journal of physiology. Cell physiology*, 300(2), pp.C295–C307. Available at:



- <http://www.ncbi.nlm.nih.gov/pubmed/21123736>.
- Liu, L. et al., 2012. Comparison of next-generation sequencing systems. *Journal of Biomedicine and Biotechnology*, 2012.
- Llompart, F.M. et al., 2013. Biological traits and growth patterns of pejerrey *Odontesthes argentinensis*. *Journal of Fish Biology*, 82(2), pp.458–474.
- Llompart, F.M., 2011. *La ictiofauna de Bahía San Blas (Provincia de Buenos Aires) y su relación con la dinámica de las pesquerías deportiva y artesanal*.
- Löytynoja, A. & Goldman, N., 2008. Phylogeny-aware gap placement prevents errors in sequence alignment and evolutionary analysis. *Science (New York, N.Y.)*, 320, pp.1632–1635. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/18566285>.
- Lu, A. & Guindon, S., 2013. Performance of standard and stochastic branch-site models for detecting positive selection among coding sequences. *Molecular Biology and Evolution*, 31(2), pp.484–495.
- Madsen, S.S., Kiilerich, P. & Tipsmark, C.K., 2009. Multiplicity of expression of Na<sup>+</sup>,K<sup>+</sup>-ATPase alpha-subunit isoforms in the gill of Atlantic salmon (*Salmo salar*): cellular localisation and absolute quantification in response to salinity change. *Journal of Experimental Biology*, 212(1), pp.78–88.
- Mannella, C.A., 1998. Conformational changes in the mitochondrial channel protein, VDAC, and their functional implications. *Journal of structural biology*, 121(2), pp.207–18. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/9615439>.
- Marioni, J.C. et al., 2008. RNA-seq: An assessment of technical reproducibility and comparison with gene expression arrays. *Genome Research*, 18(9), pp.1509–1517.
- Markova-Raina, P. & Petrov, D., 2011. High sensitivity to aligner and high rate of false positives in the estimates of positive selection in the 12 *Drosophila* genomes. *Genome Research*, 21(6), pp.863–874.
- Marshall, W.S. & Grosell, M., 2005. Ion Osmoregulation , and Acid – Base Balance. In *The Physiology of Fishes*. pp. 177–230.
- Martin, J. a. & Wang, Z., 2011. Next-generation transcriptome assembly. *Nature Reviews Genetics*, 12(10), pp.671–682. Available at: <http://www.nature.com/doi/10.1038/nrg3068><http://www.ncbi.nlm.nih.gov/pubmed/21897427>.
- Mayden, R.L. et al., 2009. Reconstructing the phylogenetic relationships of the earth's most diverse clade of freshwater fishes-order Cypriniformes (Actinopterygii: Ostariophysii): A case study using multiple nuclear loci and the mitochondrial genome. *Molecular Phylogenetics and Evolution*, 51(3), pp.500–514. Available at: <http://dx.doi.org/10.1016/j.ympev.2008.12.015>.
- McClure, M. et al., 2012. Purification of CFTR for mass spectrometry analysis: Identification of palmitoylation and other post-translational modifications. *Protein Engineering, Design and Selection*, 25(1), pp.7–14.
- Metzker, M.L., 2010. Sequencing technologies - the next generation. *Nature reviews. Genetics*, 11(1), pp.31–46. Available at: <http://dx.doi.org/10.1038/nrg2626>.
- Mi, H. et al., 2016. PANTHER version 10: Expanded protein families and functions, and analysis tools. *Nucleic Acids Research*, 44(D1), pp.D336–D342.
- Miall, A.D. et al., 2008. Chapter 9 The Western Interior Basin. In *Sedimentary Basins of the World*. Elsevier, pp. 329–362. Available at: <http://dx.doi.org/10.1016/S1874->

5997(08)00009-9.

- Mirande, J.M., 2016. Combined phylogeny of ray-finned fishes (Actinopterygii) and the use of morphological characters in large-scale analyses. *Cladistics*, pp.1–18.
- Miyazaki, H. et al., 1999. Molecular Cloning of CLC Chloride Channels in *Oreochromis Mossambicus* and Their Functional Complementation of Yeast CLC Gene Mutant. *Biochemical and Biophysical Research Communications*, 255(1), pp.175–181. Available at: <http://www.sciencedirect.com/science/article/pii/S0006291X99901663>.
- Moresco, A., 2006. *Pós-graduação em oceanografia biológica biologia reprodutiva do peixe-rei *Odontesthes argentinensis* (atherinopsidae), da região marinha lagoa dos patos.*
- Mozzi, A. et al., 2016. The evolutionary history of genes involved in spoken and written language: beyond FOXP2. *Scientific reports*, 6, p.22157. Available at: <http://www.nature.com/articles/srep22157%5Cnpapers3://publication/doi/10.1038/srep22157>.
- Murrell, B. et al., 2012. Detecting individual sites subject to episodic diversifying selection. *PLoS Genetics*, 8(7).
- Nakatani, M. et al., 2011. Evolutionary history of Otophysi (Teleostei), a major clade of the modern freshwater fishes: Pangaeen origin and Mesozoic radiation. *BMC evolutionary biology*, 11(117). Available at: <http://www.biomedcentral.com/1471-2148/11/177>.
- NCBI Resource Coordinators, 2015. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res*, 43(Database issue), pp.D6–D17. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/23193264>.
- Near, T.J. et al., 2013. Phylogeny and tempo of diversification in the superradiation of spiny-rayed fishes. , 110(31), pp.12738–12743.
- Near, T.J. et al., 2012. Resolution of ray-finned fish phylogeny and timing of diversification. *Proceedings of the National Academy of Sciences*, 109(34), pp.13698–13703.
- Nelson, D.L. & Cox, M.M., 2008. Water. In *Lehninger. Principles of Biochemistry Fifth Edition*. pp. 43–70.
- Nelson, J.S., 2006. *Fishes of the world* Fourth Edi., Edmonton: Sons, Inc.
- Nielsen, R. & Yang, Z., 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics*, 148(3), pp.929–936.
- Noletto, R.B. et al., 2009. Genome size evaluation in tetraodontiform fishes from the neotropical region. *Marine Biotechnology*, 11(6), pp.680–685.
- Norman, J.D., Ferguson, M.M. & Danzmann, R.G., 2014. An integrated transcriptomic and comparative genomic analysis of differential gene expression in Arctic charr (*Salvelinus alpinus*) following seawater exposure. *Journal of Experimental Biology*, 217(22), pp.4029–4042. Available at: <http://jeb.biologists.org/cgi/doi/10.1242/jeb.107441>.
- Oellermann, M. et al., 2015. Positive selection in octopus haemocyanin indicates functional links to temperature adaptation. *BMC Evolutionary Biology*, 15, pp.1–18. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/26142723>.
- Ozsolak, F. & Milos, P.M., 2011. RNA sequencing: advances, challenges and

- opportunities. *Nature Reviews Genetics*, 12(2), pp.87–98. Available at: <http://www.nature.com/doi/10.1038/nrg2934>.
- Pasquier, J. et al., 2016. Gene evolution and gene expression after whole genome duplication in fish: the PhyloFish database. *BMC genomics*, 17(1). Available at: <http://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-016-2709-z>.
- Paszkiwicz, K. & Studholme, D.J., 2010. De novo assembly of short sequence reads. *Briefings in Bioinformatics*, 10(5), pp.457–472.
- Pearson, W.R., 2013. An Introduction to Sequence Similarity (“Homology”) Searching. *Curr Protoc Bioinformatics*, 1(10). Available at: <http://edupediapublications.org/journals/index.php/ijr/article/view/964>.
- Pietsch, C. & Bottjer, D.J., 2014. The importance of oxygen for the disparate recovery patterns of the benthic macrofauna in the Early Triassic. *Earth-Science Reviews*, 137, pp.65–84. Available at: <http://dx.doi.org/10.1016/j.earscirev.2013.12.002>.
- Pohl, M. et al., 2015. Multigene phylogeny of cyprinodontiform fishes suggests continental radiations and a rogue taxon position of Pantanodon. *Vertebrate Zoology*, 65(1), pp.37–44. Available at: [www.senckenberg.de](http://www.senckenberg.de).
- Pond, S.L.K. et al., 2006. Automated phylogenetic detection of recombination using a genetic algorithm. *Molecular Biology and Evolution*, 23(10), pp.1891–1901.
- Porter, M.L., Pérez-Losada, M. & Crandall, K.A., 2005. Model-based multi-locus estimation of decapod phylogeny and divergence times. *Molecular Phylogenetics and Evolution*, 37(2), pp.355–369.
- Reimand, J. et al., 2007. g:Profiler—a web-based toolset for functional profiling of gene lists from large-scale experiments. *Nucleic Acids Research*, 35, pp.193–200.
- Ribas, L. & Piferrer, F., 2013. The zebrafish (*Danio rerio*) as a model organism, with emphasis on applications for finfish aquaculture research. *Reviews in Aquaculture*, 5.
- Rodrigue, N., Philippe, H. & Lartillot, N., 2010. Mutation-selection models of coding sequence evolution with site-heterogeneous amino acid fitness profiles. *Proceedings of the National Academy of Sciences*, 107(10), pp.4629–4634. Available at: <http://www.pnas.org/content/107/10/4629.abstract>.
- Romano, C. et al., 2016. Permian-Triassic Osteichthyes (bony fishes): Diversity dynamics and body size evolution. *Biological Reviews*, 91(1), pp.106–147.
- Roux, J. et al., 2014. Patterns of positive selection in seven ant genomes. *Molecular Biology and Evolution*, 31(7), pp.1661–1685.
- Rubin, A.F. & Green, P., 2009. Mutation patterns in cancer genomes. *Proceedings of the National Academy of Sciences of the United States of America*, 106(51), pp.21766–21770. Available at: <http://www.pnas.org/content/106/51/21766.abstract>.
- Sampaio, L.A., 2006. Production of “pejerrey” *Odontesthes argentinensis* fingerlings: A review of current techniques. *Biocell*, 30(1), pp.121–123.
- Schredelseker, J. et al., 2014. High resolution structure and double electron-electron resonance of the zebrafish voltage-dependent anion channel 2 reveal an oligomeric population. *Journal of Biological Chemistry*, 289(18), pp.12566–12577.
- Setiamarga, D.H.E. et al., 2008. Interrelationships of Atherinomorpha (medakas, flyingfishes, killifishes, silversides, and their relatives): The first evidence based on whole mitogenome sequences. *Molecular Phylogenetics and Evolution*, 49(2),

- pp.598–605. Available at: <http://dx.doi.org/10.1016/j.ympbev.2008.08.008>.
- Seton, M. et al., 2012. Global continental and ocean basin reconstructions since 200Ma. *Earth-Science Reviews*, 113(3–4), pp.212–270. Available at: <http://dx.doi.org/10.1016/j.earscirev.2012.03.002>.
- Shen, J. et al., 2016. Two pulses of oceanic environmental disturbance during the Permian-Triassic boundary crisis. *Earth and Planetary Science Letters*, 443, pp.139–152.
- Shenoy, S.K. & Lefkowitz, R.J., 2005. Angiotensin II-stimulated signaling through G proteins and beta-arrestin. *Science's STKE : signal transduction knowledge environment*, 2005(311). Available at: <http://www.ncbi.nlm.nih.gov/pubmed/16304060>.
- Shimizu, H. et al., 2015. Mitochondrial Ca<sup>2+</sup> uptake by the voltage-dependent anion channel 2 regulates cardiac rhythmicity. *eLife Sciences*, 4(4).
- Shoshan-Barmatz, V. et al., 2010. VDAC, a multi-functional mitochondrial protein regulating cell life and death. *Molecular Aspects of Medicine*, 31(3), pp.227–285. Available at: <http://dx.doi.org/10.1016/j.mam.2010.03.002>.
- Somoza, G.M. et al., 2008. Historical aspects, current status and prospects of pejerrey aquaculture in South America. *Aquaculture Research*, 39(7), pp.784–793.
- Song, H. et al., 2014. Anoxia/high temperature double whammy during the Permian-Triassic marine crisis and its aftermath. *Scientific reports*, 4(4), p.4132. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3928575&tool=pmcentrez&rendertype=abstract>.
- Sparks, J.S. & Smith, W.L., 2004. Phylogeny and biogeography of cichlid fishes (Teleostei: Perciformes: Cichlidae). *Cladistics*, 20, pp.501–517.
- Stager, M. et al., 2014. Signatures of natural selection in the mitochondrial genomes of Tachycineta swallows and their implications for latitudinal patterns of the “pace of life.” *Gene*, 546(1), pp.104–111. Available at: <http://dx.doi.org/10.1016/j.gene.2014.05.019>.
- Strüssmann, C.A. et al., 2010. Implications of climate change for the reproductive capacity and survival of New World silversides (family Atherinopsidae). *Journal of Fish Biology*, 77(8), pp.1818–1834.
- Strüssmann, C. a. et al., 1996. Temperature effects on sex differentiation of two South American atherinids, *Odontesthes argentinensis* and *Patagonina hatcheri*. *Environmental Biology of Fishes*, 47(2), pp.143–154.
- Sun, Y. et al., 2016. Fish-T1K (Transcriptomes of 1,000 Fishes) Project: large-scale transcriptome data for fish evolution studies. *GigaScience*, 5(18). Available at: <http://gigascience.biomedcentral.com/articles/10.1186/s13742-016-0124-7>.
- Tagliacollo, V.A. et al., 2016. Model-based total evidence phylogeny of Neotropical electric knifefishes (Teleostei, Gymnotiformes). *Molecular Phylogenetics and Evolution*, 95, pp.20–33. Available at: <http://dx.doi.org/10.1016/j.ympbev.2015.11.007>.
- Takei, Y. et al., 2014. Diverse mechanisms for body fluid regulation in teleost fishes. *American journal of physiology: Regulatory, Integrative, & Comparative Physiology*, 307(7), pp.R778–R792. Available at: <http://ajpregu.physiology.org/content/307/7/R778>.

- Tamura, K. et al., 2013. MEGA6: Molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution*, 30(12), pp.2725–2729.
- Tang, C.H. et al., 2009. Relative changes in the abundance of branchial Na<sup>+</sup>/K<sup>+</sup>-ATPase alfa-isoform-like proteins in marine euryhaline milkfish (*Chanos chanos*) acclimated to environments of different salinities. *Journal of Experimental Zoology*, 311(7), pp.522–530.
- Tang, C.H. & Lee, T.H., 2007. The effect of environmental salinity on the protein expression of Na<sup>+</sup>/K<sup>+</sup>-ATPase, Na<sup>+</sup>/K<sup>+</sup>/2Cl<sup>-</sup> cotransporter, cystic fibrosis transmembrane conductance regulator, anion exchanger 1, and chloride channel 3 in gills of a euryhaline teleost, *Tetraodon nigrovir*. *Comparative Biochemistry and Physiology - A Molecular and Integrative Physiology*, 147(Part A), pp.521–528.
- Taylor, P. et al., 2011. A brief utilization report on the Illumina HiSeq 2000 sequencer. *Mycology*, 2(3), pp.169–191. Available at: <http://www.tandfonline.com/doi/abs/10.1080/21501203.2011.615871>.
- Teranishi, K., Mekuchi, M. & Kaneko, T., 2013. Expression of sodium/hydrogen exchanger 3 and cation-chloride cotransporters in the kidney of Japanese eel acclimated to a wide range of salinities. *Comparative Biochemistry and Physiology - A Molecular and Integrative Physiology*, 164(2), pp.333–343. Available at: <http://dx.doi.org/10.1016/j.cbpa.2012.11.011>.
- Thomas, T., Gilbert, J. & Meyer, F., 2012. Metagenomics - a guide from sampling to data analysis. *Microbial Informatics and Experimentation*, 2(3). Available at: <http://www.microbialinformaticsj.com/content/2/1/3>.
- Ticino, M.C. & Bürgin, T., 1995. Actinopterygian fishes (Osteichthyes, Actinopterygii) from the Kalkschieferzone Uppermost Ladinian) near Meride (Canton Ticino, Southern Switzerland). *Birkhäuser Verlag. Basel*, 88(3), pp.803–826.
- Tine, M. et al., 2011. Salinity-related variation in gene expression in wild populations of the black-chinned tilapia from various West African coastal marine, estuarine and freshwater habitats. *Estuarine, Coastal and Shelf Science*, 91(1), pp.102–109. Available at: <http://dx.doi.org/10.1016/j.ecss.2010.10.015>.
- Tine, M. et al., 2008. Transcriptional responses of the black-chinned tilapia *Sarotherodon melanotheron* to salinity extremes. *Marine Genomics*, 1(2), pp.37–46.
- Tintori, A., 1990. The Actinopterygian fish *Prohalecites* from the Triassic of Ca' del Frate Itali. *Palaeontology*, 33(1), pp.155–174.
- Tipsmark, C.K. et al., 2008. Osmoregulation and expression of ion transport proteins and putative claudins in the gill of Southern Flounder (*Paralichthys lethostigma*). *Comparative Biochemistry and Physiology - A Molecular and Integrative Physiology*, 150(3), pp.265–273.
- Tsagkogeorga, G. et al., 2015. A phylogenomic analysis of the role and timing of molecular adaptation in the aquatic transition of cetartiodactyl mammals. *Royal Society open science*, 2(9). Available at: <http://rsos.royalsocietypublishing.org/content/2/9/150156.abstract>.
- Tsang, L.M. et al., 2008. Phylogeny of Decapoda using two nuclear protein-coding genes: Origin and evolution of the Reptantia. *Molecular Phylogenetics and Evolution*, 48(1), pp.359–368.
- Tse, W.K.F., Au, D.W.T. & Wong, C.K.C., 2006. Characterization of ion channel and transporter mRNA expressions in isolated gill chloride and pavement cells of seawater acclimating eels. *Biochemical and Biophysical Research*

- Communications*, 346(4), pp.1181–1190.
- Via, S., 2009. Natural selection in action during speciation. *Proceedings of the National Academy of Sciences of the United States of America*, 106 Suppl, pp.9939–9946.
- Wang, Y. et al., 2000. Cn3D: Sequence and structure views for Entrez. *Trends in Biochemical Sciences*, 25(6), pp.300–302.
- Wang, Z., Gerstein, M. & Snyder, M., 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nature reviews. Genetics*, 10(1), pp.57–63.
- Ward, N. & Moreno-Hagelsieb, G., 2014. Quickly finding orthologs as reciprocal best hits with BLAT, LAST, and UBLAST: How much do we miss? *PLoS ONE*, 9(7).
- Watanabe, S. et al., 2011. Electroneutral cation-Cl<sup>-</sup> cotransporters NKCC2beta and NCCbeta expressed in the intestinal tract of Japanese eel *Anguilla japonica*. *Comparative Biochemistry and Physiology - A Molecular and Integrative Physiology*, 159(4), pp.427–435. Available at: <http://dx.doi.org/10.1016/j.cbpa.2011.04.009>.
- Weissing, F.J., Edelaar, P. & van Doorn, G.S., 2011. Adaptive speciation theory: A conceptual review. *Behavioral Ecology and Sociobiology*, 65(3), pp.461–480.
- Wilson, M.V.H. et al., 1992. Paleontological Society Cretaceous Esocoides (Teleostei): Early Radiation of the Pikes in North American Fresh Waters. *Journal of Paleontology*, 66(5), pp.839–846.
- Wilson, R. et al., 1996. Intestinal base excretion in the seawater-adapted rainbow trout: a role in acid-base balance? *The Journal of experimental biology*, 199, pp.2331–2343.
- Winguth, C. & Winguth, A.M.E., 2012. Simulating Permian-Triassic oceanic anoxia distribution: Implications for species extinction and recovery. *Geology*, 40(2), pp.127–130.
- Xu, B. & Yang, Z., 2013. PamIX: A graphical user interface for PAML. *Molecular Biology and Evolution*, 30(12), pp.2723–2724.
- Xu, Q. et al., 2016. Analysis of the erythropoietin of a Tibetan Plateau schizothoracine fish (*Gymnocypris dobula*) reveals enhanced cytoprotection function in hypoxic environments. *BMC Evolutionary Biology*, 16(11). Available at: <http://www.biomedcentral.com/1471-2148/16/11>.
- Yang, Z., 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Molecular Biology and Evolution*, 15(5), pp.568–573. Available at: <http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=9580986&retmode=ref&cmd=prlinks%5Cnpapers3://publication/uuid/A5AE74C0-B088-4F85-96F1-A50AA0D91191>.
- Yang, Z., 2000. Maximum likelihood estimation on large phylogenies and analysis of adaptive evolution in human influenza virus A. *Journal of molecular evolution*, 51(5), pp.423–432.
- Yang, Z., 2007. PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, 24(8), pp.1586–1591.
- Yang, Z. & Nielsen, R., 2002. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Molecular Biology and Evolution*, 19(6), pp.908–917. Available at:

<http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=12032247&retmode=ref&cmd=prlinks%5Cnpapers3://publication/uuid/C5C0A1D1-15D5-4FDD-9C5A-9F62E5EB0E59>.

Yang, Z., Wong, W.S.W. & Nielsen, R., 2005. Bayes empirical Bayes inference of amino acid sites under positive selection. *Molecular Biology and Evolution*, 22(4), pp.1107–1118.

Zhang, Z. & Chen, J., 2016. Atomic Structure of the Cystic Fibrosis Transmembrane Conductance Regulator. *Cell*, 167(6), pp.1586–1597.

Zhang, J. et al., 2011. The impact of next-generation sequencing on genomics. *Journal of Genetics*, 38(3), pp.95–109.

Zonneveld, J.-P., Gingras, M.K. & Beatty, T.W., 2010. Diverse Ichnofossil Assemblages Following the P-T Mass Extinction, Lower Triassic, Alberta and British Columbia, Canada: Evidence for Shallow Marine Refugia on the Northwestern Coast of Pangaea. *Palaio*, 25(6), pp.368–392. Available at: <http://palaio.sepmonline.org/cgi/doi/10.2110/palo.2009.p09-135r>.

Zou, M. et al., 2012. Integrating multi-origin expression data improves the resolution of deep phylogeny of ray-finned fish (Actinopterygii). *Scientific Reports*, 2(625).