UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE CIENCIAS BIOLÓGICAS

TESIS DOCTORAL

Aplicación de técnicas moleculares en biología evolutiva: los peces continentales de Europa y México como caso de estudio

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

Diuski Keri Corona Santiago

Director

José Ignacio Doadrio Villarejo

Madrid, 2018

UNIVERSIDAD COMPLUTENSE DE MADRID

Facultad de Ciencias Biológicas

Aplicación de técnicas moleculares en biología evolutiva: Los peces continentales de Europa y México como caso de estudio

Memoria presentada para optar al grado de Doctor por

Diushi Keri Corona Santiago

Bajo la dirección del Doctor
José Ignacio Doadrio Villarejo

Madrid – Abril 2017

MINISTERIO DE
ECONOMÍA Y
COMPETITIVIDAD

CSIC
CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS

mncn
museonacionaldecienciasnaturales

José Ignacio Doadrio Villarejo Profesor en Investigación del Consejo Superior de Investigaciones Científicas.

Certifica

Que la presente memoria titulada **Aplicación de técnicas moleculares en biología evolutiva: Los peces continentales de Europa y México como caso de estudio**, que para optar el grado de Doctor presenta Diushi Keri Corona Santiago, ha sido realizada bajo mi dirección en el Departamento de Biodiversidad y Biología Evolutiva del Museo Nacional de Ciencias Naturales – CSIC (MADRID). Esta memoria está además adscrita académicamente a la Facultad de Ciencias Biológicas de la Universidad Complutense de Madrid. Considerando que representa trabajo suficiente para constituir una Tesis Doctoral, autorizamos su presentación.

Y para que así conste, firmamos el presente certificado

El director:

José Ignacio Doadrio Villarejo
Profesor en Investigación
Museo Nacional de Ciencias Naturales-CSIC

El doctorando:

Diushi Keri Corona Santiago

C/ JOSÉ GUTIERREZ ABASCAL,2
28006 MADRID ESPAÑA
TEL:91.411.13.28
Ext.
FAX:

Índice

Resumen

Los ecosistemas dulceacuícolas mantienen alrededor del 12% de la fauna animal del planeta. El estudio de las relaciones filogenéticas, a través de marcadores moleculares, de las especies permite inferir los procesos históricos, tanto geológicos como climáticos, y sus efectos en la biodiversidad. Por tanto, los estudios en biología evolutiva proveen información esencial para comprender los cambios pasados y futuros en la ictiofauna continental y sus hábitats. Esta tesis doctoral tiene como objetivo general el aplicar técnicas moleculares para responder a diversas preguntas en biología evolutiva de peces continentales. Concretamente: a) la descripción de la historia evolutiva de tres complejos de especies de peces continentales de México y la Península Ibérica; b) un estudio macroevolutivo sobre la evolución del viviparismo en el orden Cyprinodontiformes y los cambios estructurales y funcionales asociados al origen de esta novedad evolutiva; c) aportar nuevos criterios para la selección de marcadores mitocondriales para futuros estudios filogenéticos y d) describir nuevos genomas mitocondriales. Esta tesis comprende un conjunto de estudios realizados entre Octubre de 2013 y Marzo del 2017.

Tres complejos de especies fueron estudiados que presentan peculiaridades evolutivas, estrechas relaciones filogenéticas y taxonomía incierta. De México, se estudió el complejo de especies vivíparas *Allotoca diazi,* y el complejo de especies tetraploides *Catostomus plebeius-nebuliferus*. De la Península Ibérica, el nuevo complejo de origen hibridogenético *Squalius* sp. Para ello se utilizaron distintos marcadores moleculares tanto mitocondriales como nucleares, desde genes que pueden resolver relaciones interespecíficas, hasta regiones hipervariables como microsatélites que puedes dar información a nivel de poblaciones e inclusive de individuos. La información genética se integró con herramientas analíticas para describir patrones demográficos, de diversidad genética, estimar tiempos de divergencia, reconstruir áreas ancestrales, y modelizar el nicho ecológico. Finalmente, se aportaron criterios para la conservación para cada uno de los complejos de especies estudiados.

De acuerdo a los objetivos propuestos en esta tesis, la descripción de la historia evolutiva de los complejos *Allotoca diazi*, *Catostomus plebeius-nebuliferus* y

Squalius sp, deja en claro la necesidad de tomar en cuenta cuestiones básicas en biología como es la definición de *especie*, dado que en los tres casos, es difícil acomodar los linajes identificados genéticamente a cualquiera de las definiciones explicitas o implícitas de especie. Los conflictos entre conceptos de especie son observables en los primeros tres capítulos, y aparecen básicamente al definir criterios operativos para identificar estos taxa, incluso cuando se utilizan conceptos de índole evolutiva.

El estudio de la evolución de la viviparidad en Cyprinodontiformes basado en mitogenomas, permitió determinar y corroborar varias hipótesis. Nuestros resultados apoyan la hipótesis de que la viviparidad en este orden tiene como ancestro el oviparismo, pero que el ovoviviparismo no constituye un paso intermedio hacia el viviparismo. Este tipo de cuestiones evolutivas pueden ser abordadas a partir de la obtención de una gran cantidad de datos, y la búsqueda de herramientas analíticas que nos permitan explicar las presiones selectivas o los cambios estructurales en el genoma mitocondrial asociados al surgimiento de novedades evolutivas.

A su vez, el capítulo 4, aportó información acerca de la importancia e implicaciones que tiene el determinar y seleccionar marcadores moleculares adecuados en filogenias a grandes niveles taxonómicos. Así como determinar cuantos y qué genes son necesarios para resolver una filogenia.

Finalmente se describieron dos mitogenomas, uno de ellos perteneciente a *Luciobarbus rifensis*, un endemismo del norte de África que fue descrito recientemente, y el otro, de *Xenotoca variata* el cual fue utilizado para desarrollar el estudio sobre la evolución de la viviparidad. Estos trabajos significaron un primer acercamiento a la aplicación de datos genómicos y un primer paso para la exploración e identificación de nuevos marcadores moleculares que fueron desarrollados en el capítulo 4.

La presente tesis doctoral soporta las bases tradicionales en filogeografía y filogenética a través de la interpretación cuantitativa de cada genealogía y la información referente a la diversidad genética para el estudio de las poblaciones. Así mismo, aporta información consistente con trabajos previos sobre temas de macroevolución y evolución molecular proponiendo que marcadores moleculares pueden ser utilizados a la luz de diversos supuestos y preguntas evolutivas.

Abstract

Freshwater ecosystems maintain about 12% of the animal fauna of the planet. The study of phylogenetic relationships through molecular markers of the species allows inferring historical processes, both geological and climatic, and their effects on biodiversity. Therefore, studies in evolutionary biology provide essential information for the understanding of current and future changes in the continental ichthyofauna and their habitats. Therefore, this doctoral thesis has as general objective to apply molecular techniques to solve questions in evolutionary biology in fish. Specifically the description of the evolutionary history of three complexes of continental fish species in Mexico and the Iberian Peninsula, as well as a macroevolutionary study on the evolution of viviparism in the order Cyprinodontiformes and the structural and functional changes associated to the origin of this evolutionary novelty, to provide new criteria for the selection of mitochondrial markers for future phylogenetic studies, and to describe mitochondrial genomes. This thesis includes a set of studies conducted during October 2013 to March 2017.

Three species complexes were studied that show evolutionary peculiarities and close phylogenetic relationships and taxonomic controversies. From Mexico, the complex of viviparous species *Allotoca diazi*, and the complex of tetraploid species *Catostomus plebeius-nebuliferus*. From the Iberian peninsula, the new complex of hybridogenic origin *Squalius* sp. For this studies, different molecular markers were used both mitochondrial and nuclear, from genes that are considered to be able to solve interspecific relations, to hypervariable regions such as microsatellites that can give information at the level of populations and even individuals. Genetic information was integrated with analytical tools to describe demographic patterns, genetic diversity, time divergence estimation, ancestral area reconstruction, and ecological niche modeling. Finally, conservation criteria were provided for each of the species complexes studied.

According to the objectives proposed in this thesis, the description of the evolutionary history of the complexes *Allotoca diazi*, *Catostomus plebeius-nebuliferus* and *Squalius* sp, makes clear the need to take into account basic questions in biology

such as the definition of species, given that in all three cases it is difficult to fit the genetically identified lineages to any of the explicit or implicit definitions. Conflicts between concepts of species are observable in the first three chapters, and appear basically when defining operational criteria to identify these taxa even when using concepts of an evolutionary nature.

The study of the evolution of viviparity in Cyprinodontiformes based on mitogenomas, allowed to determine and corroborate several hypotheses such as the one that explains that viviparity has a common ancestor viviparity, whereas ovoviviparismo is not an intermediate step toward viviparity. This type of evolutionary questioning can be approached from obtaining a large amount of data, as well as the search for analytical tools to explain specific issues such as selective pressures or structural changes in the mitochondrial genome associated with the emergence of evolutionary novelties.

Chapter 4, in turn, provided information about the implications of determining and selecting molecular markers in phylogenies at high taxonomic levels, what are the most important factors and what and how many genes are required to solve a phylogeny of this type.

Finally, two mitogenomas were described, one of them belonging to *Luciobarbus rifensis*, an endemism from North Africa that was recently described, and the other from *Xenotoca variata* which was used to develop the study on the evolution of viviparity. These works meant a first approach to the application of genomic data and a first step for the exploration and identification of new molecular as was investigated in Chapter 4.

The present PhD thesis supports the traditional bases in phylogeography and phylogenetics through the quantitative interpretation of each genealogy and the information concerning genetic diversity for the study of the populations, as well as provides consistent information with previous works on subjects in macroevolution and molecular evolution suggesting that molecular markers can be used in light of diverse bases and evolutionary questions.

Introducción General

Los conocimientos en el estudio de la diversidad y clasificación de los peces continentales, han aumentado rápidamente debido al desarrollo de las técnicas moleculares y su continuo avance metodológico que permite la realización de estudios basados en un número mayor y más complejo de datos (Betancur *et al.,* 2013). Estos avances han permitido el mejor conocimiento y descripción de la historia evolutiva de diversos grupos taxonómicos en el mundo y su relaciones filogenéticas además de permitir identificar nuevos linajes evolutivos independientes y determinar especies crípticas, además de caracterizar genéticamente las poblaciones de un grupo taxonómico y sus cambios a través del tiempo (Helfman *et al.,* 2009).

Como fue predicho por Greenwood *et al.,* (1966), el descubrimiento y desarrollo de los marcadores moleculares, especialmente de secuencias de genes mitocondriales, abrió una nueva ventana al estudio de la evolución de peces proporcionando un criterio común de información filogenética a través de vastos niveles taxonómicos (Miya *et al.,* 2003; Betancur *et al.,* 2013). En la última década el desarrollo de estas técnicas han permitido el uso de aproximaciones integrativas combinando marcadores moleculares y datos morfológicos (Stepien y Kocher, 1997). Así como, la aplicación de aproximaciones multi-loci a través de secuenciación de nueva generación (McCormack *et al.,* 2013). El incremento en la disponibilidad de marcadores moleculares, con diferentes características, para atender distintas cuestiones en biología evolutiva en especies no modelo ha asumido procesos cada vez más complejos para generar datos (Edwards, 2008). Esta información es destinada a su vez para resolver y explicar procesos evolutivos aun más complejos en que los datos morfológicos o los proporcionados por u sólo gen mitocondrial no aportan suficiente información. Este es el caso de estudios filogeográficos, de genética de poblaciones o de procesos evolutivos en los que la hibridación está involucrada (Ekblom y Galindo, 2010).

La filogeografía comúnmente se ha utilizado como una herramienta para esclarecer patrones evolutivos a nivel intraespecífico. Sin embargo, en la última década se ha utilizado para describir procesos demográficos históricos como: flujo génico, tamaño efectivo poblacional, procesos y secuencias de colonización, cuellos de botella y también para determinar barreras entre especies y los procesos que las conforman (McCormack *et al.,* 2013). Por otro lado, la filogeografía tiene gran importancia en la conservación, puesto que permite reconocer la estructura genética de los organismos e identificar unidades de conservación (Avise *et al.,* 2000; Freeland, 2005). Es por ello que el beneficio de la inclusión de múltiple información genética debe conducir a una "verdadera señal filogenética común" que muestre la evolución de la biota y no la de una región de ADN en particular (Comas *et al.,* 2007).

Los ecosistemas dulceacuícolas mantienen alrededor del 12% de la fauna animal del planeta (Brucet *et al.,* 2013). La causa principal es el aislamiento de su hábitat por barreras geográficas que ha generado alta diversidad de especies aunque también ha aumentando el riesgo de extinción en un mundo cambiante (Olden *et al.,* 2010). La diversidad de los organismos acuáticos se reduce alarmantemente y sin precedentes debido a procesos como pérdida y fragmentación de hábitat, alteración hidrológica, cambio climático, sobreexplotación, contaminación e introducción de especies exóticas (Revenga y Echeverria, 2001; Dudgeon *et al.,* 2006). Es por ello que la determinación de los procesos involucrados en el origen y evolución de la biodiversidad y su modificación a través del tiempo, es crucial para la aplicación de eficientes estrategias de conservación de estos ecosistemas (Kerr, Kharouba y Currie, 2007). El estudio de las relaciones filogenéticas permite inferir los procesos históricos, tanto geológicos como climáticos, y sus efectos en la biodiversidad (Oberprieler, 2005). De esta forma, los estudios de biología evolutiva a través de marcadores moleculares proveen información esencial para el entendimiento de los cambios actuales y futuros en la ictiofauna continental y sus hábitats.. Sin embargo, no todas los preguntas en biología evolutiva son fáciles de abordar puesto que las especies de animales no siguen un patrón evolutivo único y éstas no se originan por un proceso histórico específico (Olden *et al.,* 2010).

Marcadores moleculares: de la genética a la genómica

La examinación de marcadores moleculares con "apropiada" tasa de cambio y señal filogenética, permite obtener información acerca de casi cualquier población y de los procesos evolutivos a través del tiempo, independientemente de que una variante genética sea influenciada en el futuro por la biología, migración, tamaño efectivo, etc (Sunnucks, 2000). Pero, no todos los genes o fragmentos de ADN son idóneos para ser usados en las reconstrucciones filogenéticas, y no todos los marcadores son útiles para el análisis genético de un grupo dado de organismos (Patwardhan, Ray y Roy, 2014). El método para conocer la capacidad de resolver relaciones evolutivas de una secuencia de ADN dentro de un particular grupo taxonómico, incluye el estudio de la capacidad del gen de recuperar relaciones filogenéticas bien establecidas dentro de un clado de edad similar (Graybeal, 1994). Sin embargo, la selección de nuevos marcadores moleculares depende de factores como la intensidad de saturación, ambigüedades en los alineamientos, debidos a nivel de divergencia entre taxa y ligado al modelo de substitución nucleotídica estimado, longitud de la secuencia, número de sitios variables, tamaño de muestra y la existencia de presiones selectivas (Meiklejohn *et al.,* 2014).

El marcador mitocondrial más extensamente usado en los estudios evolutivos de peces es el gen Citocromo b *cytb*, que tiene generalmente una longitud de 1140 pares de bases. Se ha reportado del *cytb* su capacidad para recuperar relaciones filogenéticas entre taxa estrechamente relacionados pero que pierde resolución en los nodos profundos de las filogénias (Patwardhan *et al.,* 2014). En el caso de marcadores nucleares, el uso depende del nivel al que el estudio se desarrolle y el nivel de profundidad taxonómica en el que se lleva a cabo. Los microsatelites, repeticiones de bases en tándem en loci hipervariables son muy utilizados en estudios de taxa estrechamente relacionados o en genética de poblaciones (Ellegren, 2000).

Las ventajas de las aproximaciones utilizando muchos genes frente a aquellas basadas en un solo gen son *a priori* evidentes, y se podrían evitar las historias evolutivas basadas en un único locus a favor de una “verdadera” señal filogenética común. Sin embargo, en la práctica, la inclusión de muchos genes para compensar la falta de tamaño de muestra normalmente no resuelve los problemas que puede observarse cuando se utiliza un único gen (Comas *et al.,* 2007). Aún así, la secuenciación de nueva generación abrió una nueva ventana en la última década para inferir la historia de poblaciones y especies como base de nuevos estudios filogenéticos y filogeográficos), permitiendo aproximaciones multilocus con sets de datos aún mayores (Edwards 2008; Thomson *et al.,* 2010; McCormack *et al.,* 2013). Sin embargo, la dificultad de la aplicación de estas herramientas en filogenética y filogeografía reside principalmente en el lento desarrollo de las técnicas para especies no modelo. Así como, la necesidad de gran numero de muestras por especies, la ausencia de protocolos consensuados para la preparación de librerías para un tipo de estudio en particular, la disponibilidad de recursos informáticos capaces de manejar esa cantidad de datos, y el rápido cambio de las tecnologías de secuenciación (McCormack *et al.,* 2013).

Tomando en cuenta todas las ventajas que los marcadores moleculares poseen, la presente tesis doctoral estuvo enfocada en realizar estudios aplicando diversas técnicas moleculares con el fin de responder a diferentes cuestiones evolutivas en grupos de peces que viven en América y en Europa. Los primeros 3 capítulos están destinados a describir la historia evolutiva de tres "complejos" de especies. Estos complejos tienen distintas características, y el término complejo sólo es usado porque el origen y evolución de una especie está involucrada y permanece sin resolver. Sin embargo, en los tres trabajos los complejos de especies se suscitan en diferentes espacios y tiempos, involucrando procesos evolutivos e históricos diferentes desde el final del Mioceno hasta el presente. Además, cada complejo representa un modelo de estudio bastante singular desde el punto de vista evolutivo. El complejo *Allotoca diazi* (Meek, 1902), presentado en el primer capítulo, es el resultado de posibles translocaciones durante los últimos 2.000 años y los procesos geológicos y climáticos Pleisto-Holocénicos en el centro de México. El segundo capítulo de esta tesis tiene como objeto resolver la historia evolutiva del complejo *Catostomus plebeius-nebuliferus* el cual está configurado por los procesos climáticos y geológicos del oeste de México durante el Plio-Pleistoceno. Finalmente el tercer capítulo hace referencia la complejo complejo *Squalius* sp endémico de la Península Ibérica. Este constituye un complejo poliploide híbrido con estrategias reproductivas peculiares como la hibridogénesis que se configuró posiblemente durante el final del Mioceno y el Plioceno.

El cuarto capítulo fue generado a raíz del capítulo 6, en el cuál se obtuvo el genoma mitocondrial completo de una especie vivípara *Xenotoca variata* (Bean, 1887) perteneciente a la subfamilia Goodeinae endémica del centro de México

(Domínguez-Domínguez *et al.,* 2012). La obtención tanto del genoma mitocondrial como de secuencias nucleares obtenidas a través de secuenciación masiva, estuvo encaminado a futuros estudios genéticos del género *Xenotoca* Hubbs y Turner, 1939 que aparece como polifilética en estudios filogenéticos previos (Doadrio y Domínguez, 2004). Una vez obtenido el genoma mitocondrial de *X. variata* se realizó un estudio de relaciones filogenéticas a nivel familiar con otros mitogenomas disponibles del orden Cyprinodontiformes, cuyas relaciones filogenéticas aún no estaban esclarecidas. Además, el estudio mitogenómico permitió conocer procesos de selección en los genes mitocondriales del orden y asociar dichos procesos al origen y evolución del viviparismo en éste grupo de peces.

Los cuatro primeros capítulos tienen en común que abordan preguntas evolutivas sobre el origen y evolución de grupos taxonómicos concretos con diferentes características morfológicas y genéticas como el viviparismo (complejo *Allotoca diazi*), la polipoidización (complejo *Catostomus plebeius-nebuliferus*), la hibridación (complejo *Squalius* sp.) y rasgos generales en la evolución de los vertebrados como lo es el viviparismo desde un punto de vista genómico. A través de esto, podemos observar cómo distintas herramientas moleculares y su manejo a través de diversos recursos bioinformáticos, son capaces de explicar y reconstruir la historia evolutiva de distintos grupos de peces. Además, permiten conocer si a través de unos pocos marcadores moleculares se pueden responder de la misma forma que con muchos marcadores a preguntas evolutivas generales y obtener resultados robustos.

Los capítulos 5 y 6 conforman trabajos en los que se incursiona en la secuenciación masiva para obtener loci idóneos para el estudio de las especies a las que pertenecen: *Lucobarbus rifensis* Doadrio, Casal-López y Yahyaoui, 2015 un endemismo del norte de África descrito recientemente, y *Xenotoca variata*, un endemismo mexicano como se mencionó anteriormente. En dichos capítulos se describe el mitogenoma de ambas especies como primer paso para obtener marcadores moleculares mitocondriales, mediante la separación de éste genoma del genoma parcial nuclear secuenciado a la vez. La generación de este tipo de marcadores moleculares ayudará a los futuros estudios genéticos y evolutivos de las especies.

Complejo Allotoca diazi (Meek, 1902)

Consiste en tres especies vivíparas estrechamente relacionadas genética y morfológicamente las cuales se encuentran en una fuerte controversia taxonómica (Doadrio y Domínguez, 2004; Domínguez-Domínguez *et al.,* 2012). Forman parte del género más amplio de la subfamilia Goodeinae (Goodeidae), endémica de México, la cual ha sido objeto de numerosos estudios evolutivos y ecológicos debido a sus características únicas de reproducción y desarrollo-nutrición embrionario (Hubbs y Turner, 1937). La primera de ellas es *Allotoca diazi* (Meek, 1902), especie que se encuentra restringida al lago de Pátzcuaro, uno de los mayores lagos de México. La segunda es *A. meeki* (Álvarez, 1959) endémica de la cuenca del lago Zirahuén y que actualmente sólo se encuentra en el manantial de Opopeo y tributarios del mismo, y por último, *A. catarinae* (De Buen, 1942), restringida al río Cupatitzio en la cuenca del Balsas, una de las más grandes de México. Su actual distribución se ha tratado de explicar a través de dos hipótesis biogeográficas acerca de la conexión y desconexión de los cuerpos de agua que habita el complejo (De Buen, 1943; Álvarez,1972). Las divergencias genéticas entre los miembros de este complejo son reducidas y ha dificultado el esclarecimiento de su taxonomía y la descripción de su historia evolutiva a través de secuencias de ADN. Por este motivo, en el estudio realizado en la presente tesis, se obtuvieron genotipos de microsatélites que fueron diseñados en trabajos previos para distintas especies de la familia Goodeidae (Boto y Doadrio, 2003; Hamill *et al.,* 2007). La alta tasa de mutación de estos marcadores (Nielsen y Palsbøll, 1999; Ellegren, 2000) ha permitido establecer los tiempos de divergencia entre los tres miembros del complejo, así como también caracterizar el pool genético de las especies para formar criterios para su conservación.

Complejo Catostomus plebeius-nebuliferus

Las especies reconocidas que conforman este complejo son *Catostomus plebeius* Baird y Girard, 1854 y *C. nebuliferus* Garman, 1851, distribuida la primera de ellas en el río Mezquital, Piaxtla, Fuerte, y Yaqui; en el sistema hidrográfico Guzmán (Santa María, Casas Grandes, y Del Carmen); Río Grande (Río Conchos en México y Río Grande en Estados Unidos EU); y en el río Mimbres en Nuevo México EU (Miller, Minckley y Norris, 2005). Mientras que *C. nebuliferus* se encuentra distribuida sólo en las cuencas endorreicas del río Nazas y Aguanaval. En trabajos previos también ha sido sugerido que poblaciones, tales como Mezquital, Piaxtla y Fuerte, pueden representar linajes evolutivos independientes (Ferris, Buth y Whit, 1982, Crabtree y Buth, 1987; McPhee, Osborne y Turner, 2008). La condición tetraploide del complejo, ha dificultado el estudio y esclarecimiento de las relaciones y divergencias de estas poblaciones, desde un punto de vista molecular (Uyeno y Smith, 1972). Por este motivo, su taxonomía y relaciones evolutivas son controvertidas y a algunos autores consideran a *C. nebuliferus* como un sinónimo de *C. plebeius* o incluyen a este complejo en el subgénero parafilético *Pantosteus* Cope, 1875 (Smith, 1992; Doosey *et al.,* 2010; Laitinen, 2012; Unmack *et al.,* 2014). En esta tesis nos planteamos como objetivo discernir la historia evolutiva del complejo, explorar su diversidad genética y estructura, asumiendo que su evolución está asociada a la historia geológica de la Sierra Madre Occidental y los procesos climáticos pleistocénicos (Ceballos, Arroyo-Cabrales y Ponce, 2010). Para ello, se utilizó un marcador mitocondrial y se amplificó un intrón de un gen nuclear ortólogo el cual se ha reportado su nivel de resolución a nivel interespecífica a lo largo de la familia Catostomidae (Clements, Bart y Hurley, 2012). Esto ha permitido estimar los tiempos de divergencia entre las poblaciones del complejo y entre los miembros de éste. Los resultados de esta tesis sugieren la necesidad de establecer un nuevo estatus taxonómico de los miembros del complejo, debido a la estructura genética significativa pero divergencias (en el caso del marcador nuclear) menores a la media estimada entre especies de la familia.

Complejo Squalius sp

El complejo sin describir taxonómicamente se distribuye únicamente en algunas cabeceras de tributarios del río Guadiana en la Península Ibérica. Posee características peculiares que lo hacen uno de los dos sistemas poliploides más complejos entre los vertebrados (Gregory y Mable, 2005). Este complejo hibridogenético fue reconocido como un grupo unisexual conformado por hembras triploides y diploides, debido a múltiples eventos de hibridación con *S. alburnoides* y *S. pyrenaicus* durante la configuración del sistemas hidrográficos que actualmente habita (Cunha y Doadrio 2009). El resultado de esos eventos fue la adquisición de un genoma A proveniente de *S. alburnoides* y un genoma P proveniente de *S. pyrenaicus*, además un genoma R genéticamente divergente más relacionado filogenéticamente con las especies *Squalius aradensis* (Coelho, Bogutskaya, Rodrigues & Collares-Pereira,1998) y *Squalius torgalensis* (Coelho, Bogutskaya, Rodrigues & Collares-Pereira,1998) endémico de las cuencas del sur de Portugal. Sin embargo el genoma R no es idéntico a estas especies y es indicativo de que dicho genoma pertenece a una especie que se desconoce y que está extinta actualmente. Este mismo patrón es reconocido también en el complejo híbrido *S. alburnoides* (Steindachner, 1866) (Cunha *et al.,* 2011). Por ello este nuevo complejo híbrido resulta interesante desde el punto de vista evolutivo. Sin embargo, en el trabajo previo existían incongruencias con respecto a las distancias genéticas observadas en los dos marcadores moleculares obtenidos (Cunha y Doadrio 2009). Además, sólo una población fue analizada y algunas preguntas evolutivas permanecían sin resolver. Así pues, este capítulo fue dedicado a describir la actual distribución del complejo, sus poblaciones desde el punto de vista genético, su estructura genética y los procesos evolutivos que pudieron estar involucrados en el surgimiento de este complejo. Para ello, se utilizó un marcador mitocondrial y uno nuclear, este último desarrollando una nueva estrategia para la obtención del locus ortólogo que nos permitiera describir las relaciones filogenéticas del genoma R. De esta forma se obtuvieron los alelos de origen hibrido que nos permitieran sugerir un primer modelo de reproducción del complejo *Squalius* sp, y establecer los primeros criterios para una formal descripción taxonómica en el futuro así como sentar las bases para su conservación.

Evolución de la viviparidad en Cyprinodontiformes

Este capítulo fue desarrollado tomando en cuenta que la viviparidad (vivíparos y ovovivíparos) asume una variedad de compromisos feto-maternales incluyendo el desarrollo, morfológicos, tróficos, osmoregulatorios, respiratorios, endocrinológicos e inmunológicos en los cuales el genoma mitocondrial debería estar involucrado (Wourms; 1981; Bai, Shaeley y Attardi, 2000; Kawaguchi *et al.,* 2014). Los cambios estructurales y funcionales surgidos durante el origen de los grupos vivíparos de peces, deberían ser posibles de observar y de estimar si dichos cambios están asociados a un proceso de selección positiva. Para la realización de este capítulo, se utilizó el genoma mitocondrial descrito en el capítulo 6 además de mitogenomas disponibles en GenBank para estudiar los cambios estructurales en los mitogenomas de grupos viviparos. En concreto se estudió el de los orden Ciprinodontiformes donde la viviparidad aparece repetidas veces durante la evolución (Meyers y Lydeard, 1993). Además, sus relaciones filogenéticas presentan inconsistencias cuando son basadas en datos moleculares o morfológicos (Parenti, 1981; Costa 1998; 2011; Webb, 1998; Hertwig, 2008; Setiamarga *et al.,* 2008; Pohl *et al.,* 2015). En este capítulo se pretendió sugerir una hipótesis fiogenética para el orden basada en mitogenomas, además de sugerir los cambios estructurales en el mitogenoma que pudieron estar asociados en el origen del viviparismo. En resumen, los mismo datos fueron utilizados para explorar la capacidad de resolución de cada gen del mitogenoma y buscar cuál o cuales genes pueden generar una misma topología con alto soporte a aquella obtenida utilizando el mitogenoma. De esta forma su aplicación debe contribuir de manera significativa a futuros estudios integrativos con genes nucleares o datos morfológicos. Por lo tanto, este cuarto capítulo representa una primera incursión a métodos genómicos por mi parte, para responder a cuestiones macroevolutivas y de evolución molecular en vertebrados.

Otros estudios genéticos

Los capítulos 5 y 6 consisten en descripciones de genomas mitocondriales, uno de ellos (*Xenotoca variata*, capítulo 6) utilizado para el desarrollo del capítulo 4 y el segundo (*Luciobarbus rifensis,* capítulo 5) como primer paso para futuros estudios y exploración de nuevos marcadores moleculares de ese grupo formado por especies tetraploides.

Finalmente, dado que los trabajos realizados durante esta tesis doctoral abordan distintos grupos taxonómicos y distintas preguntas evolutivas, los objetivos propuestos para esta tesis fueron establecidos de acuerdo a la búsqueda de resolver cuestiones evolutivas en grupos taxonómicos de Europa y México aplicando distintas estrategias experimentales a través de técnicas moleculares:

- Describir la historia evolutiva de los 3 complejos: *Allotoca diazi*, *Catostomus plebeius-nebuliferus* y *Squalius* sp.
- Esclarecer o aportar criterios genéticos a problemas taxonómicos.
- Aportar hipótesis evolutivas a través de información genética y la integración de información histórica.
- Generar criterios de conservación mediante la caracterización genética de las poblaciones de peces estudiadas.
- Proporcionar una nueva hipótesis filogenética para el Orden Ciprinodontiformes
- Describir los cambios estructurales y funcionales en el mitogenoma que pueden estar asociados en el origen y evolución de la viviparidad.
- Describir genomas mitocondriales para la búsqueda de nuevos marcadores moleculares o su aplicación en futuros estudios ictiofaunisticos.

Referencias

Álvarez J. 1972. Ictiología Michoacana V. Origen y distribución de la ictiofauna dulceacuícola michoacana. *Anales de la Escuela Nacional de Ciencias Biológicas*. 19, 155-161.

Avise JC. 2000. *Phylogeography: the history and formation of species*. Harvard university press.

Bai Y, Shakeley RM, Attardi G. 2000. Tight control of respiration by NADH dehydrogenase ND5 subunit gene expression in mouse mitochondria. *Molecular and Cellular Biology*. 20(3), 805-815.

Betancur RR, Broughton RE, Wiley EO, Carpenter K, López JA, Li C, et al. 2013. The tree of life and a new classification of bony fishes. *PLOS Currents Tree of Life*.

Boto L, Doadrio I. 2003. Polymorphic microsatellites in two different species of the genus *Zoogoneticus* Meek, 1902 (Goodeidae, Actynopterygii). *Molecular Ecology Notes*. 3, 70–72.

Brucet S, Pédron S, Mehner T, Lauridsen TL, Argillier C, Winfield IJ, et al. 2013. Fish diversity in European lakes: geographical factors dominate over anthropogenic pressures. *Freshwater Biology*. 58(9), 1779-1793.

Ceballos G, Arroyo-Cabrales J, Ponce E. 2010. Effects of Pleistocene environmental changes on the distribution and community structure of mammalian fauna of Mexico. *Quaternary Research.* 73: 464–473.

Clements MD, Bart HLJr, Hurley DL. 2012. A different perspective on the phylogenetic relationships of the Moxostomatini (Cyprinoformes: Catostomidae) based on cytochrome-b and Growth Hormone intron sequences. *Molecular Phylogenetics and Evolution.* 63: 159–167.

Comas I, Moya A, González-Candelas F. 2007. From phylogenetics to phylogenomics: the evolutionary relationships of insect endosymbiotic γ-Proteobacteria as a test case. *Systematic Biology*. 56(1), 1-16.

Costa WJ. 1998. Phylogeny and classification of the Cyprinodontiformes (Euteleostei: Atherinomorpha): A reappraisal. In: Malabarba, L.R.; Reis, R.E.; Vari, R.P.; Lucena, Z.M. Lucena, C.A.S. (Eds.), *Phylogeny and Classification of Neotropical Fishes.* Edipucrs, Porto Alegre. p. 519-536.

Costa WJ. 2011. Phylogenetic position and taxonomic status of *Anablepsoides*, *Atlantirivulus*, *Cynodonichthys*, *Laimosemion* and *Melanorivulus* (Cyprinodontiformes: Rivulidae). *Ichthyological Exploration of Freshwaters*. 22(3), 233.

Crabtree CB, Buth DG. 1987. Biochemical systematics of the catostomid genus *Catostomus*: assessment of *C. clarki*, *C. plebeius*, and *C. discobolus* including the Zuni sucker, *C. d. yarrow*. *Copeia.* 843–854.

Cunha C, Doadrio I, Abrantes J, Coelho MM. 2011. The evolutionary history of the allopolyploid *Squalius alburnoides* (Cyprinidae) complex in the northern Iberian Peninsula. *Heredity*. 106(1), 100-112.

Cunha C, Doadrio I. 2009. Evidencias de un nuevo complejo híbrido unisexuado del género *Squalius* (Teleostei, Cyprinidae) en el río Estena (Parque Nacional de Cabañeros): una aproximación filogenética. *Proyectos de Investigación en Parques Nacionales*. 357-375.

De Buen F. 1943. Los lagos michoacanos. I. Caracteres generales. El lago de Zirahuén. *Revista de la Sociedad Mexicana de Historia Natural*. 4(3-4), 211-232.

Doadrio I, Domínguez O. 2004. Phylogenetic relationships within the fish family Goodeidae based on cytochrome b sequence data. *Molecular phylogenetics and evolution*. 31(2), 416-430.

Domínguez-Domínguez O, Mercado-Silva N, Lyons J, Grier HJ. 2005. The viviparous goodeid fishes. *Viviparous fishes, MC Uribe and HJ Grier (eds.). New Life Publications, Homestead, Florida*, 525-569.

Doosey MH, Bart HLJr, Saitoh K, Miya M. 2010. Phylogenetic relationships of Catostomid fishes (Actinopterygii: Cypriniformes) based on mitochondrial ND4/ND5 gene sequences. *Molecular Phylogenetics and Evolution.* 54: 1028–1034.

Dudgeo D, Arthington AH, Gessner MO, Kawabata ZI, Knowler D J, Lévêque C, et al. 2006. Freshwater biodiversity: importance, threats, status and conservation challenges. *Biological Reviews*. 81(2), 163-182.

Edwards SV. 2008. Is a new and general theory of molecular systematics emerging?. *Evolution*. 63(1), 1-19.

Ekblom R, Galindo J. 2011. Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity*. 107(1), 1-15.

Ellegren, H. (2000). Microsatellite mutations in the germline: implications for evolutionary inference. *Trends in Genetics*. *16*(12), 551-558.

Ferris SD, Buth DG, Whitt GS. 1982. Substantial genetic differentiation among populations of *Catostomus plebeius*. *Copeia*. 444–449.

Freeland J. 2005. The evolution of population biology: past, present and future. *Journal of Biogeography*. 32(11), 2039-2040.

Graybeal A. 1994. Evaluating the phylogenetic utility of genes: a search for genes informative about deep divergences among vertebrates. *Systematic Biology*. 43(2), 174-193.

Greenwood PH, Rosen DE, Weitzman SH, Myers GS. 1966. Phyletic studies of teleostean fishes, with a provisional classification of living forms. Bulletin of the AMNH; v. 131, article 4.

Gregory TR, Mable BK. 2005. Polyploidy in animals. *The Evolution of The Genome*. 171, 427-517**.**

Hamill RM, Webb SA, Macías Garcia C, Graces JA, Magurran AE, Ritchie MG. 2007. Comparison of genetic diversity at microsatellite loci in near-extinct and non-endangered species of Mexican goodeine fishes and prediction of cross-amplification within the family. *Journal of Fish Biology*. 70, 16–32.

Helfman G. Collette BB, Facey DE, Bowen BW. 2009. *The diversity of fishes: biology, evolution, and ecology*. John Wiley & Sons.

Hertwig ST. 2008. Phylogeny of the Cyprinodontiformes (Teleostei, Atherinomorpha): the contribution of cranial soft tissue characters. *Zoologica Scripta*. 37(2), 141-174.

Hubbs CL, Turner CL. 1939. *Studies of the Fishes of the Order Cyprinodontes, XVI. A Revision of the Goodeidae*. University of Michigan Press.

Johnson N, Revenga C, Echeverria J. 2001. Managing water for people and nature. *Science*. 292(5519), 1071-1072.

Kawaguchi M, Tomita K, Sano K, Kaneko T. 2015. Molecular events in adaptive evolution of the hatching strategy of ovoviviparous fishes. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*. 324(1), 41-50.

Kerr JT, Kharouba HM, Currie DJ. 2007. The macroecological contribution to global change solutions. *Science*. 316(5831), 1581-1584.

Laitnen NJ. 2012. The phylogeogrtaphy of the Mountain Sucker [*Catostomus* (*Pantosteus*) *platyrhynchus*]. Unpublished D. Phil. Thesis. Brigham Young University.

McCormack JE, Hird SM, Zellmer AJ, Carstens BC, Brumfield RT. 2013. Applications of next-generation sequencing to phylogeography and phylogenetics. *Molecular Phylogenetics and Evolution*. 66(2), 526-538.

McPhee MV, Osborne MJ, Turner TF. 2008. Genetic diversity, population structure, and demographic history of the Rio Grande Sucker, *Catostomus* (*Pantosteus*) plebeius, in New Mexico. *Copeia.* 191–199.

Meiklejohn KA, Danielson MJ, Faircloth BC, Glenn TC, Braun EL, Kimball RT. 2014. Incongruence among different mitochondrial regions: a case study using complete mitogenomes. *Molecular Phylogenetics and Evolution*. 78, 314-323.

Meyer A, Lydeard C. 1993. The evolution of copulatory organs, internal fertilization, placentae and viviparity in killifishes (Cyprinodontiformes) inferred from a DNA phylogeny of the tyrosine kinase gene X-src. *Proceedings of the Royal Society of London B: Biological Sciences.* 254(1340), 153-162.

Miller RR, Minckley WL, Norris SM. 2005. *Freshwater fishes of Mexico*. University of Chicago Press.

Miya M, Takeshima H, Endo H, Ishiguro NB, Inoue JG, Mukai T, et al. 2003. Major patterns of higher teleostean phylogenies: a new perspective based on 100 complete mitochondrial DNA sequences. *Molecular Phylogenetics and Evolution*. 26(1), 121-138.

Nielsen R, Palsbøll PJ. 1999. Single-locus tests of microsatellite evolution: multi-step mutations and constraints on allele size. *Molecular phylogenetics and evolution*. 11(3), 477-484.

Oberprieler C. 2005. Temporal and spatial diversification of circum-Mediterranean Compositae-Anthemideae. *Taxon*. 54(4), 951-966.

Olden JD, Kennard MJ, Leprieur F, Tedesco PA, Winemiller KO, García‐Berthou. 2010. Conservation biogeography of freshwater fishes: recent progress and future challenges. *Diversity and Distributions*. *16*(3), 496-513.

Parenti LR. 1981. A taxonomic revision of the Andean killifish genus *Orestias* (Cyprinodontiformes, Cyprinodontidae). *American Museum of Natural History*.

Patwardhan A, Ray S, Roy A. 2014. Molecular markers in phylogenetic studies-A Review. *Journal of Phylogenetics & Evolutionary Biology*.

Pohl M, Milvertz FC, Meyer A, Vences M. 2015. Multigene phylogeny of cyprinodontiform fishes suggests continental radiations and a rogue taxon position of *Pantanodon*. *Vertebrate Zoology*. 65(1), 37-44.

Setiamarga DH, Miya M, Yamanoue Y, Mabuchi K, Satoh TP, Inoue JG, Nishida M. 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), 598-605.

Smith GR. 1992. Phylogeny and biogeography of the Catostomidae, freshwater fishes of North America and Asia, In: Mayden RL. Ed. *Systematics, historical ecology and North American freshwater fishes*. Stanford University Press, Stanford.

Stepien CA, Kocher TD. 1997. Molecules and morphology in studies of fish evolution. *Molecular systematics of fishes*. 1-11.

Sunnucks P. 2000. Efficient genetic markers for population biology. *Trends in Ecology & Evolution*. 15(5), 199-203.

Thomson RC, Wang IJ, Johnson JR. 2010. Genome‐enabled development of DNA markers for ecology, evolution and conservation. *Molecular Ecology*. *19*(11), 2184-2195.

Unmack PJ, Dowling TE, Laitinen NJ, Secor CL, Mayden RL, Sciozawa DK, Smith GR. 2014. Influence of introgression and geological processes on phylogenetic relationships of western North American Mountain Suckers (*Pantostues*, Catostomidae). *PloS ONE.* 9, e90061.

Uyeno T, Smith GR. 1972. Tetraploid origin of the karyotype of catostomid fishes.

Webb SA. 1998. *A phylogenetic analysis of the Goodeidae (Teleostei: Cyprinodontiformes)* (Doctoral dissertation).

Wourms JP, Grove BD, Lombardi J. 1981. 1 The Maternal-Embryonic Relationship in Viviparous Fishes. *Fish Physiology.* 11, 1-134.

Capítulo 1: Evolutionary history of the live-bearing endemic *Allotoca diazi* species complex (Actinopterygii, Goodeinae): Evidence of founder effect events in the mexican Pre-Hispanic period

DIUSHI K. CORONA-SANTIAGO[1], OMAR DOMÍNGUEZ-DOMÍNGUEZ[2] and IGNACIO DOADRIO[1]

[1] Departamento de Biodiversidad y Biología Evolutiva, Museo Nacional de Ciencias Naturales, CSIC, Madrid, España

[2] Laboratorio de Biología Acuática, Facultad de Biología, Universidad Michoacana de San Nicolás de Hidalgo, Morelia, Michoacán, México

Published in: PLOS ONE http://dx.doi.org/10.1371/journal.pone.0124138

Abstract

The evolutionary history of Mexican ichthyofauna has been strongly linked to natural events, and the impact of pre-Hispanic cultures is little known. The live-bearing fish species *Allotoca diazi*, *Allotoca meeki* and *Allotoca catarinae* occur in areas of biological, cultural and economic importance in central Mexico: Pátzcuaro basin, Zirahuén basin, and the Cupatitzio River, respectively. The species are closely related genetically and morphologically, and hypotheses have attempted to explain their systematics and biogeography. Mitochondrial DNA and microsatellite markers were used to investigate the evolutionary history of the complex. The species complex shows minimal genetic differentiation. The separation of *A. diazi* and *A. meeki* was dated to 400-7000 years ago, explained by geological and climate events. A bottleneck and reduction of genetic diversity in *Allotoca diazi* was detected, attributed to recent climate fluctuations and anthropogenic activity. The isolation of *A. catarinae* occurred ~1900 years ago. No geological events are documented in the area during this period, but the date is contemporary with P'urhépecha culture settlements. This founder effect represents the first evidence of fish species translocation by a pre-Hispanic culture of Mexico. The response of the complex to climate fluctuation, geological changes and human activity in the past and the future according to the ecological niches predictions indicates areas of vulnerability and important information for conservation. The new genetic information showed that the *Allotoca diazi* complex consist of two genetic groups with an incomplete lineage sorting pattern: Pátzcuaro and Zirahuén lakes, and an introduced population in the Cupatitzio River.

Keywords: evolutionary history, founder effect, live-bearing fish, P'urhépecha culture

Introduction

The diversity and distribution of freshwater ichthyofauna has been strongly linked to natural events, mainly geological process and historical climatic fluctuations. It is therefore that, taking into account the biogeographic history of each region in the world; the freshwater fish have been used as a model for the study of biogeography, speciation, paleohydrology and evolution. However, translocation and introduction of fish species is well documented in many ancient cultures having an important role in their current distribution, misunderstanding the evolutionary history of the species, In America, particularly in Mexico, the impact of pre-Hispanic cultures in the current distribution of the biota, is little known and mostly documented for birds (Christensen, 2000; Peer, 2011; Haemig, 2014). In fact, the scientific community has long assumed that translocations and introduction of species in America occurred only after European colonization (Telles, 2011).

The region in West-central Mexico, particularly between the Lerma-Santiago and Balsas rivers, is characterized by varied and rugged physiography, a product of intense tectonic and volcanic activity in the area since the Miocene (Miller, Minckley and Norris, 2005). Also, this region is an area with a long cultural history, with the P'urhépecha settlement developing one of the most stable empires at least 3000 years ago (Bravo-Espinosa, Mendoza Cantú and Medina Orozco, 2009), particularly due to the richness and abundance of lacustrine resources. The main resources currently exploited by settlements around the lakes are fish, as well as turtles, salamanders, clams, and wetland plants (Williams and Weigand, 1996; Guzmán, Polaco and Pollard, 2001). Even the area was formerly known by the Náhuatl culture as Mechuacan, “a place of fishermen,” and is currently called Michoacán.

Approximately 100 fish species are in the West-central Mexico, 70% of which are endemic and one of the most representative endemic fish groups of this region is the subfamily Goodeinae (Miller *et al.*, 2005). The group presents high species richness in a small area (c. 41 species) and unique characteristics associated with breeding strategies and embryo development, such as internal fertilization, matrotrophy, and sexual selection (Doadrio and Domínguez, 2004; Webb *et al.*, 2004; Domínguez-Domínguez, Doadrio and Perez-Ponce de Leon, 2006; Domínguez-Domínguez, 2008; Corona-Santiago and Domínguez-Domínguez, 2013). The family presents a high diversification apparently influenced by viviparity, with vicariance and adaptive radiation being the most important factors. The speciation rate of goodeids has not been constant and has been impacted by multiple extinctions, estimating that at least 25 000 years is required to establish an evolutionary lineage (Ritchie *et al.*, 2007).

The genus *Allotoca* is the most diverse within the Goodeinae, and is the subject of taxonomic controversy, with the need for taxonomic revision to validate species (Doadrio and Domínguez, 2004; Webb *et al.*, 2004; De Buen, 1942; Álvarez, 1959; Meyer, Radda and Domínguez-Domínguez, 2001). The *A. diazi* complex comprises three recognized species (Miller *et al*., 2005): *Allotoca diazi* (Meek, 1902), endemic to Pátzcuaro basin, *A. meeki* (Álvarez, 1959), endemic in the Zirahuén basin, and *Allotoca catarinae* (De Buen, 1942), restricted to the Cupatitzio River, an upper tributary of the Balsas River basin (Fig. 1). *Allotoca diazi* was extirpated from Pátzcuaro Lake and is currently restricted to the Chapultepec Spring, a tributary of Pátzcuaro Lake, while *A. meeki* was extirpated from Zirahuén Lake and currently restricted to Opopeo Spring, a tributary of the Zirahuén Lake (Miller *et al*., 2005). The lakes are located in the Michoacán-Guanajuato Corridor, an area of over 1000 volcanic cones active from the Pliocene to the present. To explain the evolution of these three sister species and other co-distributed taxa, two biogeographic hypotheses have been proposed. The first argues the existence of a tributary connecting the Lerma River with Cuitzeo, Pátzcuaro, and Zirahuén lakes (De Buen, 1943) around 700 000 years ago (Israde-Alcántara and Garduño-Monroy, 1999). The second hypothesis indicates the existence of a tributary connecting the Cupatitzio River with Zirahuén and Pátzcuaro lakes, reaching Zacapu (Álvarez, 1972).

Additionally, the application of modern biogeographic, genetic, systematic, and bioinformatics methodologies in the study of evolutionary processes within the Goodeinae has questioned recent connections of the Lerma and Cupatitzio Rivers with the Zirahuén, Pátzcuaro, Cuitzeo, and Zacapu basins (Doadrio and Domínguez, 2004; Domínguez-Domínguez, 2008; Domínguez-Domínguez *et al.*, 2008; Pérez-Rodríguez *et al.*, 2009). The taxonomic status of the three species has also been questioned (Corona-Santiago *et al*., 2013; Domínguez-Domínguez *et al.*, 2010), and the possibility of pre-Hispanic translocation of exploited species has been proposed (Domínguez-Domínguez, 2008).

In order to elucidate and analyze the two biogeographic hypotheses about the connections or disconnections between the Zirahuén, Pátzcuaro and Cupatitzio basins, and the possibility of a species translocation, we examined the evolutionary history of *Allotoca diazi* complex using two different molecular markers described for the Goodeinae Subfamily (Doadrio and Domínguez, 2004; Boto and Doadrio, 2003; Hamill *et al.*, 2007): a conserved molecular marker of mitochondrial DNA (Cytochrome *b*, *cytb* gene) and a less conserved microsatellite nuclear markers. We obtained these molecular markers to explore the genetic differentiation at intraspecific and interspecific level applying phylogeographic and phylogenetic approaches. The results of the genetic study were analyzed based on geological, paleoclimatic and anthropogenic records from central Mexico to access how natural and artificial historical processes are involved in the evolutionary history of *A. diazi* complex. In addition, we used all the genetic information and ecological niche predictions for identify operational conservation units.

Materials and methods

Ethics statements

This research was a part of the multidisciplinary work for restoration and conservation of the Cupatitzio River, and the complete work, field and laboratory protocols were approved by the Mexican Government, SEMARNAT (The Ministry of Environment and Natural Resources for Mexico) and CONACyT (National Council of Sciences And Technology) (Permit No. FOMIX-MICH-2009-C02-115897). All individuals were anesthetized using Tricaine mesylate (MS-222) to minimize suffering.

Fish sampling and DNA isolation

One hundred and twenty specimens were collected from six populations throughout the distribution range of the species complex (Table 1S). Our sampling did not recover any *A. diazi* or *A. meeki* from Pátzcuaro Lake and Zirahuén lacustrine basin, respectively. This is consistent with reports that these species were extirpated from these basins as a result of the introduction of exotic species (Miller *et al.*, 2005). Capture was by electrofishing between March and August 2011. Pectoral or caudal fin clips, were obtained and preserved in absolute ethanol and frozen at -20°C. Some specimens of the species complex were identified and deposited in the fish collection of the Universidad Michoacana de San Nicolás de Hidalgo CPUM in México; the remaining individuals were returned to the water. Approximately 2 mm^2 of tissue was used for DNA extraction. Digestion was performed with ATL QIAGEN® Buffer and Proteinase K and purifying by BioSprint DNA Blood Kit QIAGEN® according to the manufacturer's instructions.

Locus amplification, sequencing, and genotyping

Polymerase Chain Reaction (PCR) was performed to amplify the molecular markers. To amplify the *cytb* gene, we used the primers GLuDG (Palumbi *et al.*, 1991) and H16460 (Perdices *et al.*, 2002). The PCR consisted of a 12 µl volume reaction with a final concentration of 0.2 µM of each primer, 0.25 mM of each dNTP, 1.5 M of $MgCl_2$, and 1 unit of *Taq* DNA Polymerase. The PCR procedure consisted of 2 min at 94ºC followed by 35 cycles of 45 s at the 94ºC for DNA denaturation, 1 min at 46ºC for primer alignment, 1.5 min at 72ºC for synthesis, and a final extension of 5 min at 72ºC. The PCR products were checked by electrophoresis in agarose gel 1.5% and submitted to Macrogen Inc. (Korea) for sequencing.

For microsatellites, we searched polymorphic loci of the *A. diazi* complex using 20 primers designed for Goodeinae (Boto and Doadrio, 2003; Hamill *et al.*, 2007). Seven polymorphic microsatellite loci were selected (Table 2S). The 7 loci were amplified with QIAGEN® Multiplex PCR Kit according with the manufacturer specifications, and the products were submitted for genotyping (Secugen S. L. Corp, Spain).

mtDNA analysis

Manual alignment was implemented in Mega v5.2 (Tamura *et al.*, 2011) and examined chromatographically. The genetic diversity was estimated by calculating nucleotide diversity (π), haplotype diversity (h) and proportion of segregating sites (Θ) using the software DNAsp v5.0 (Librado and Rozas, 2009). Neutrality tests (Tajima's *D* and Fu's FS) to *cytb* were implemented using Arlequin v3.5.1.3 (Excoffier and Lischer, 2009). The genetic *P*-distances separating the three species were obtained (D_p) with Mega v5.2, and a bootstrapping process was implemented with 1000 repetitions. We conducted an analysis of geographic correspondence by network haplotype reconstruction using Network v4.6.1.0 (Bandelt, Forster and Röhl, 2000), applying the median-joining method.

Tree-building algorithms were used for phylogenetic reconstruction using *Allotoca zacapuesis*, a sister group of the species complex (Doadrio and Domínguez, 2004), as an outgroup. Neighbor-joining clustering was implemented in Mega v5.2 based on D_p with bootstrapping with 1000 replicates. Maximum likelihood reconstruction was conducted with RAxMLGUI v1.3 (Stamatakis, 2006; Silvestro and Michalak, 2012), using the generalized time reversible (GTR) + gamma + proportion of invariable sites model (Stamatakis, 2006) and 1000 bootstrap repetitions. The evolutionary substitution model selected was the transitional model (Posada, 2003) + invariable site proportion + gamma (TMI+I+G) using jModeltest v1.7 (Posada, 2008) and selected by the Akaike information criterion (AIC) (Table 3S). Phylogenetic reconstruction was by Bayesian inference (BI) with the software MrBayes v3.1.2 (Huelsenbeck and Ronquist, 2001), using the above selected evolutionary substitution model and implementing 2 reactions for 4 Markov Chains Monte Carlo (MCMC) processes with 3 million generations, sampling every 100 generations. We evaluated the convergence of the log-likelihood (-InL) value of the 2 reactions, with 10% of reconstructions discarded as burn-in to construct the consensus tree (σ=0.0002). The posterior probabilities obtained based on a confidence limit of 95% (highest posterior density-HPD) were used to evaluate the support values of nodes.

Microsatellite analysis

Quality of the samples genotyped and polymorphism in microsatellite loci was analyzed using GeneMapper v 4. Allele number by loci and species (N_a), effective number of alleles (N_{ae}), null allele proportion, and linkage disequilibrium were obtained using Arlequin v3.5.1.3 and Genepop v4.1.2 (Rousset, 2008). Observed (H_o) and expected (H_e) heterozygosity was determined, and the endogamy coefficient was obtained based on Wright's fixation index (F_{IS}) (Wright, 1951) to calculate the Hardy-Weinberg (HW) equilibrium across each locus and species using the software Arlequin v3.5.1.3, implementing 100 000 permutations.

The genetic structure level among the three species of the *A. diazi* complex was estimated with paired fixation indices (F_{ST}). Significance was assessed by AMOVA with 20 000 permutations implemented in the software Arlequin v 3.5.1.3 using multiple hypothetical arrangements based on phylogenetic and biogeographic information. A Bonferroni correction (Rice, 1989) was applied to each *P*-value obtained in the paired test, HW analysis and linkage disequilibrium estimation. Bayesian clustering was conducted with the software Structure v2.3.3 (Pritchard, Sthepens and Donnely, 2000). We performed a series of independent runs from values of K=1-10 populations assuming correlated allele frequencies and an admixture model (Hubisz *et al.*, 2009) with 10 runs for each K value to assess convergence of results without geographic information. For each K value, the MCMC was run with a burn-in of 100 000 steps and chain length of 500 000. Following the method described by Evanno *et al.* (Evanno, Regnaut and Goudet, 2005), ΔK was calculated to find the K value that best explained the genetic differentiation level.

To determine whether the difference in allele size contributed to genetic structure and if each locus is consistent with a strict Stepwise Mutation model, we estimated the R_{ST} values to assess deviation of equilibrium between R_{ST} and F_{ST} values. R_{ST} values greater than F_{ST} imply influence of the size differences of alleles on interpopulation genetic differentiation; thus we can assume that loci mutate faster by multi-step changes (Hardy, 2003; Cutter, 2005; Seyfert *et al.*, 2008; Haasl and Payseur, 2010). The R_{ST} estimation was performed with SPAGeDi v1.3a (Hardy and Vekemans, 2002) without geographic information, applying 20 000 permutations to obtain a test of differences among observed and expected R_{ST} and F_{ST} values.

Isolation-with-migration and demographic history analyses

This approach allows determination of gene flow before and after isolation within a population and assesses phylogenetic relationship scenarios in recent isolation events within closely related populations or species. Divergence times of populations of members of the *A. diazi* complex were calculated using three analyses in IMa2 software (Hey, 2010), which has the advantage of implementing models of gene flow. Pairwise analysis was conducted to find the confidence limits of the time to divergence among the three species, combining the molecular markers to substantiate and support the exchange rates within Markov chains. A molecular clock was calibrated using the mutation rate for *cytb* estimated for teleosts of 0.76-2.2%/million years ($4.33*10^{-6}$-$1.25*10^{-5}$subst/locus/generation) (Zardoya and Doadrio, 1999; Machordom and Doadrio, 2001; Near and Benard, 2004) and the Hasegawa-Kishino-Yano (HKY) evolutionary substitution model (Hasegawa, Kishino and Yano, 1985). The mutation rate used for calibration with microsatellites was $0.9*10^{-4}$-$1.5*10^{-4}$ locus/gamete/generation, estimation obtained for the Zebra fish *Danio rerio* (Shimoda *et al.*, 1999) and congruent with the mutation rate for vertebrates (Weber and Wong, 1993; Yue, David and Orban, 2007). The SMM evolutionary model was implemented for microsatellites. Ten independent runs were conducted for each scenario using 150 Markov chains with a burn-in of 500 000 for a total of 5 000 000, sampling topologies every 10 000 steps to recover 100 000 topologies from all runs. It was assumed that effective population size and the migration rate was different in each period.

The demographic history of the members of the species complex was obtained constructing Bayesian skyline plots (BSP) using Beast 1.8.0 (Drummond *et al.*, 1973) for *cytb* marker. Chains were run for 70 000 000 steps sampling each 1000 and 10% was discarded as burn-in under HKY substitution model, a relaxed molecular clock calibrated with the information of *cytb* mentioned above. All operators were optimized automatically and the results were analyses and visualized using Tracer 1.5 (Rambaut and Drummond, 2007). Effective sample size (ESS) was used to evaluate the strength of the analyses, and we determined the marginal distributions and HPD values in both analyses.

A test to identify recent bottlenecks was conducted using Bottleneck v1.2.0.2 (Cornuet and Luikart, 1996) with three evolutionary models for microsatellites: Infinite Allele Model (IAM) (Kimura and Crow, 1964), Two-Phase Model (Di Rienzo *et al.*, 1994), and Stepwise Mutation Model (SMM) (Kimura and Ohta, 1978). The test was implemented with 10 000 iterations based on a confidence limit of 95%. A mode-shift test was conducted to detect significant deviations in the allelic frequencies in populations that have experienced a recent bottleneck (Luikart *et al.*, 1998).

Identification of operational conservation units and predictions of future distribution

The genetic, biological, and ecological information obtained in this investigation, along with relevant published information (Doadrio, Perdices and Machordom, 1996; Domingues *et al.*, 2006), was used for identification of conservation units. In order to predict a future scenario for range of distribution of the natural members of the species complex, a discriminant analysis of presence-only data through maximum entropy using the program Maxent v3.3.3 (Phillips, Anderson and Schapire, 2006; Phillips and Dudík, 2008) was conducted for the present and for two periods in the future, years 2041 to 2060 and 2061 to 2080. The 6.0 Representative Concentration Pathways (RCP6.0) (Fujino *et al.*, 2006) was selected from downscaling Intergovernmental Panel on Climate Change 5 scenario (IPPC5) (Coupled Model Intercomparison Project 5) (Taylor, Stouffer and Meehl, 2012) implementing the Community Climate System Model 4 (CCSM4) (Moss *et al.,* 2010). We implemented independent runs of 5000 iterations with 19 climatic variables (http://www.worldclim.org/bioclim), bootstrapping of 100 replicates to obtain 95%

confidence intervals, regularization multiplier (RM) values from 0.01 to 10, evaluating area under the ROC curve (AUC) values and using the omission rates to discard identical AUC of different RM values. Collinear climatic variables were discarded (Table 4S).

Results

mtDNA analysis

One hundred nine sequences of the mitochondrial gene *cytb* (1037pb) were obtained for the sampled populations of the three species (GenBank accession numbers: KJ776467-KJ776575) (Table 1S). Overall, *A. diazi* and *A. meeki* nucleotide diversity was 0.0005±0.0002, haplotype diversity was 0.65±0.13, and mean proportion of segregating sites per locus was 1.4±0.3. *Allotoca catarinae* showed null genetic diversity in four sampled populations with a unique haplotype. *Allotoca diazi* was the most genetically diverse with five haplotypes (n=35), followed by *A. meeki* with three haplotypes (n=22), of their respective and unique populations (Table 1). No significant deviation from neutrality was found for the *A. diazi* complex. The mean of the genetic distances of *A. catarinae* with respect to *A. diazi* and *A. meeki* was 0.6% and between *A. diazi* and *A. meeki* was 0.3%. The three phylogenetic approximations were congruent. The topologies were similar with all DNA sequences and haplotypes. The tree based on haplotypes is presented in Fig. 2. The consensus tree was formed by two haplogroups: a non-monophyletic group of *A. diazi*/*A. meeki* and a group with the complete of *A. catarinae*. The haplotype network recovered the same two genetic groups, with four mutational steps separating them (Fig. 3). Spatial congruence was not found in haplogroup *A. diazi*/*A. meeki*. Two haplotypes of *A. diazi* were shared with *A. meeki*, and the unique haplotype of *A. catarinae* is not shared with the other members of the species complex.

Microsatellite analysis

For the 7 microsatellites loci amplified in populations of the *A. diazi* complex, we obtained a total of 96 genotypes. The average number of alleles per locus was 10.4±3.5; the number of genetic variants contributing to heterozygosity in each locus of the species complex was 5.8±2.2; and the proportion of null alleles was 0.06±0.03 ($P > 0.05$). Significant deviation from linkage equilibrium was not found at most loci, LD=0.06±0.03 ($P > 0.05$). Significant linkage disequilibrium was observed in two loci of *A. diazi* (Table 5S). Significant deviation from HW equilibrium associated with lower than expected heterozygosity was found in *A. diazi* and *A. catarinae* ($P < 0.05$) (Table 2).

The observed R_{ST} values for all loci were not significantly different from expected (Ho:R_{ST}=pR_{ST}) or from the F_{ST} values (Table 6S), suggesting that allele size difference did not contribute to the genetic structure of the *A. diazi* complex. Low but significant genetic differentiation was detected (F_{ST}=0.113 $P < 0.05$) via AMOVA analysis (S7 Table). Genetic structure by BI recovered two genetic groups (*A. diazi+A.meeki – A. meeki*) with a high proportion of individuals assigned to other groups (0.3-0.7) (Fig. 4). However, significant genetic differentiation was detected between the three members of the species complex. The lowest F_{ST} value of genetic differentiation was found between *Allotoca diazi* and *A. meeki* (0.084), while genetic differentiation of *A. catarinae* from *A. diazi* and *A. meeki* was 0.133 and 0.109 (P<0.05), respectively (Table, 3). The number of migrants value between members of the species complex were not significant.

Genetic structure by BI recovered two genetic groups (K=2 *A. catarinae* + *A. diazi/A.meeki*) with high proportion of individuals assigned to another group (0.3±0.7) (Figs 3, 1S of the Supplementary information).

Estimate of divergence times and demographic history

The *A. diazi* complex divergence from its common ancestor *A. zacapuensis* was estimated to have occurred in the past million years. The isolation event of *A. diazi* from *A. meeki* occurred in the past 7000-400 years (HPD≥95%) with the peak of probability estimated as 1700 years ago (Table 4 and Fig. 2S). Divergence time and isolation of *A. catarinae* from *A. diazi* and *A. meeki* was calculated at 8000-700 years ago, with the highest probability being 1900 years ago for *A. diazi* and 2500 years ago for *A. meeki*.

The migration rate estimated among the members of the *A. diazi* complex was fewer than 5.5 individuals/generation and was not significant (Table 3). BSP analysis using *cytb* revealed demographic decline in the last 10 000 years for *A. catarinae* (HPD≥95%) (Fig. 4S), and recent bottlenecks for the three members of the species complex were detected out of the HPD confidence range. However, recent bottlenecks were identified in the three species based on different evolutionary models using microsatellites. The bottlenecks detected in *A. diazi* and *A. catarinae* via SMM were supported by the F_{IS} coefficients: 0.141 and 0.151, respectively. The bottleneck detected in *A. meeki* was associated with higher than expected heterozygosity and allele size differences in the IAM analysis.

Operational conservation units and predictions of future distribution

We included in the analysis the species considered as natural members of the species complex: *Allotoca diazi* and *A. meeki*. The species *Allotoca catarinae* was excluded of the analysis because could be result of translocation. The predictions were performed taking into account that the gene flow between species will be null.

The ecological niches area predicted with present environment conditions for *Allotoca diazi* and *A. meeki* is bigger than their real current distribution. In the case of the ecological niche models for two periods in the future, high AUC values for were obtained for both species (AUC=0.940-0.995 RM=0.01-0.1, see Table 4S). In the first period between the years 2041 to 2060, the model suggest that the ecological niches available for the two species will be located at higher altitudes than they are currently distributed (Figs 5, 6). In the second period of the future (2061-2080), the area of ecological niches predicted will increase for *A. diazi* but the probability of occurrence in the Pátzcuaro Lake is low. In the same period, the ecological niches predicted to *A. meeki* could decrease drastically. According with the genetic and future predictions, we propose the existence of two OCUs corresponding to natural populations in Pátzcuaro and Zirahuén basins.

Discussion

Allotoca diazi complex

The isolation of the *Allotoca diazi* complex from its common ancestor *A. zacapuensis*, does not contradict the hypothesis of Álvarez (1972) with respect to the connection between the Zacapu and Pátzcuaro basins during the Pleistocene, resulted from the formation of the El Zirate mountain and the northern Pátzcuaro Lake shoreline during the Late Pleistocene (Robles-Camacho *et al.*, 2010).

The *Allotoca diazi* complex consists of a genetic group with low genetic structure and its isolation within the members was more recent, during the Holocene. The isolation of *A. diazi*, *A. meeki*, and *A. catarinae* is reflected in the low genetic differences, non-monophyletic patterns, shared haplotypes, and genetic groups assignment, and we considered that the species complex consist in an incomplete lineage sorting pattern (Maddison, 1997; Arbogast, *et al.*, 2002; Yu *et al.*, 2011; Rosenfeld, Payne and DeSalle, 2012). The genetic distances among the three species in *cytb* are smaller than reported to sister Goodeinae species (1.7-11%) (Doadrio and Domínguez, 2004) and sister species, congeneric species, and cofamilial genera within and across the major vertebrate taxonomic classes (~2%) (Brown, George and Wilson, 1979; Avise *et al.*, 1987; Avise, 1998; Bradley and Baker, 2001). The genetic variation estimated in the species complex based on the *cytb* gene is attributed to genetic drift and mutation (D_T and FS values did not show significant deviation from neutrality). Low genetic variation was detected as a genetic pool arrangement through AMOVA analysis. However, BI detected this minimum mark of genetic structure as two genetic groups. The results obtained are supported by the minimal morphological differences previously reported (De Buen, 1942; Meyer *et al.*, 2001; Regan, 1906; Hubbs, 1926; Hubbs, 1932; Hubbs and Turner, 1939; Hieronimus, 1995).

Allotoca diazi and A. meeki isolation

The hypothesis of connection and disconnection of Pátzcuaro and Zirahuén lakes during the last 700 000 years (De Buen, 1943) is supported by stratigraphic

(Israde-Alcántara *et al.*, 2008); limnologic data (Bradbury, 2000; Israde-Alcántara, Garduño-Monroy and Ortega, 2002); shared native icthyofauna such as *Chirostoma estor*, *Chirostoma pátzcuaro* (Atherinopsidae), *Algansea lacustris* (Cypronidae), *Alloophorus robustus*, *Skiffia lermae*, *Allotoca dugesii*, and *Goodea atripinnis* (Goodeidae) (Villaseñor-Gómez, 2005); and genetic data (Domínguez-Domínguez, 2008; Bloom *et al.*, 2009).

Shared haplotypes of *A. diazi* and *A. meeki* and the high proportion of individuals assigned to other species (0.5 to 0.7) by microsatellite data, suggest recent gene flow. Microsatellite information showed low level of genetic differentiation (F_{ST}=0.084). Low genetic diversity was estimated for *A. diazi* in the *cytb* gene (h=0.78, π=0.0030) and even lower for *A. meeki* (h=0.52, π=0.0007). The migration rate between the two species is not significant any time.

We estimated that the isolation of *A. diazi* from *A. meeki* occurred in the past 400-7000 years, which is consistent with geological, climatic, and anthropogenic events involving Pátzcuaro and Zirahuén lakes. Tecto-volcanic events giving rise to the separation of Pátzcuaro and Zirahuén lakes began 8000 years ago with the formation and activity of the La Tasa volcano southwest of Lake Pátzcuaro and climate fluctuations causing decline in water level and drying of streams (Álvarez, 1972; Israde-Alcántara, *et al.*, 2005; Garduño-Monroy, *et al.*, 2009; Ortega *et al.*, 2010; Vázquez *et al.*, 2010). Fluctuations in the level of Lake Pátzcuaro during the Holocene associated with climate change, human activity (Israde-Alcántara, *et al.*, 2005; Garduño-Monroy *et al.*, 2009; Street-Perrott, Perrott and Harkness, 1989; O'Hara, 1993; O'Hara, Metcalfe and Street-Perrott, 1994; O'Hara and Metcalfe, 1995; Fisher *et al.*, 2003; Fisher, 2005), and occurrence of tsunamis (Garduño-Monroy *et al.*, 2011) could be involved to the isolation and demographics changes of *A. diazi* and *A. meeki*. These demographic changes were reflected in genetic diversity including the positive D_T value, which was found to be low in *A. diazi*, and associated with recent bottlenecks that promoted inbreeding (F_{IS}=0.141), reduced heterozygosity, and the loss of allelic richness. In contrast, a bottleneck identified in *A. meeki* is attributed to the rapid loss of allelic diversity; however, the R_{ST} value and the BSP analysis (in the HPD range) indicates that the bottleneck detected by IAM were overestimated. The population dynamics of *A. meeki* can be explained by the variations in nucleotide diversity and proportion of segregating sites ($\pi < \Theta s$), which is interpreted as a recent population reduction followed by expansion, as demonstrated by the estimated value of D_T=-0.86 and non-significant positive values of Fu's FS estimates (Hudson , Kreitman and Aguadé, 1987; Tajima, 1989; Fu, 1997).

Based in our results, we conclude that natural processes, such as geologic events and climate fluctuations, are the main factors responsible for the isolation of *A. diazi* from *A. meeki* during the Holocene, not enough time to diverge and showing an incomplete lineages sorting.

Allotoca catarinae

Divergence times and isolation of *A. catarinae* from *A. diazi* and *A. meeki* was calculated to be within the past 8000-700 years and 8000-800 years, respectively. The peak of probability for isolation of *A. caratinae* was 1900 years ago. The low number of native species fish shared (Domínguez-Domínguez, 2008; Villaseñor-Gómez, 2005), and the geological and climatic records not support the connection and disconnection of the region in ancient periods: as the formation of Cupatitzio River ~30 mya (López Ramos, 1981; Gómez-Tuena, Orozco-Esquivel and Ferrari, 2005), or in recent periods: as the activity and collapse of Tancítaro volcano 792 000±22 000 years ago (Ownby, *et al.*, 2007; Morelli *et al.*, 2010), and the origin of Zirahuén Lake 17 000 years ago (Ortega *et al.*, 2010; Vázquez, *et al.*, 2010; Tamayo and West, 1964), contrasting the Álvarez hypothesis (Álvarez, 1972).

We propose that the existence of this taxon in the Cupatitzio River could be the result of a pre-Hispanic introduction, causing a founder effect, which resulted in rapid fixation of a single *cytb* haplotype in populations, along with a loss of genetic diversity. The lack of shared haplotypes between Cupatitzio populations and Pátzcuaro-Zirahuén populations cloud this interpretation, but the specimens sampled in the Pátzcuaro basin were from a small spring 10 km east of the lake and represent the only known local population of the species, which has been extirpated form the main basin (Miller *et al.*, 2005). We speculate that the ancestral haplotype introduced into Cupatitzio came from the main basin, and ancestral haplotype is lost. The haplotypes collected from the small spring (50 m^2 approximately) were isolated from the main basin, and the genetic drift fixed rare haplotypes. In addition, null genetic diversity was found in *A. catarinae,* associated with recent bottlenecks. *Allotoca catarinae* is the only one in which bottlenecks were detected in several models of evolution, and also supported with the BSP analysis. These bottlenecks drive in all cases to low heterozygosity (F_{IS}=0.151).

Relevant anthropogenic events in the region of Lakes Pátzcuaro and Zirahuén and the Cupatitzio River are congruent with our estimated divergence times. There is evidence that fishing was a major activity among the P'urhépecha. Fish, specifically goodeines and atherinopsids, had importance as food, even today goodeids are important part of the food supply for natives in the region, as well as culturally, in all socioeconomic strata, especially for Cazonci, the supreme ruler of the P’urhépecha state. Offerings of fish found in tombs at the archaeological site Uricho southwest of Lake Pátzcuaro, dating to the Late and Epi-Classic periods (~500-900 A.D.) (Williams and Weigand, 1996; Guzmán *et al*., 2001).

The “islander”, “Chichimeca” and the “Coringuaro”, considered as first human groups to settle in the region of Uruapan and south of Michoacán, probably had influence on the management of aquatic resources and translocation of species may have occurred (De Alcalá, 2000; Roskamp, 1998). However, the information about their biological impact is less well documented. The rise of the P'urhépecha culture occurred around 1350, during the formation of the empire, with establishment of new villages, and species translocations possibly took place (De Alcalá, 2000; Roskamp, 1998; Goggin, 1943; Nichols and Pool, 2012). The expansion of the P'urhépecha Empire occurred during the second period with establishment around 1400 (second period) and 1450-1530 (third period) (Fig. 3S). During these periods, settlements ranged from the region of the La Palma River, to the Marqués River region, in the Balsas Depression. Evidence from the P'urhépechas settlements in the Cupatitzio River subbasin and the Tepalcatepec-Balsas River suggests that, in this period, *A.diazi* translocation may have occurred. We conclude that the biogeographic history of *A. catarinae* was possibly determined by a founder effect mediated by pre-Hispanics in the past 3000 years, and taking into account this inference we cannot rule out species translocations in Pátzcuaro and Zirahuén lakes.

Conservation implications

The factors involved in the genetic diversity and demographic history of *A. diazi* and *A. meeki*, were not only the founder effect (*A. meeki*) and climate fluctuations (*A. diazi*), but also anthropogenic activity (Vázquez *et al.*, 2010; Street-Perrott *et* al., 1989; O'Hara, 1993; O'Hara *et al.*, 1994; O'Hara and Metcalfe, 1995; Fisher *et al.*, 2003; Fisher, 2005).

The two OCU’s proposals (Pátzcuaro and Zirahuén lakes) are areas with a high degree of alteration, including eutrophication, introduction of exotic species, drought, and overfishing (Domínguez-Domínguez *et al*., 2006; Domínguez-Domínguez, 2008; Domínguez-Domínguez *et al.*, 2008; Domínguez-Domínguez *et al.*, 2010; Lyons *et al.*, 1998; Zambrano, 2011; Ramírez-Herrejón *et al.*, 2014), which led to the extirpation of *A. diazi* from the main Pátzcuaro basin and *A. meeki* from Zirahuén Lake. Neither Pátzcuaro nor Zirahuén are designated protected areas. In addition, the results of the future ecological niche modeling for *A. diazi and A. meeki* show a reduction of potential habitats questioning the survival of the species, taking into account its genetic vulnerability to climate fluctuations and human activity, which indicates the necessity of the implementation of effective conservation and management strategies of this micro-endemic species with a long history of economic and cultural importance to the lacustrine settlements.

Acknowledgements

Thank to Rodolfo Pérez, Fernando Mar, Eloisa Torres and Silvia Perea for their help in sampling trips and valuable comments to this work. Also, thank to the two anonymous reviewers and editor for their observations and contributions to this work. ODD is greateful to Chester Zoo Garden, Fish Aqualab Conservation Project and to the Programa de Mejoramiento del Profesorado grant 2009 for its financial support. DKCS.

References

Álvarez J. 1959. Contribución al conocimiento del género *Neoophorus* (Pisc., Goodeidae). *Ciencia: Revista Hispano-Americana de Ciencias Puras y Aplicadas*. 19, 13–22.

Álvarez J. 1972. Ictiología Michoacana V. Origen y distribución de la ictiofauna dulceacuícola michoacana. *Anales de la Escuela Nacional de Ciencias Biológicas*. 19, 155-161.

Arbogast BS, Edwards SV, Wakeley J, Beerli P, Slowinski JB. 2002. Estimating divergence times from molecular data on phylogenetic and population genetic timescales. *Annual Review of Ecology and Systematics*. 33, 707–740.

Avise JC, Arnold J, Ball RM, Bermingham E, Lamb T, Neigel JE *et al.* 1987. Intraspecific phylogeography: The mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecology and Systematics*. 18, 489–522.

Avise JC. 1998. The history and purview of phylogeography: A personal reflection. *Molecular Ecology*. 7, 371–379.

Bandelt H-J, Forster P, Röhl A. 2000. Median-Joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*. 16, 37–48.

Bloom DD, Piller KR, Lyons J, Mercado-Silva N, Medina-Nava M. 2009. Systematics and biogeography of the silverside tribe Menidiini (Teleostomi: Atherinopsidae) based on the mitochondrial ND2 gene. *Copeia*. 2, 408-417.

Boto L, Doadrio I. 2003. Polymorphic microsatellites in two different species of the genus *Zoogoneticus* Meek, 1902 (Goodeidae, Actynopterygii). *Molecular Ecology Notes*. 3, 70–72.

Bradbury JP. 2000. Limnologic history of the Lago de Pátzcuaro, Michoacán, Mexico from the past 48,000 years: impacts of the climate and man. *Paleogeography, Palaeoclimatology, Palaeoecology*. 163, 69-95.

Bradley RD, Baker RJ. 2001. A test of the genetic species concept: cytochrome-b sequences and mammals. *Journal of Mammalogy*. 82, 960–973.

Bravo-Espinosa M, Mendoza Cantú ME, Medina Orozco LE. 2009. Escenarios de erosión bajo diferentes manejos agrícolas en la cuenca del lago de Zirahuén, Michoacán, México. *Investigaciones Geográficas*. 73–84.

Brown WM, George M, Wilson AC. 1979. Rapid evolution of animal mitochondrial DNA. *PNAS*. 76, 1967–1971.

Christensen AF. 2000. The fifteenth- and twentieth-century colonization of the basin of Mexico by the grat-tailed grackle (*Quiscalus mexicanus)*. *Global Ecology and Biogeogrpahy*. 9, 415-420.

Cornuet JM, Luikart G. 1996. Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics*. 144, 2001–2014.

Corona-Santiago DK, Domínguez-Domínguez O. 2013. Goodeids phylogeny. In Kempkes M, Köck M, Stawikouski R (eds): Contributions to the biology and conservation of splitfins (Goodeidae). *Westarp Wissenschaften*, Germany. 681.

Cutter AD. 2005. Molecular evolution inferences from the *C. elegans* genome. In The *C. elegans* Research Community (eds): WormBook-The Online Review of *C. elegans* Biology. *Pasadena* (CA).

De Alcalá FJ. *Relación de Michoacán* (Fondo de Cultura Económica, Ed,). Fondo 2000, 1540;88.

De Buen F. 1942. Una nueva especie del Neoophorus diazi (Meek). *Anales del Instituto de Biología*. 13, 342-249.

De Buen F. 1943. Los lagos michoacanos. II. Pátzcuaro. *Revista de la Sociedad Mexicana de Historia Natural*. 4, 211-232.

Di Rienzo A, Peterson AC, Garza JC*,* Valdes AM, Slatkin M, Freimer NB*.* 1994. Mutational processes of simple-sequence repeat loci in human populations. *PNAS*. 91, 3166–3170.

Doadrio I, Domínguez O. 2004. Phylogenetic relationships within the fish family Goodeidae based on Cytochrome *b* sequence data. *Molecular Phylogenetics and Evolution*. 31, 416–430.

Doadrio I, Perdices A, Machordom A. 1996. Allozymic variation of the endangered killifish *Aphanius iberus* and its application to conservation. *Environmental Biology of Fishes*. 45, 259–271.

Domingues VS, Santos RS, Brito A, Almada VC. 2006. Historical population dynamics and demography of the eastern Atlantic pomacentrid *Chromis limbata*

(Valenciennes, 1833). *Molecular Phylogenetics and Evolution*. 40, 139–147.

Domínguez-Domínguez O, Alda F, Pérez-ponce de León G García-Garitagoitia JL, Doadrio I. 2008b. Evolutionary history of the endangered fish *Zoogoneticus quitzeoensis* (Bean, 1898) (Cyprinodontiformes: Goodeidae) using a sequential approach to phylogeography based on mitochondrial and nuclear DNA data. *BMC Evolutionary Biology*. 8, 161.

Domínguez-Domínguez O, Doadrio I, Perez-Ponce de Leon G. 2006. Historical biogeography of some river basins in central Mexico evidenced by their goodeine freshwater fishes: A preliminary hypothesis using secondary Brooks parsimony analysis. *Journal of Biogeography*. 33, 1437–1447.

Domínguez-Domínguez O, Pedraza-Lara C, Gurrola-Sánchez N, Perea S, Pérez-Rodríguez R, Israde-Alcántara I *et al.* 2010. Historical Biogeography of the Goodeinae (Cyprinodontiforms). In Uribe-Aranzabal MC, Grier H (eds): Viviparous fishes II. *New Life Publications*, Florida. 13-30.

Domínguez-Domínguez O. 2008a. Filogeografía de *Zoogoneticus quitzeoensis*, *Xenotoca variata* y *Alloophorus robustus* (Cyprinodontiformes: Goodeidae) en el centro de México: implicaciones taxonómicas y de conservación. Doctoral Dissertation. Instituto de Biología, UNAM. Distrito Federal, México.

Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecuar Biology and Evolution*. 29, 1969-1973.

Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, 8, 2611-2520.

Excoffier L, Lischer HEL. 2009. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*. 10, 564–567.

Fisher CT, Pollard HP, Israde-Alcántara I, Garduño-Monroy VH, Banerjee SK. 2003. A reexamination of human-induced environmental change within the Lake Pátzcuaro Basin, Michoacán, México. *PNAS*. 100, 4957–4962.

Fisher CT. 2005. Demographic and Landscape Change in the Lake Patzcuaro Basin, Mexico: Abandoning the Garden. *American Anthropologist*. 107, 87–95.

Fu YX. 1997. Statistical Tests of Neutrality of Mutations Against Population Growth, Hitchhiking and Background Selection. *Genetics*. 147, 915–925.

Fujino J, Nair R, Kainuma M, Masui T, Matsuoka Y. 2006. Multi-gas mitigation analysis on stabilization scenarios using AIM global model. *Multigas Mitigation and Climate Policy*. *The Energy Journal*. 3 (Special Issue).

Garduño-Monroy VH, Chávez-Hernández J, Aguirre-González J, Vázquez-Rosas R, Mijares-Arellano H, Israde-Alcántara I *et al.* 2009. Zonificación de los periodos naturales de oscilación superficial en la ciudad de Pátzcuaro, Mich., México, con base en microtremores y estudios de paleosismología. *Revista Mexicana de Ciencias Geológicas*. 26, 623–637.

Garduño-Monroy VH, Soria-Caballero DC, Israde-Alcántara I, Hernández VMM, Rodríguez-Ramírez A, Ostroumov M *et al.* 2011. Evidence of tsunami events in the Paleolimnological record of Lake Pátzcuaro, Michoacán, Mexico. *Geofísica Internacional*. 50-2, 147–161.

Goggin JM. 1943. An archaeological survey of the Río Tepalcatepec basin, Michoacán, México. *American Antiquity*. 9, 44–58.

Gómez-Tuena A, Orozco-Esquivel MT, Ferrari L. 2005. Petrogénesis ígnea de la Faja Volcánica Transmexicana. *Boletín de la Sociedad Geológica Mexicana*. 57, 227–283.

Guzmán AF, Polaco ÓJ, Pollard HP. 2001. Ofrendas de peces asociadas a entierros del Clásico-Epiclásico en Urichu, Michoacán, México. *Archaeofauna*. 149–162.

Haasl RJ, Payseur BA. 2010. The number of alleles at a microsatellites defines the allele frequency spectrum and facilitates fast accurate estimation of θ. *Molecular Biology and Evolution*. 12, 2702-2715.

Haemig PD. 2014. Aztec introduction of the great-tailed grackle in ancient Mesoamerica: Formal defense of the Sahaguntine historical account. *NeoBiota*. 22, 59-75.

Hamill RM, Webb SA, Macías Garcia C, Graces JA, Magurran AE, Ritchie MG. 2007. Comparison of genetic diversity at microsatellite loci in near-extinct and non-endangered species of Mexican goodeine fishes and prediction of cross-amplification within the family. *Journal of Fish Biology*. 70, 16–32.

Hardy OJ, Charbonnel N, Fréville H, Heuertz M. 2003. Microsatellite allele sizes: A simple test to assess their significance on genetic differentiation. *Genetics*. 163, 1467–1482.

Hardy OJ, Vekemans X. 2002. SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes*. 2, 618–620.

Hasegawa M, Kishino H, Yano T. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution*. 22, 160–174.

Hey J. 2010. Isolation with migration models for more than two populations. *Molecular Biology and Evolution*. 27, 905–920.

Hieronimus H. 1995. "Leven baren" wat is dat wigenlijk?. *Poecilia Nieuws*. 2, 22-27.

Hubbs CL, Turner CL. 1939. Studies of the fishes of the Order Cyprinodontes, XVI: A revision of the Goodeidae.

Hubbs CL. 1926. Studies of the fishes of the Order Cyprinodontes. VI. 3–87.

Hubbs CL.1932. Studies of the fishes of the order Cyprinodontes, XII: A new genus related to Empetrichthys. 1–5.

Hubisz MJ, Falush D, Stephens M, Pritchard JK. 2009. Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources*. 9, 1322–1332.

Hudson RR, Kreitman M, Aguadé M. 1987. A test of neutral molecular evolution based on nucleotide data. *Genetics*. 116, 153–159.

Huelsenbeck JP, Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*. 17, 754–755.

Israde-Alcántara I, Garduño-Monroy VH, Fisher CT, Pollard HP, Rodríguez-Pascua MA. 2005. Lake level change, climate, and the impact of natural events: The role of seismic and volcanic events in the formation of the Lake Patzcuaro Basin, Michoacan, Mexico. *Quaternary International*. 135, 35–46.

Israde-Alcántara I, Garduño-Monroy VH, Ortega MR. 2002. Paleoambiente lacustre del cuaternario tardio en el centro del lago de Cuitzeo. *Hidrobiología*. 12, 61-78.

Israde-Alcántara I, Garduño-Monroy VH. 1999. Lacustrine record in a volcanic intra-arc setting: The evolution of the Late Neogene Cuitzeo basin system (central-western Mexico, Michoacan). *Palaeogeography, Palaeoclimatology, Palaeoecology*. 151, 209–227.

Israde-Alcántara I, Wade M, Garduño-Monroy VH, Barron J. 2008. Estratigrafía y encuadramiento geodinámico de las cuencas lacustres del centro de México. Unión Mexicana de Estudios del Cuaternario. *Fondo de Cultura Económica*, México, DF.

Kimura M, Crow JF. 1964. The number of alleles that can be maintained in a finite population. *Genetics*. 49, 725–738.

Kimura M, Ohta T. 1978. Stepwise mutation model and distribution of allelic frequencies in a finite population. *PNAS*. 75, 2868–2872.

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

López Ramos E. 1981. Paleogeografia y tectonica del Mesozoico de Mexico. *Revista Mexicana de Ciencias Geológicas*. 5, 158–177.

Luikart G, Allendorf FW, Cornuet JM, Sherwin WB. 1998. Distortion of allele frequency distributions provides a test for recent population bottlenecks. *The Journal of Heredity*. 89, 238–247.

Lyons J, González-Hernández G, Soto-Galera E, Guzmán-Arroyo M. 1998. Decline of freswater fishes and fisheries in selected drainages of west-central Mexico. *Fisherie*. 23, 10-18.

Machordom A, Doadrio I. 2001. Evidence of a cenozoic Betic-Kabilian connection based on freshwater fish phylogeography (Luciobarbus, Cyprinidae). *Molecular Phylogenetics and Evolution*. 18, 252–263.

Maddison WP. 1997. Gene Trees in Species Trees. *Systematic Biology*. 46, 523–536.

Meek SE. 1902. A contribution to the ichthyology of Mexico. *Zoological Series, Field Columbian Museum*. 3, 63–128.

Meyer MK, Radda AC, Domínguez-Domínguez O. 2001. Notes on the genera *Neoophorus* Hubbs & Turner, 1937, with a description of a new species of *Allotoca* from Laguna de Zacapu, Michoacán, Mexico (Teleostei, Cyprinodontiformes: Goodeidae). *Annalen des Naturhistorischen Museums in*

Wien. 103 B, 453–460.

Miller RR, Minckley WL, Norris SM. *Freshwater fishes of México*. University Of Chicago Press. 2005.

Morelli S, Monroy V, Gigli G, Falorni G. 2010. The Tancítaro debris avalanche: characterization, propagation and modeling. *Journal of Volcanology*. 193, 93–105.

Moss RH, Edmonds JA, Hibbard KA, Manning MR, Rose SK, van Vuuren DP *et al.* 2010. The next generation of scenarios for climate change research and assessment. *Nature*. 463, 747–756.

Near TJ, Benard MF. 2004. Rapid allopatric speciation in logperch darters (Percidae: Percina). *Evolution*. 58, 2798–2808.

Nichols DL, Pool CA. 2012. The Oxford Handbook of Mesoamerican Archaeology - Google Books. *Oxford Univesity Press*.

O'Hara SL, Metcalfe SE, Street-Perrott FA. 1994. On the arid margin: The relationship between climate, humans and the environment. A review of evidence from the highlands of central Mexico. *Chemosphere*. 29, 965–981.

O'Hara SL, Metcalfe SE. 1995. Reconstructing the climate of Mexico from historical records. *The Holocene*. 5, 485–490.

O'Hara SL. 1993. Historical evidence of fluctuations in the level of Lake Pátzcuaro, Michoacán, Mexico over the last 600 years. *Geographical Journal*. 159, 51–62.

Ortega B, Vázquez G, Caballero M, Israde I. 2010. Late Pleistocene: Holocene record of environmental changes in Lake Zirahuen, Central Mexico - Springer. *Journal of Paleolimnology*. 44, 745-760.

Ownby S, Delgado Granados H, Lange RA, Hall CM. 2007. Volcán Tancítaro, Michoacán, Mexico, 40Ar/39Ar constraints on its history of sector collapse. *Journal of Volcanology and Geothermal Research*. 161, 1–14.

Palumbi SR, Martin AP, Romano S *et al*. *Simple fool's guide to PCR*. University of Hawaii, Honolulu. 1991.

Peer BD. 2011. Invasion of the emperor's grackle. *Ardeola*. 58, 405-409.

Perdices A, Bermingham E, Montilla A, Doadrio I. 2002. Evolutionary history of the genus *Rhamdia* (Teleostei: Pimelodidae) in Central America. *Molecular Phylogenetics and Evolution*. 25, 172–189.

Pérez-Rodríguez R, de León GP-P, Domínguez-Domínguez O, Doadrio I. 2009. A new species of *Algansea* (Actinopterygii: Cyprinidae) from the Ameca River basin, in Central Mexico. *Revista mexicana de biodiversidad*. 80, 483–490.

Phillips SJ, Anderson RP, Schapire RE. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modeling*. 190, 231–259.

Phillips SJ, Dudík M. 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography*. 2008;31, 161–175.

Posada D. 2003. Selecting models of evolution. The phylogenetic handbook. A practical approach to DNA and protein phylogeny. *Cambridge University Press, Cambridge*. 256–282.

Posada D. 2008. jModelTest: Phylogenetic Model Averaging. *Molecular Biology and Evolution*. 25, 1253–1256.

Pritchard JK, Sthepens M, Donnely P. 2000. Inference of populations structure using multilocus genotype data. *Genetics*. 155, 945-959.

Rambaut A, Drummond A. 2007. Tracer v1.5. Available: http. tree bio ed ac uk/software/tracer.

Ramírez-Herrejón JP, Zambrano L, Mercado-Silva N, Torres-Téllez A, Pineda-garcía F, Caraveo-Patiño J *et al.* 2014. Long term changes in the fish fauna of lago de Pátzcuaro in central México. *Latin American Journal of Aquatic Research*. 42, 137-149.

Regan CT. 1906-1908. Pisces. *Biología Centrali-Americana*. 8, 1-203.

Rice WR. 1989. Analyzing tables of statistical tests. *Evolution*. 223–225.

Ritchie MG, Hamil RM, Graves JA, Magurran AE, Webb SA, Macías García C. 2007. Sexual differentiation: population genetic divergence and sexual dimorphism in mexican goodeid fish. *Journal of Evolutionary Biology*. 20, 2048-2055.

Robles-Camacho J, Corona-Chávez P, Morales-Gámez M, Guzmán AF, Polaco ÓJ, Domínguez-Vázquez G *et al.* 2010. Estratigrafía y paleoambiente asociados a un Gomphoteriidae (*Cuvieronius hyodon*) en Tzintzuntzan, Michoacán, México. *Revista Mexicana de Ciencias Geológicas*. 27, 530–544.

Rosenfeld JA, Payne A, DeSalle R. 2012. Random roots and lineage sorting. *Molecular Phylogenetics and Evolution*. 64, 12–20.

Roskamp H. *La historiografía indígena de Michoacán.* Research School Cnws. 1998.

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

Seyfert AL, Cristescu MEA, Frisse L, Schaack S, Thomas WK, Lynch M. 2008. The rate and spectrum of microsatellites mutation in *Caenorhabditis elegans* and *Daphnia pulex*. *Genetics*. 178, 2113-2121.

Shimoda N, Knapik EW, Ziniti J, Sim C, Yamada E, Kaplan S *et al.* 1999. Zebrafish genetic map with 2000 microsatellite markers. Genome. 58, 219–232.

Silvestro D, Michalak I. 2012. raxmlGUI: a graphical front-end for RAxML. *Organisms Diversity & Evolution*. 12, 335–337.

Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*. 22, 2688–2690.

Street-Perrott FA, Perrott RA, Harkness DD. 1989. Anthropogenic soil erosion around Lake Patzcuaro, Michoacan, Mexico, during the Preclassic and late Postclassic-Hispanic periods. *American Antiquity*. 54, 759–765.

Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*. 123, 585–595.

Tamayo LT, West RC. 1964. The hidrology of middle America. In Wanchope R, West RC (eds): Handbook of Middle America Indians. *University of Texas Press*, US.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*. 28, 2731–2739.

Taylor KE, Stouffer RJ, Meehl GA. 2012. An overview of CMIP5 and the experiment design. *Bulletin of the american Meteorological Society*. 93, 485-498.

Telles JL. 2011. The unknown extent of ancient bird introductions. *Ardeola*. 58, 399-404.

Vázquez G, Ortega B, Davies SJ, Aston BJ. 2010. Registro sedimentario de los últimos *ca.* 17000 años del lago de Zirahuén, Michoacán, México. Boletín de la *Sociedad Geológica Mexicana*. 62, 325–343.

Villaseñor-Gómez LE. *La biodiversidad en Michoacán: Estudio de estado*. 2005.

Webb SA, Graves JA, Macias-Garcia C, Magurran AE, Foighil DÓ, Ritchie MG. 2004. Molecular phylogeny of the livebearing Goodeidae (Cyprinodontiformes). *Molecular Phylogenetics and Evolution*. 30, 527–544.

Weber JL, Wong C. 1993. Mutation of human short tandem repeats. *Human molecular genetics*. 2, 1123–1128.

Williams E, Weigand PC. *Las cuencas del occidente de México*. Colegio de Michoacán. 1996.

Williams E. *El antiguo Occidente de México: Un area cultural mesoamericana. Fundación para el avance de los estudios mesoamericanos* FAMSI. INC. 2003.

Wright S. 1951. The genetical structure of populations. *Annals of Eugenics*. 15, 323–354.

Yu Y, Than C, Degnan JH, Nakhleh L. 2011. Coalescent histories on phylogenetic networks and detection of hybridization despite incomplete lineage sorting. *Systematic Biology*. 60, 138–149.

Yue GH, David L, Orban L. 2007. Mutation rate and pattern of microsatellites in common carp (*Cyprinus carpio* l.). *Genetica*. 129, 329–331.

Zambrano L, Cordova-Tapia F, Ramírez-Herrejón JP, Mar-Silva V, BustamanteL, Camargo T *et al*. 2011. Las especies exóticas en el lago de Pátzcuaro, Michoacán, México. In Huerto-Delgadillo R, Vargas-Velázquez S, Ortiz-Paniagua CF (eds): Estudio Ecosistémico del lago de Pátzcuaro. Aportes en gestión ambiental para el fomento del desarrollo sustentable. Instituto Mexicano de Tecnología del Agua, México. 133-156.

Zardoya R, Doadrio I. 1999. Molecular evidence on the evolutionary and biogeographical patterns of european Cyprinids. *Journal of Molecular Evolution.* 49, 227–237.

Figures

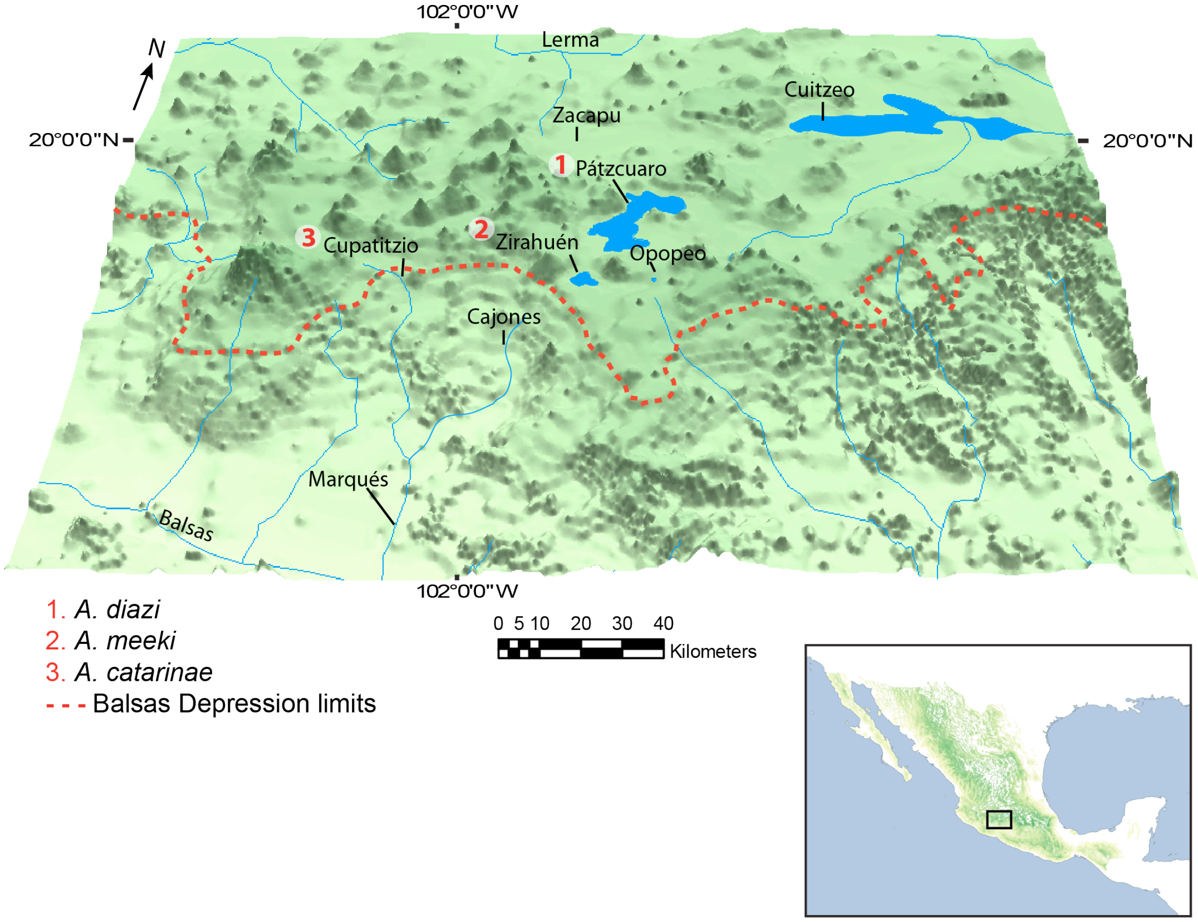


Figure 1. Geographical distribution of *Allotoca diazi* complex.

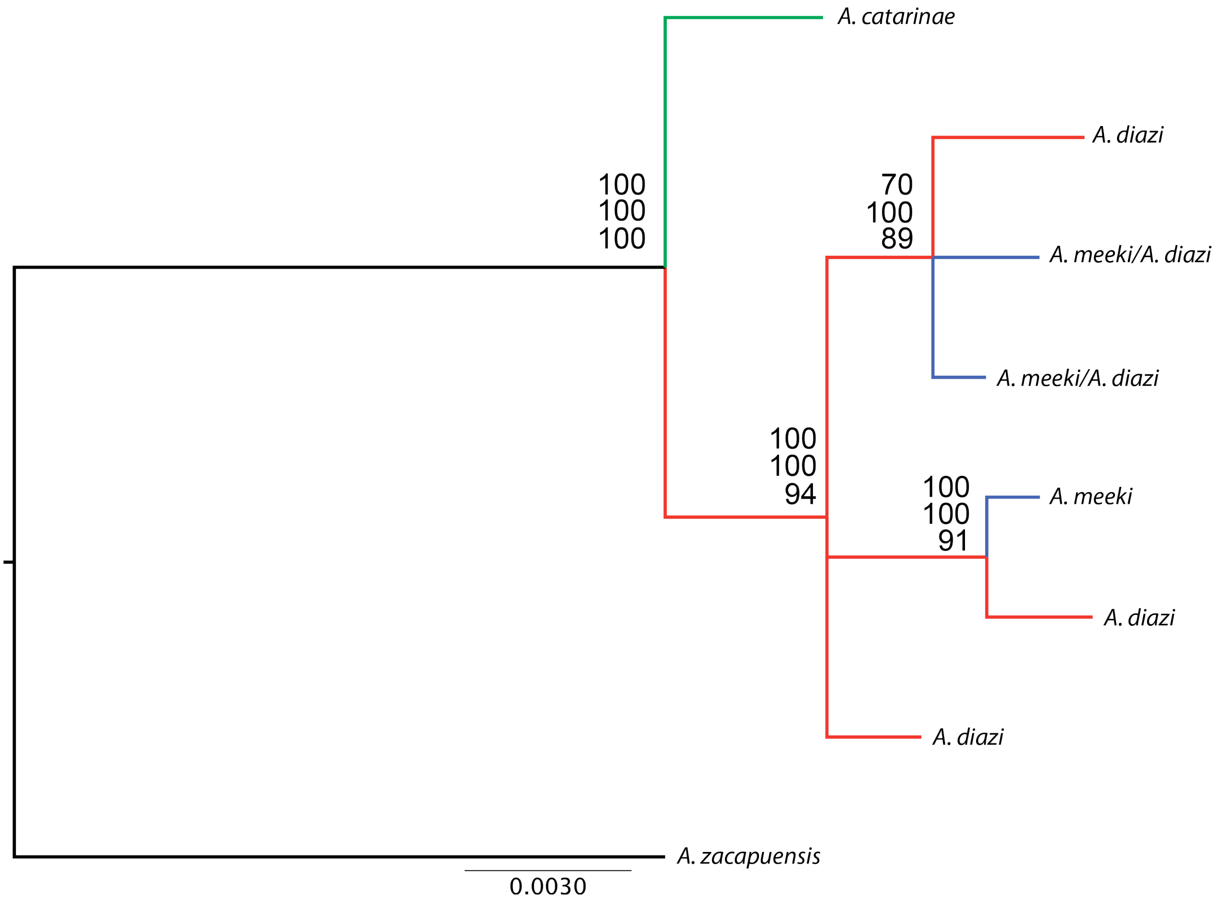


Figure 2. Phylogenetic inference based on haplotypes using Neighbor-joining, Maximum likelihood and Bayesian inference of *cytb* gene. v*Allotoca diazi* (n=35 Hn=5), v*A. meeki* (n=22 Hn=3) and v*A. catarinae* (n=52). Support values are represented by numbers above the nodes for analyses in the order mentioned above.

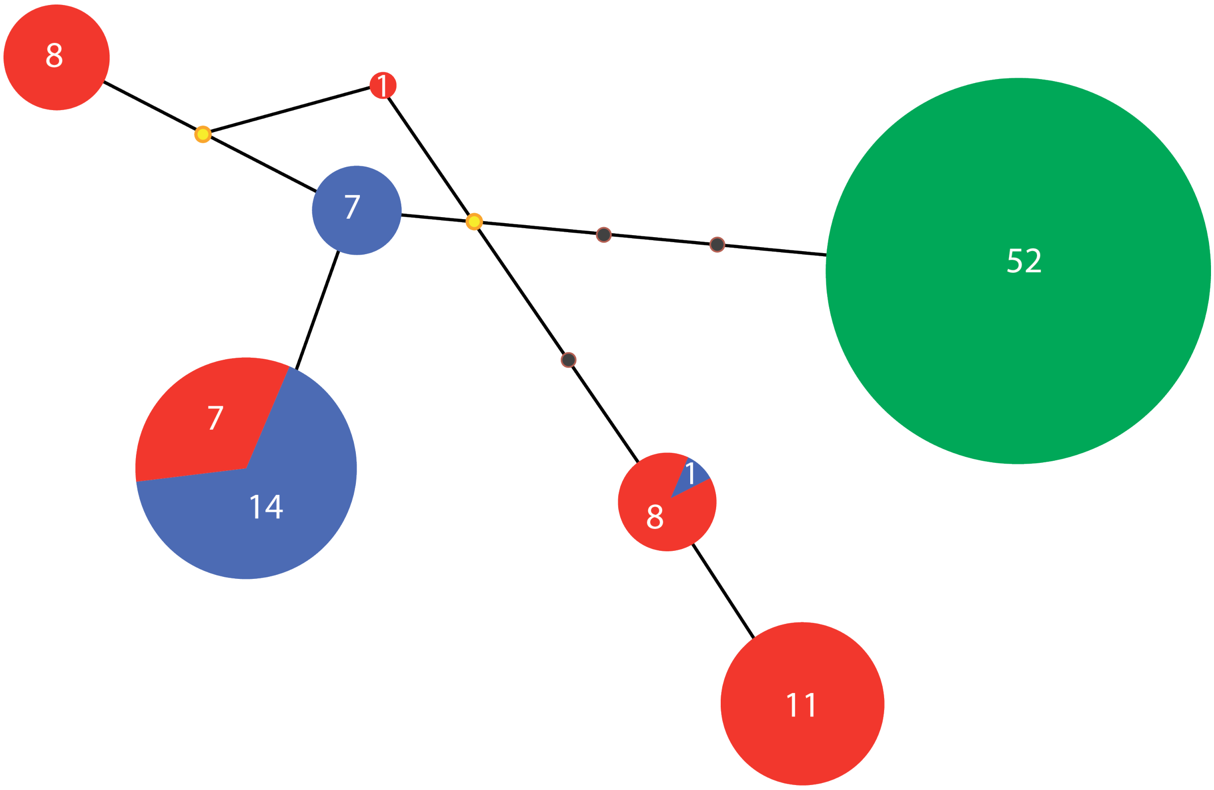


Figure 3. Haplotype network based on *cytb* gene: vA. *diazi* (n=35 Hn=5), vA. *meeki* (n=22 Hn=3) and vA. *catarinae* (n=52). Values represent the number of individuals per haplotype. The yellow circles represent the median-vectors.

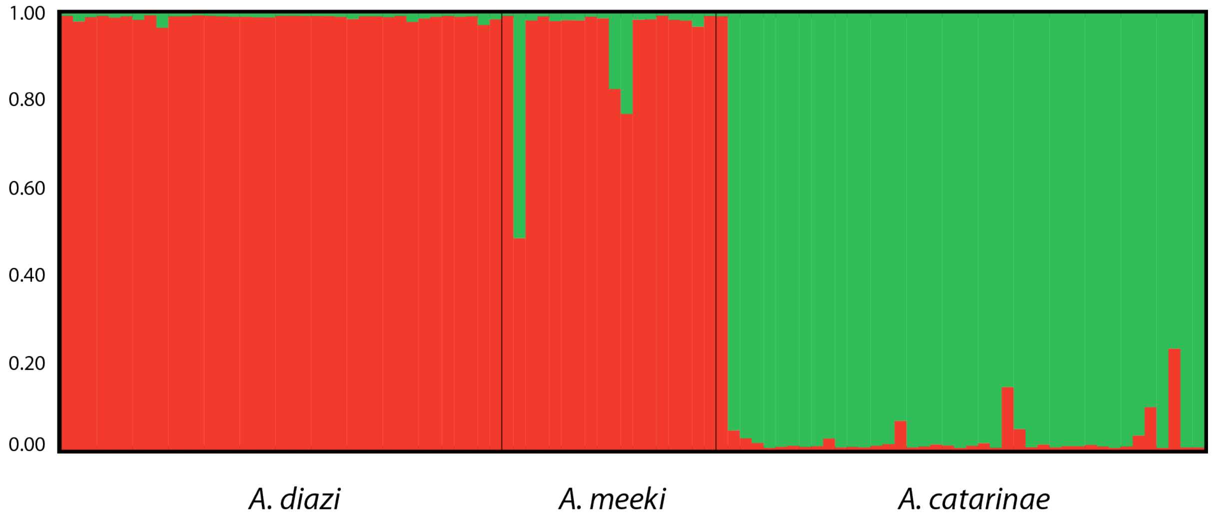


Figure 4. Genetic structure by Bayesian inference based on microsatellite loci.

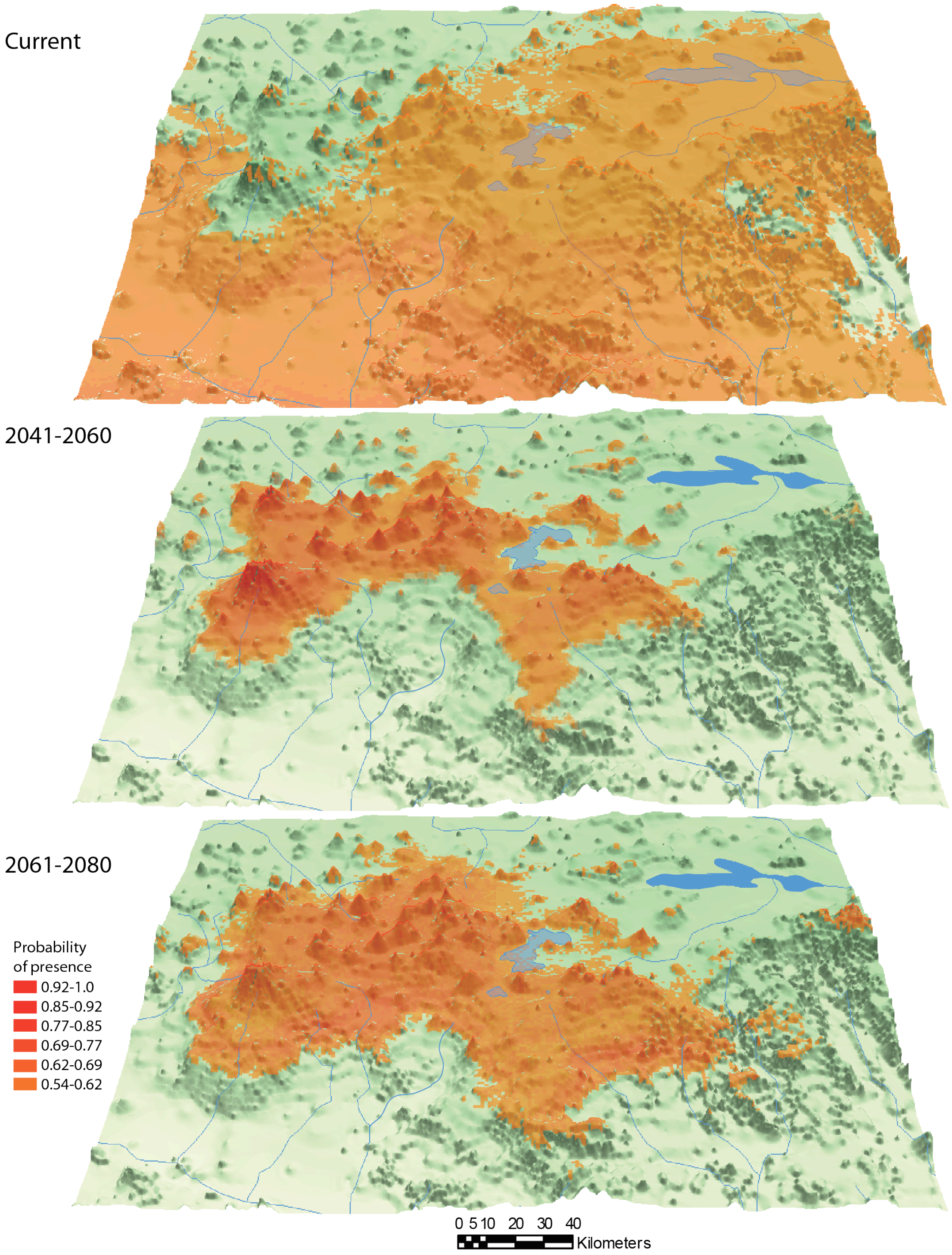


Figure 5. Ecological niche projections for *Allotoca diazi* in present and two future periods.

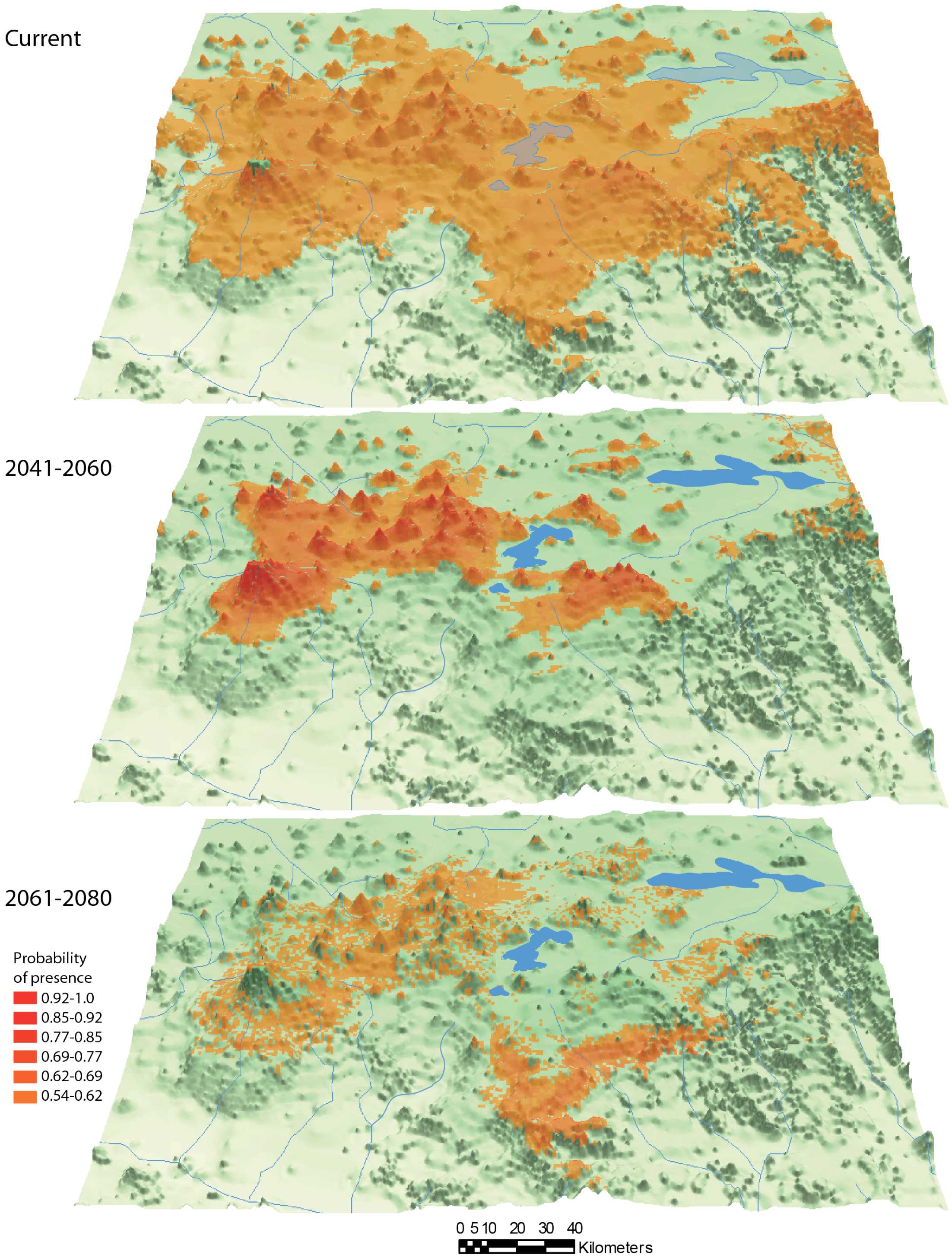


Figure 6. Ecological niche projections for *Allotoca meeki* in present and two future periods.

Tables

Table 1. Genetic diversity parameters and neutrality test based on *cytb* gene sequences

Species	**n**	**Hn**	**Populations**	**π**	**h**	**Θ_S**	**Θ_L**	**FS**	**D_T**
A. diazi	35	5	1	0.0030	0.78	0.0016	1.70	2.07ns	1.61ns
A. meeki	22	3	1	0.0007	0.52	0.0010	1.10	0.69ns	-0.87ns
A. catarinae	52	1	4	0	0	0	0	-	-

n=sample size, Hn=haplotype number, π=nucleotide diversity, h=haplotype diversity, Θ_S=proportion of segregating sites per site, Θ_L=segregate sites per locus, FS=Fu's FS test, D_T=Tajima's *D* test, ns=no significant deviation of neutrality test (P>0.05).

Table 2. Genetic diversity, HW equilibrium, and bottleneck test for microsatellite loci

Species	Locus	H_o	H_e	HW	F_{IS}	BNK IAM/TPM/SMM
A. diazi	XC18	0.84	0.86	--	0.141	-/-/*
	ZT1.6	0.84	0.78	--		
	ZT1.7	0.73	0.88	<0.001		
	IW196	0.62	0.76	<0.05		
	XC25	0.30	0.46	<0.001		
	AS2	0.60	0.81	<0.001		
	ZT1.9	0.70	0.83	<0.05		
A. meeki	XC18	0.83	0.90	--	0.088	*/-/-
	ZT1.6	0.78	0.92	--		
	ZT1.7	0.89	0.88	--		
	IW196	0.72	0.88	--		
	XC25	0.67	0.67	--		
	AS2	0.78	0.88	--		
	ZT1.9	0.83	0.90	--		
A. catarinae	XC18	0.71	0.89	<0.001	0.151	*/-/*
	ZT1.6	0.66	0.74	--		
	ZT1.7	0.81	0.87	<0.05		
	IW196	0.59	0.61	--		
	XC25	0.22	0.20	--		
	AS2	0.59	0.79	<0.001		
	ZT1.9	0.66	0.93	<0.001		

Ho= observed heterozygosity; He=expected heterozygosity; HW= Hardy Weinberg deviation test; FIS=endogamy coefficient (Bold letters-*P*<0.05). Identifying bottleneck (BNK) with Wilcoxon test (*P*<0.05) based on infinite allele model (IAM), two phases model (TPM), and stepwise model (SMM) (* significant; - non-significant). The *P*-values were corrected with Bonferroni method.

Table 3. Estimated genetic differentiation based on microsatellites

	A. diazi	***A. meeki***	***A. catarinae***
A. diazi	--	5.5	3.3
A. meeki	0.084*	--	4.1
A. catarinae	0.133*	0.109*	--

Pairwise F_{ST} values (under the diagonal) and number of migrants (Nm) (above the diagonal). * Significance value *P*<0.05 after Bonferroni correction.

Table 4. Divergence times based on mitochondrial *cytb* gene and microsatellite loci in an Isolation-with-Migration model

	HiPk years	**Divergence times years (HPD)**	**2Nm**
A. diazi			
A. diazi/A. meeki	1702	414-6760	0.0-1.8ns
A. meeki			
A. diazi/A. catarinae	1886	690-8048	0.3-1.4ns
A. catarinae			
A. meeki/A. catarinae	2529	830-8030	0.4-1.4ns

HiPk=highest peak ns=no significance (*P* > 0.05); number of migrants (Nm); Highest Posterior Density (HPD) interval.

Supplementary information

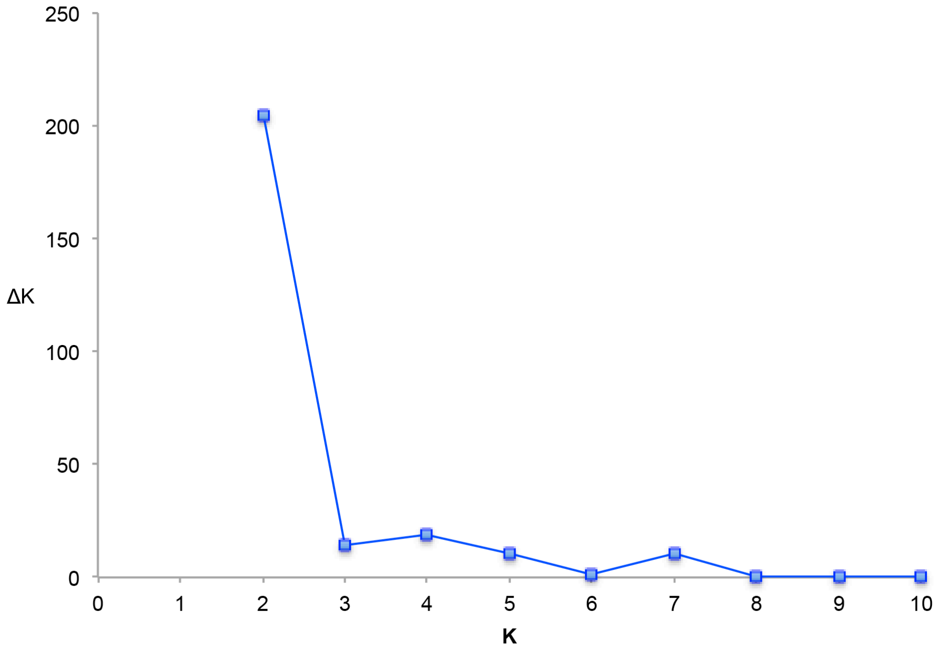


Figure 1S. Number of groups (K) to explain the genetic structure by ΔK criteria.

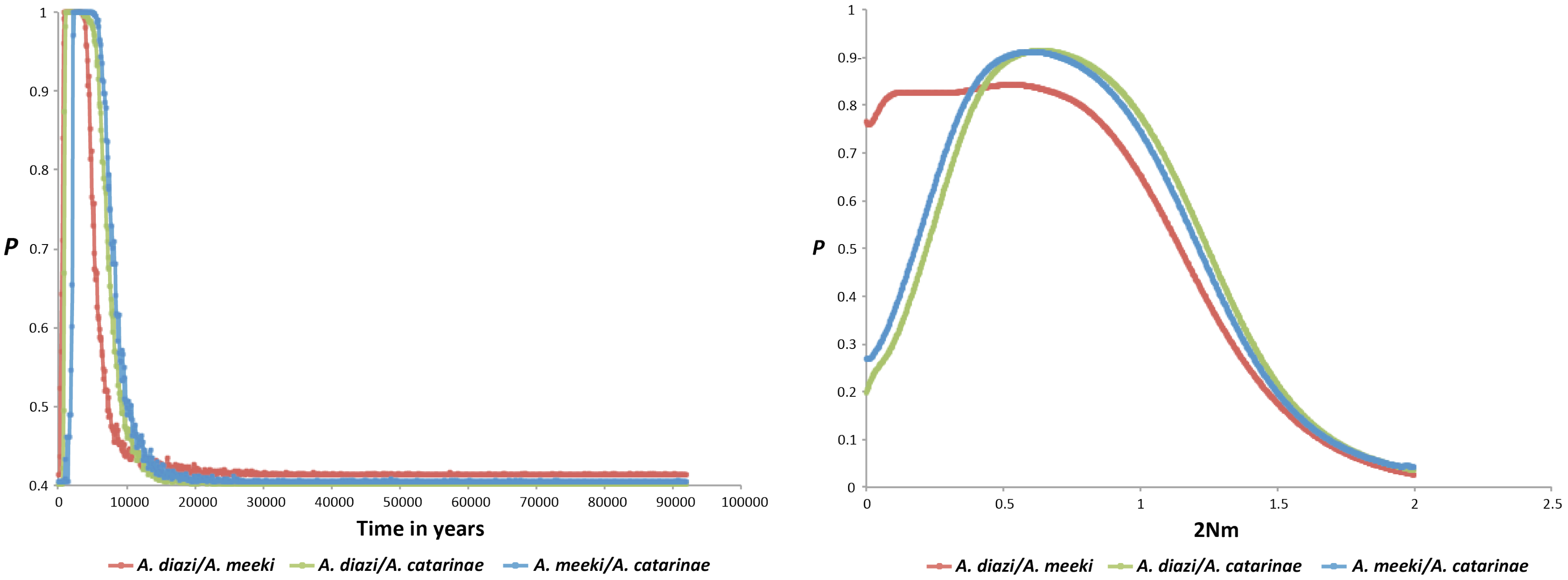


Figure 2S. Marginal distribution with Isolation-with-Migration model of three scenarios of ancestry based on mitochondrial *cytb* gene and microsatellite loci (ESS>100 000).

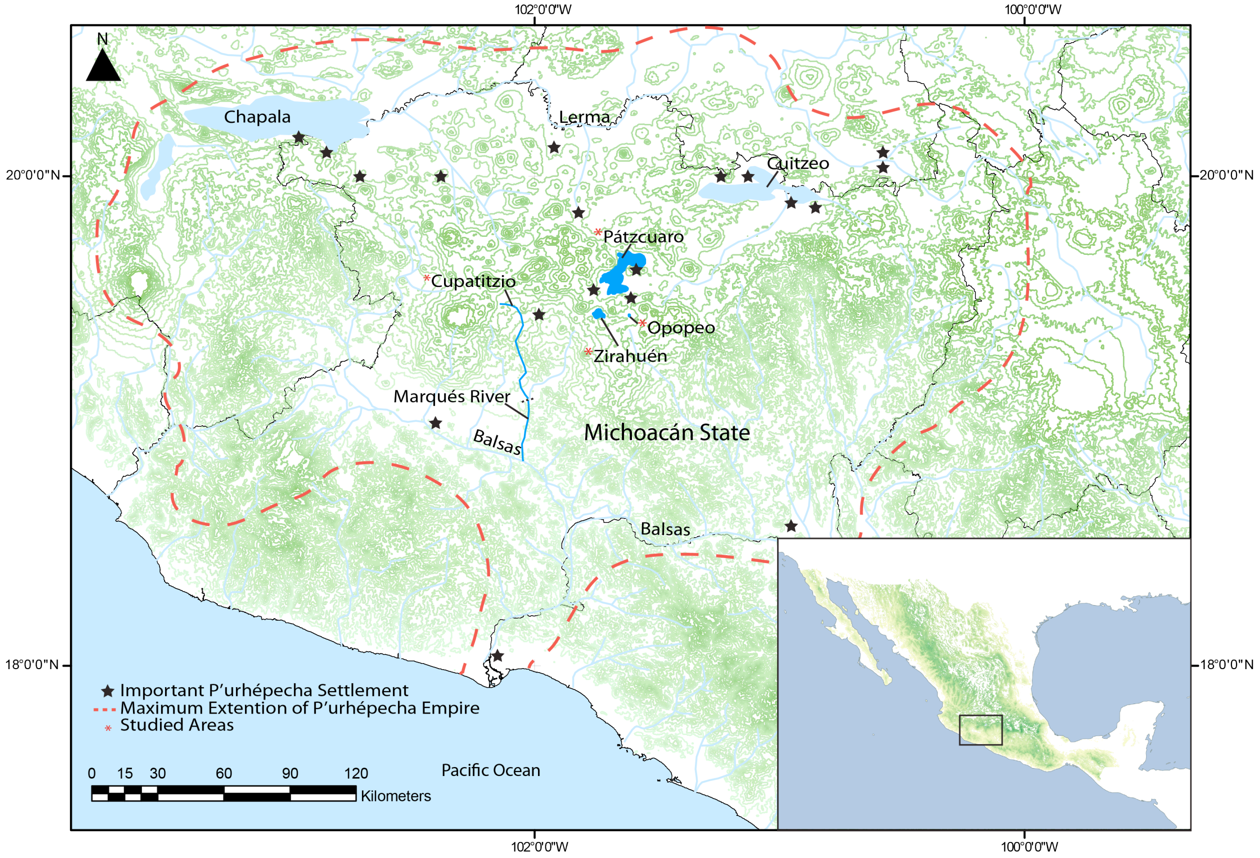


Figure 3S. Extension of the P'urhépecha Empire during the period ca. 1450-1520 A.D. Modified from [1].

.

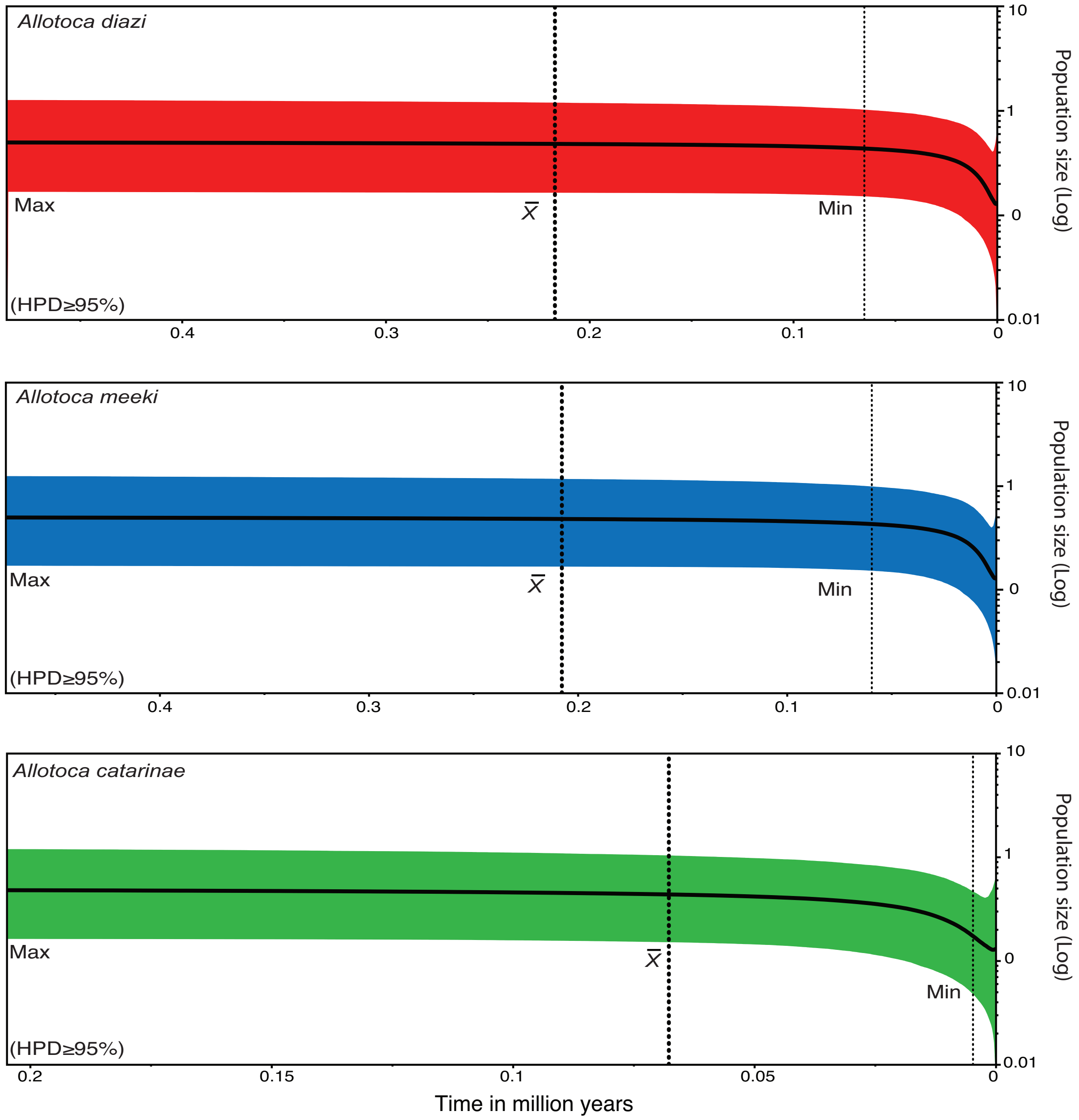


Figure 4S. Demographic history of the members Allotoca diazi complex using Bayesian skyline plots from *cytb* sequences. Dotted lines represent the location of the upper bound (Max), the mean ($\bar{X}$) and lower bound of the HPD≥95%.

Table 1S. Localities and sample size.

Species	n *cytb*	n Microsatellites	Localities	GenBank accession number	Coordinates
A. catarinae	13	9	Santa Catarina Dam	KJ776467-KJ776479	N 19°25’18.39’’ W 102°00’59.9’’
	15	15	Urban Ecological Park	KJ776493-KJ776507	N 19°23’23.7’’ W 102°00’51.8’’
	11	9	El Sauco Stream	KJ776508-KJ776518	N 19º20’58.4’’ W 104º04’50.6’’
	13	9	Cupatitzio Dam	KJ776480-KJ776492	N 19°20’33.4’’ W 102°08’00.8’’
A. diazi			Pátzcuaro Lake (possibly extirpated)		
	35	37	Chapultepec Spring	KJ776519- KJ776553	N 20º34’19.1’’ W 101º31’29.3’’
A. meeki			Zirahuén Lake (extirpated)		N 19º27’00’’ W 101º45’00’’
	22	22	Opopeo	KJ776554- KJ776575	N 19º24’22’’ W 101º36’13.5’’

Table 2S. Selected polymorphic microsatellite loci.

Locus	Repetition	Primer sequence	Length
ZT1.6	$(TG)_{21}$	F GCACCTCCTAATACCTCACA	240-306
		R AAGCAGCAGTACTGTGAGCT	
ZT1.7	$(GGAT)_{12}$	F TTGCTATATTGTCAGTCT	130-250
		R ACTGAAATACTGTTTGTGT	
ZT1.9	$(GGAT)_{30}$	F CTGATGGTTTGAATCTATGC	392-487
		R TGAAGAAGCTGTCTGAGAGG	
XC18	$(CA)_{14}$	F AGGTCATCCGGAGAAAGGT	519-625
		R TTTCACAATCAAACCCCATGT	
XC25	$(AC)_{16}$	F CCTATTTGGGCTCCTGCTG	178-196
		R CTTTCTTCCGTCCGGACTCT	
AS2	$(AC)_{10}G(CA)_{22}$	F CGAAAGCTGTTGTCCTCCT	220-275
		R TTGTGATGCAAAAGCAGTCGT	
IW196	$(CA)_{13}$	F GGGGAGACAGCAAATGAGAA	197-251
		R TATAGCGGGGCTTTTCACAC	

Table 3S. Evolutionary substitution model from mitochondrial *cytb* gene by Akaike Information Criterion (AIC).

Criterion	**Position**	**Evolutionary Model**	**-lnL**	**Substitution rate model Ti/Tv**	**Nucleotide empiric base frequency**	**Gamma distribution**	**Invariable sites proportion**
AIC	complete fragment	TIM+I+G	-1702.006	--	*A*=0.2476 *C*=0.2928 *G*=0.1349 *T*=0.3248	0.1490	0.9430

-lnL=log likelihood, Ti=Transitions, Tv=Transversions.

Table 4S. Variables included in ecological niche models.

Species	Period in years ~	RM	BIOvariables included	AUC
A. diazi	Present	0.01	1-15 and 17-19	0.876
	2041-2060	0.01	7, 10, 14 and 19	0.940
	2061-2080	0.01	5, 10, 11, 12, and 15	0.986
A. meeki	Present	1	1-15 and 17-19	0.932
	2041-2060	0.01	5, 7, 9 and 19	0.965
	2061-2080	0.01	5, 9, 10 and 19	0.995

RM= Regularization multiplier value selected according to the of AUC value.

Table 5S. Estimated allelic diversity by locus and species.

Locus		**N_a**	**N_{ae}**	**P_{an}**	**LD**
XC18		14.0±3.6	7,6±0.9	0.05	--
ZT1.6		10.0±3.6	5.9.0±3.3	0.04	--
ZT1.7		12.0±2.0	7.0±0.3	0.04	--
IW196		9.3±2.1	4.4±2.1	0.06	--
XC25		5.0±3.0	2.0±0.8	0.07	--
AS2		7.3±1.5	5.0±1.6	0.12	--
ZT1.9		15.0±1.7	8.6±3.3	0.07	--
Total		10.4±3.5	5.8±2.2	0.06±0.03	
Species	**n**	**Na**	**Nae**	**Pna**	**DL**
A. diazi	35	9.9±3.0	4.9±1.8	0.10	AS2-ZT1.9*
A. meeki	22	11.3±2.3	6.9±2.1	0.01	--
A. catarinae	52	10.0±6.3	5.5±3.9	0.05	--
Total	96	10.4±0.8	6.2±1.0		

n=sample size, N_a=mean of alleles, N_{ae}=mean number of effective alleles, P_{an}=null allele proportion, LD=linkage disequilibrium test with significant value after Bonferroni correction (*=$P > 0.05$).

Table 6S. Genetic structure based on R_{ST} values compared with F_{ST} of permutation analysis.

Locus	R_{ST}	pR_{ST} (95% I. C.)	F_{ST}
XC18	0.072 ns	0.095 (-0.015-0.338)	0.102*
ZT1.6	0.527 ns	0.139 (-0.012-0.441)	0.146*
ZT1.7	0.266 ns	0.088 (-0.014-0.310)	0.091*
IW196	0.228 ns	0.162 (-0.015-0.513)	0.189*
XC25	0.158 ns	0.083 (-0.016-0.230)	0.088*
AS2	-0.001 ns	0.113 (-0.015-0.356)	0.113*
ZT1.9	0.141 ns	0.048 (-0.018-0.183)	0.051*
All loci	0.178 ns	0.099 (0.026-0.205)	0.113*

H_o $R_{ST}=pR_{ST}$, ns=non-significant, *significant ($P < 0.05$) after Bonferroni correction, R_{ST}=observed values, pR_{ST}=expected values, IC=95% confidence interval.

Table 7S. Genetic structure inferred via Analysis of Molecular Variance AMOVA based on microsatellite data.

Groups	FST	FSC	FCT	% among groups	% within groups	*P*
(*A. diazi*) (*A. meeki*) (*A. catarinae*)	0.113	-	-	-	-	<0.05
(*A. diazi*, *A. meeki*) *A. catarinae*	0.117	0.086	0.038	3.27	8.28	0.33
***A. diazi* (*A. meeki*, *A. catarinae*)**	0.118	0.106	0.011	1.09	10.43	0.67
(*A. diazi*, *A. catarinae*) *A. meeki*	0.120	0.127	-0.036	3.55	13.19	1

References

Williams E. 2003. El antiguo Occidente de México: Un area cultural mesoamericana. Fundación para el avance de los estudios mesoamericanos FAMSI. INC.

Capítulo 2: Phylogenetic and phylogeographic analysis of the *Catostomus plebeius-nebuliferus* complex (Actinopterygii: Catostomidae)

DIUSHI KERI CORONA-SANTIAGO[1], LLANET TOVAR-MORA[2], RODOLFO PÉREZ-RODRÍGUEZ[2], YVONNE HERRERÍAS-DIEGO[3], JOSÉ RAMÓN PARDOS-BLAS[1], OMAR DOMÍNGUEZ-DOMÍNGUEZ[2] and IGNACIO DOADRIO[1]

[1] Departamento de Biodiversidad y Biología Evolutiva, Museo Nacional de Ciencias Naturales, CSIC. c/José Gutiérrez Abascal, 2. 28006 Madrid, Spain.

[2] Laboratorio de Biología Acuática, Facultad de Biología, Universidad Michoacana de San Nicolás de Hidalgo, Morelia, Michoacán, México.

[3] Facultad de Biología, Universidad Michoacana de San Nicolás de Hidalgo, Morelia, Michoacán, México.

For Submission in: Molecular Phylogenetics and Evolution

Abstract

The Sierra Madre Occidental mountain range from Mexico is a region with a high number of endemic species. We conducted a phylogenetic and phylogeographic study of the *Catostomus plebeius-nebuliferus* complex, a group of primary freshwater fishes widely distributed in the Sierra Madre Occidental range that provides a satisfactory model for evaluation of the influence of geological and climatic events on species evolution. Mitochondrial *cytb* and the 3rd nuclear intron of the *GHI* gene were sequenced. Four well-differentiated lineages within the species complex were identified, with divergence ranging from 5.9 to 9.2% for *cytb* and 0.1% to 0.9% for *GHI*. Evolutionary history of the complex is characterized by vicariants post-dispersal process linked to geological episodes in the Sierra Madre Occidental range and North Central Mexico since the Pliocene. Significant intrapopulation genetic structure was observed in the Nazas, Mezquital, Fuerte, Casas Grandes, and Santa Maria drainage basins for *cytb*. Our results suggest a preliminary hypothesis to clarify the taxonomy of *Catostomus plebeius-nebuliferus* and for establishing conservation units.

Keywords: phylogeny – conservation units – evolutionary history – Mexico – Sierra Madre Occidental – taxonomy.

Introduction

The structure of current North American biodiversity is related to geological (tectonic and volcanic) and climatic changes that occurred mainly during the Neogene (*ca.* 33 Mya) and Quaternary (*ca.* past 2.5 Mya) (Riddle and Hafner, 2006; Aguirre-Díaz *et al.,* 2008). The Sierra Madre Occidental (SMOC) mountain range in Mexico is an area of high endemism of flora and fauna and is considered an important biogeographic corridor (Domínguez-Domínguez *et al.,* 2011). The formation of the SMOC during the Oligocene (*ca.* 33-23 Mya), Early Miocene (*ca.* 23-15 Mya), Late Miocene (*ca.* 13-11 Mya), and Pliocene-Quaternary (*ca.* 5.3-0 Mya) (Ferrari, Valencia-Moreno and Bryan, 2007; Aguirre-Díaz *et al.,* 2008) has been considered to have an important influence on the evolution of several taxa, particularly organisms with low dispersal ability such as freshwater fish species (Rinne, Hughes and Calamusso, 2005; Echelle *et al.,* 2005; Dominguez-Dominguez *et al.,* 2008; Domínguez-Domínguez and Pérez-Ponce de León, 2009; Domínguez-Domínguez *et al.,* 2011; Schönhuth *et al.,* 2012a; 2014; 2015; Pérez-Rodríguez *et al.,* 2016).

Although the biodiversity of northwestern Mexico is mainly related to tecto-volcanic activity of the Tertiary, the SMOC is also considered a Pleistocenic refuge (Ceballos, Arroyo-Cabrales and Ponce, 2010). The Pleistocene climatic change was associated with the expansion and contraction of the distribution of numerous species. The effects of the climate changes on epicontinental waters should have significantly affected the distribution of the fish fauna of this region to the present time (Domínguez-Domínguez *et al.,* 2011). The evolutionary history of the Catostomidae in Mexico may have been influenced by these processes, especially in the case of the Catostominae, the predominant subfamily in Mexico (Wei-Jen and Mayden, 2012).

Within the Catostominae, two closely related species of the genus *Catostomus* Le Sueur, 1817 inhabit northern Mexico including the SMOC: *Catostomus plebeius* Baird and Girard, 1854 and *Catostomus nebuliferus* Garman, 1881 (Doosey *et al.,* 2010; Wei-Jen and Mayden, 2012). *Catostomus nebuliferus* is distributed in the endorheic Nazas and Aguanaval drainages, whereas *Catostomus plebeius* is more widespread and occurs in several basins across the SMOC: the Mezquital, Piaxtla, Fuerte, and Yaqui basins; the Central Guzman hydrographic system (Santa Maria, Casas Grandes, and Del Carmen basins); the Upper Rio Grande basin (Conchos River in Mexico and Rio Grande River in USA); and the Mimbres basin in New Mexico, USA (Fig. 1) (Miller, Minckley and Norris, 2005). It is also distributed in the Gila River of the Colorado basin, but the origin of this population is uncertain and considered to have been artificially introduced (Buth and Crabtree, 1985; Pages and Burr, 2011), while others suggest this population to be the result of stream capture from the Mimbres River (Sublette, Hatch and Subblete, 1990).

As with other *Catostomus* species in Mexico, *C. plebeius* is believed to have colonized several basins during the postglacial period. Dispersal during anastomosis of rivers, followed by vicariance after river isolation, has been hypothesized as the main factor influencing genetic divergence among *Catostomus* populations, even suggesting that some populations, could represent one or more independent evolutionary lineages such as those of the Mezquital, Piaxtla, and Fuerte basins (Ferris, Buth and Whitt, 1982). Significant genetic structure has also been observed among populations of *C. plebeius* in the upper Rio Grande basin when compared Mimbres and Gila River populations (Crabtree and Buth, 1987; McPhee, Osborne and Turner, 2008). The species *Catostomus nebuliferus* has been suggested to be a synonym of *C. plebeius* (Smith, 1992). Both *C. plebeius* and *C. nebuliferus* are considered to belong to the subgenus *Pantosteus* Smith, 1966. However, the phylogenetic relationships of *Pantosteus* are unclear due to incomplete species representation in taxonomic works (Smith, 1966; Sun *et al.,* 2007; Bart *et al.,* 2010). Paraphyletic relationships have also been observed among the species of the genus *Pantosteus* (Doosey *et al.,* 2010; Laitinen, 2012; Unmack *et al.,* 2014).

Phylogeographic studies of *Campostoma ornatum* Girard, 1856, and *Codoma ornata* Girard, 1856, species of Cyprinidae co-distributed with *Catostomus plebeius-nebuliferus* throughout its ranges, have shown a complex reticulated biogeographic history (Domínguez-Domínguez *et al.,* 2011; Schönhuth *et al.,* 2015). This pattern was related to repeat events of dispersal and isolation associated with tecto-volcanic activity occurring in the SMOC since the early Pliocene, as well as to climatic fluctuations during the glacial and interglacial periods. Population structure in *Campostoma ornatum* and *Codoma ornata* is reflected in high genetic differentiation among populations inhabiting basins in the northern SMOC, including the Guzman system, and the Rio Grande River, Yaqui, Mayo, Fuerte, and Nazas basins (Domínguez-Domínguez *et al.,* 2011; Schönhuth *et al.,* 2015).

The aim of this study was 1) to examine the phylogenetic relationships and describe the evolutionary history of the *Catostomus plebeius-nebuliferus* complex in Mexico; 2) to explore its genetic diversity and structure; and 3) to determine species divergence times using nuclear and mitochondrial markers. We assume that the genetic structure and distribution of the main clades of *Catostomus plebeius-nebuliferus* are associated with the geological history of the SMOC.

Materials And Methods

Sampling and DNA isolation

Eighty-seven specimens were collected with permission of relevant authorities in July 2013 from 18 localities in 8 drainage basins (Fig. 1 and Table 1) using electrofishing and hand nets. Despite a considerable sampling effort, no specimens were obtained from the Aguanaval and Yaqui basins. Fin clips were obtained and preserved in absolute ethanol and stored at -75 °C. Several specimens were formalin-fixed for identification and deposited in the fish collection of the Universidad Michoacana de San Nicolas de Hidalgo-CPUM, Mexico. The remaining captured fish were returned to the water.

Isolation of genomic DNA was performed with BioSprint DNA Blood Kit QIAGEN according to the manufacturer's instructions. We amplified the complete cytochrome *b* mitochondrial gene (*cytb*) using the primers GLuDG (Palumbi *et al.,* 1991) and H16460 (Perdices *et al.,* 2002) and the nuclear 3rd growth hormone (*GHI*) intron of 654bp, using primers designed for *Ictiobus bubalus* (Catostomidae) (Clements, Bart and Hurley, 2012).

The PCR products were checked by electrophoresis on 1.5% agarose gel and submitted to Macrogen Inc. (Korea) and htSEQ Inc (High-Throughput Sequencing, University of Washington, USA) for sequencing.

Phylogenetic analysis and haplotype network reconstruction

Sequences were manually aligned in Mega v. 5.2 (Tamura *et al.,* 2011) and examined using chromatograms. For *GHI* all the sequences were phased using DNAsp v. 5.0 (Librado and Rozas, 2009) and a test of recombination was applied using a coalescent algorithm (10000 replicates).

For phylogenetic reconstruction, two building algorithms were used, and a sister species of the complex available for the two genes on GenBank, *Catostomus catostomus* (GenBank Assesion: AF454871 for *cytb* and GU937824 for *GHI*), was used as outgroup. Independents Maximum likelihood (ML) analyses were conducted with RAxMLGUI v. 1.3.1 (Stamatakis, 2006; Silvestro and Michalak, 2012) for both genes, performing 10 000 bootstrap repetitions and using the evolutionary substitution model estimated with PartitionFinder (Lanfear *et al.,* 2012). These model were General Time-Reversible (Tavaré, 1986) + gamma (GTR+G) (Table 1S of the Supplementary information) for the complete *cytb* and Tamura-Nei (Tamura and Nei, 1993) + gamma (TrN+G) for *GHI*. Bayesian phylogenetic (BI) reconstruction was conducted with the software MrBayes v. 3.2.6 (Ronquist and Huelsenbeck, 2003), using the above selected evolutionary substitution model and implementing two reactions for four Markov Chain Monte Carlo (MCMC) processes with 7 million generations for *cytb* and 15 million generations or *GHI*, sampling every 100 generations in both cases. We evaluated the convergence of the log-likelihood (-InL) value of the two reactions, with 10% of reconstructions discarded as burn-in, to construct the consensus tree ($\sigma < 0.005$). The posterior probabilities obtained based on a confidence limit of 95% (highest posterior density-HPD) were used to evaluate the support of nodes. Concatenated analyses of nuclear and mitochondrial data for ML and BI were performed using the same parameter previously mentioned. Haplotype networks were constructed for both genes using PopART (available at htt://popart.otago.ac.nz) and applying the Median-Joining method (Bandelt, Forster and Röhl, 2000).

Genetic variation

The genetic diversity within drainage basins was estimated for the two markers based on nucleotide diversity (π), haplotype diversity (h), and the proportion of segregating sites (Θ_S), using the software DNAsp v. 5.0. The genetic *p*-distances (D_P) between populations were obtained with Mega v. 5.2, and a bootstrapping process was implemented with 1000 repetitions.

The genetic structure at the interspecific and intraspecific level of the *Catostomus plebeius-nebuliferus* complex was estimated with paired fixation indices (Φ_{ST}) for both molecular markers. The genetic structure of different hierarchical arrangements, considering previous information, was also tested, and significance was assessed by analysis of molecular variance (AMOVA) with 10 000 permutations implemented in the software Arlequin v. 3.5.1.3 (Excoffier and Lischer, 2010). To define groups of populations that were maximally differentiated without constraint of their geographic distribution, we used spatial analysis of molecular variance of geographically homogeneous K groups (SAMOVA) with SAMOVA v. 2.0 (Dupanloup, Schneider and Excoffier, 2002), implementing 10 000 iterations from each of 100 random initial conditions, tested for each K value from 1 to 8. A Bonferroni correction (Rice, 1989) was applied to each *p*-value obtained in the paired test of genetic differentiation Φ_{ST}.

Divergence time estimation and ancestral area reconstruction

The divergence times were estimated by a coalescence method using the software BEAST v. 1.8.0 (Drummond *et al.,* 2007), based on the described evolutionary substitution model for both genes (Table 1S). A relaxed molecular clock, using a Extended Bayesian Skyline (including nuclear an mitochondrial data), was calibrated using the fossil record, which was located in the root of the *Catostomus plebeius-nebuliferus* complex, represented by fossils of *Catostomus asitus* (Smith, Stewart and Carpenter, 2013), dated from the Miocene-Pliocene (7.5-2.5 Mya) (Arbor, 2013). Another molecular clock analysis was conducted using the mutation rate for *cytb* estimated for teleosts of 0.76-2.2%/Myr in uniform prior in BEAST (Zardoya and Doadrio, 1999; Machordom and Doadrio, 2001; Perdices *et al.,* 2001; Near and Benard, 2004).

For both analyses, we implemented three independent analyses with 70 000 000 generations each one, sampling every 1000 generations. After evaluating the posterior parameter values based on effective sample size, and according to the convergence data using Tracer v. 1.6 (Rambaut *et al.,* 2014), 10% of runs were discarded and a maximum clade credibility tree was constructed combining the three analyses.

Ancestral area reconstruction was implemented in RASP v. 3.2, 20160719 (Yu *et al.,* 2015) and applying Statistical-Dispersal Vicariance Analysis (S-DIVA) (Yu, Harris and He, 2010) and the Statistical Dispersal-Extintion-Cladogenesis model (S-DEC) (Beaulieu, Tank and Donoghue, 2013). In both cases the chronograms obtained in BEAST were used but resampling each 3000 samples for use a total of 70 000 trees. Nine were considered in the analysis corresponding mainly to the basin sampled and in this work and taking into account the difficult to determine the species boundaries in the complex. The areas included were the basins: A, Palomas; B, Casas Grandes; C, Santa Maria; D, Del Carmen; E, Fuerte; F, Conchos; G, Piaxtla; H, Nazas; and I, Mezquital. An unconstrained model was applied allowing any combination of geographic range in adjacency matrix with a maximum of five adjacent hydrographic basins, and configuring the dispersal rate with the same probability for each population of the Catostomus plebeius-nebuliferus complex.

Results

A total of 92 DNA sequences of the cytb gene (1140bp) and 82 of the 3rd intron of GHI were obtained for Catostomus plebeius (n = 63 of cytb and n = 49 of GHI) and C. nebuliferus (n = 29 of cytb and n = 33 of GHI) from 22 localities in 8 drainage basins, representing most of the species distribution range (Miller et al., 2005). Eighty-seven and eighty-two were new sequences for cytb and GHI respectively; five of cytb were retrieved from GenBank (Fig. 1 and Table 1). No significant recombination was detected in the nuclear GHI sequences (p > 1).

Phylogenetic relationships and haplotype network analysis

Bayesian and maximum likelihood phylogenetic analyses for the *cytb* and *GHI* sequences recover the same topology with highly supported four main lineages (Fig. S1 and S2 of the Supplementary information), and the concatenated analysis is showed here (Fig. 2). Lineage I included samples from the Guzman hydrographic system (Casas Grandes, Santa Maria, and Del Carmen basins) and the Upper North Rio Grande system (Las Palomas). Lineage II comprised the populations of the Fuerte basin and Conchos River (Upper South Rio Grande basin). Lineage III, corresponding to the Nazas and Piaxtla basin populations and Lineage IV consisted of the population from the Mezquital basin.

Four haplogroups were recovered in the haplotype network reconstruction for both molecular markers, corresponding to the four primary lineages obtained by phylogenetic analysis, but with different relationships in *cytb*, placing the Mezquital population (Lineage IV) closely related to the Guzman System/Palomas populations (Lineage I), whereas the Nazas/Piaxtla populations (Lineage III) were more closely related to the Conchos/Fuerte populations (Lineage II) than to the Mezquital population (Fig. 3 A and B). In the case of *GHI*, the range of mutations steps (MS) between lineages was 1-2 MS, geographic correspondence was observed, and shared haplotypes were identify between populations related phylogenetically (Casas Grandes/Del Carmen/Santa Maria/Palomas, Fuerte/Conchos and Piaxtla/Nazas) (Fig. 3 B).

In contrast, no shared haplotypes were observed among lineages or among drainage basins in the case of *cytb* and high number of MS were observed between lineages.

The central groups in the haplotype network were those of the Guzman System/Palomas (Lineage I) and the Fuerte/Conchos (Lineage II), which were separated by 52 MS. Lineage I was separated from the Mezquital (Lineage IV) by 72 MS, and Lineage II was also separated from Nazas/Piaxtla (Lineage III) groups by 72 MS. The number of MS separating the populations of Fuerte from the Conchos River and the Nazas from Piaxtla basins was 3 MS in both cases. We did not include as MS the median vectors or hypothetical haplotypes implied by the haplotype network reconstruction.

Genetic diversity

Although the haplotype diversity in all basins was relative high (h = 0.829-0.981) for *cytb* and *GHI*, the nucleotide diversity and the proportion of segregating sites were low for both markers (π = 0.001-0.010 and Θ_S = 0.0005-0.010) (Table 2). The populations with the lowest nucleotide diversity (π = 0.003) were those from Fuerte. The Casas Grandes and Nazas basin populations showed the highest number of haplotypes. However, the number of haplotypes was different in each marker by population.

Genetic distances and structure

The highest absolute genetic distances (AGD) for *cytb* found between lineages were between the Mezquital (Lineage IV) and Piaxtla/Nazas basins (Lineage III) (D_P = 9.2%) (Table 3), and higher AGD than 5% between lineages were observed. In contrast, based on *GHI* the AGD were lower than one percent in all the cases and the highest distances (D_P=0.9%) were observed between Mezquital population respect to Fuerte and Del Carmen, and between Piaxtla respect to Del Carmen and Fuerte. The lowest AGD were observed between populations of Casas Grandes and Santa Maria for *cytb* (D_P=1.3%) and between Palomas respect to Casas Grandes and Santa María for *GHI* (D_P=0.1%) in the Lineage I.

The AMOVA revealed significant genetic structure among populations (Φ_{SC} = 0.9028 for *cytb*; Φ_{SC} = 0.6904 for *GHI*; $p < 0.05$) (Table 4). The highest genetic structure with SAMOVA was found when K = 8 (Φ_{CT} = 0.8784 for *cytb*; Φ_{CT} = 0.7460 for *GHI*, $p < 0.05$), and each basin was clustered in an independent group with the exception of one population of the current Upper South Rio Grande, the Conchos population, which was clustered with the Fuerte population. The genetic groups obtained using K = 4 and K = 2 in SAMOVA corresponded to the primary lineages recovered in phylogenetic analysis and the populations considered *C. plebeius* and *C. nebulieferus* currently, with $\Phi_{CT} = 0.7685$ and $\Phi_{CT} = 0.3545$ for *cytb* ($p < 0.05$). In the case of *GHI*, no significant genetic structure was observed when groups of populations considered *C. plebeius* and *C. nebuliferus* are constructed ($\Phi_{CT} = 0.7685$, $p > 0.05$).

High Φ_{ST} and significant pairwise values were observed among basins in both molecular markers, with Φ_{ST} ranging 0.461 to 0.968 for *cytb* and 0.050 to 0.907 for *GHI* (only $p < 0.5$) (Table 5). The highest value was observed between the populations of Piaxtla and Fuerte in *cytb*, while between Piaxtla and Del Carmen populations in *GHI.* The lowest but significant was between Piaxtla and Nazas in *cytb* and between Del Carmen and Santa Maria in *GHI*. In contrast to *cytb* Φ_{ST} distances, no significant Φ_{ST} between populations of Santa Maria and Casas Grandes.

Divergence times and ancestral area reconstruction

The global likelihood scores with the highest probabilities were obtained using S-DEC analysis (Posterior probability = 0.96). However total congruence in ancestral areas predicted was observed between both analyses. The scenario showed by the two models (S-DIVA and S-DEC) indicated that the *Catostomus plebeius-nebuliferus* inhabit and area (ABCDF) by the Guzman system and the Upper South and North Rio Grande (Conchos and Palomas, respectively), which two dispersal process was identified from Conchos to Fuerte and to Nazas basins, followed of a vicariant event give rise the isolation of the Guzmán system/Palomas from Conchos population (EFH) from more likely to have occurred *ca.* 3.9-5.1 Mya at the Pliocene (Highest Posterior Density HPD ≥ 95%) (Fig. 4). Both ancestral areas reconstruction suggest that a range expansion occur from Nazas to Piaxtla and to Mezquital basins followed by a vicariant event that isolated these populations from Conchos/Fuerte about the Late Pliocene-Early Pleistocene (*ca.* 1.9-4-8 Mya). Subsequently, two vicariant events fragmented the ancestor of the remaining members of the Piaxtla/Nazas/Mezquital (GHI; 1.3-3.9 Mya) and Guzman system/Palomas (ABCD; 0.6-3.6 Mya). Pleistocene vicariant events where identified separating population within lineages: Casas Grandes from Palomas *ca.* 1.1 Mya; Santa Maria from Del

Carmen *ca.* 1.2 Mya; Fuerte from Conchos *ca.* 0.6 Mya; and Nazas from Piaxtla *ca.* 1.4 Mya.

Discussion

The main forces shaping the cladogenetic events in the *Catostomus plebeius-nebuliferus* are consistent with geological processes occurring since the Pliocene associated with the formation and evolution of the SMOC, central North Mexico and the tectonic activity of the Rio Grande Rift. As has occurred with *Moxotoma* genus (Pérez-Rodríguez *et al.,* 2016), river capture may have played an important role in expansion range shaping the spatial distribution in some basins. Genetic variation within basins seems to have been influenced by ecological characteristics of the species and climatic fluctuations since the Pleistocene. Our results revealed that the complex comprises four significant genetic structure between main lineages in nuclear and mitochondrial markers indicative of a long history of isolation. However, mitochondrial information shows high divergences and several mutation steps between lineages unlike to nuclear marker that show shared haplotypes between basins (Guzman system) or low mutation steps between lineages probably due to slow mutation rate.

HISTORICAL BIOGEOGRAPHY OF *CATOSTOMUS-PLEBEIUS-NEBULIFERUS*

Pliocene events

The most recent common ancestor for all the lineages of the *Catostomus plebeius-nebuliferus* was calculated at the Pliocene (3.9-5.1 Mya, HPD ≥ 95%) (Fig. 5) in an ancestral area conformed by the proto-Guzman system/Upper Rio Grande (Palomas and Conchos) with a range expansion of the complex to Fuerte basins, probably resulted of a basin catchment, allowing the fauna interchange as *Gila pulchra* (Schönhuth *et al.,* 2014), and *Codoma ornata* (Schönhuth *et al.,* 2015) during the tecto-volcanic episodes in SMOC evolution in the extension and opening of the Gulf of California, including repeat alkaline basalt events (Henry and Aranda-Gómez, 2000; Aranda Gómez *et al*., 2005; Ferrari *et al.,* 2007); and another dispersal event to Nazas promoted by alkaline basalts events during the the Mapimi Bolson formation and Chihuahua desert evolution (Smith and Miller, 1986; Mayden, Matson and Hills 1992, Pérez-Rodríguez *et al.,* 2016).

This dispersal events were followed by a vicariant event that isolated the Lineage I associated with the volcanism in the Babicora-Bustillos sector disrupting the connection between the proto-Fuerte/Conchos and the proto-Guzman systems/Upper North Rio Grande by the periodic accumulation of sediments in the Conchos Valley (Pérez-Rodríguez *et al.,* 2016), and also promoted the formation of the lagoons in the Guzman system during the Pliocene (Hawley, 1969; Ortega-Ramírez 1990; *et al.,* 1998). This isolation event is reflected in the high genetic divergence between lineages I and II in *cytb* (D_P = 6.3-7.0%, Φ_{ST} = 0.911-0.938 $p <$ 0.05; Table 3) and the lack of shared haplotypes (Fig. 3). The formation of Lineage I and the proto-Fuerte/Conchos is consistent with that of such genetically divergent fish populations as *Cyprinella* spp. (Schönhuth *et al.,* 2010), *Campostoma ornatum* (Domínguez-Domínguez *et al.,* 2011; Schönhuth *et al.,* 2011), *Cyprinodon* spp. (Echelle *et al.,* 2005; Echelle, 2008), *Gila* spp. (Schönhuth *et al.,* 2014), *Codoma ornata* (Schönhuth *et al.,* 2015), *Moxostoma* cf. *austrinum* (Pérez-Rodríguez *et al.,* 2016) and similar to the reported divergence time of the Rio Grande *Catostomus* spp. population from the Nazas population in a previous work (Unmack *et al.,* 2014). In addition the connection between Conchos and Palomas during the Pliocene also has been reported in *Moxostoma* genus (Catostomidae) (Pérez-Rodríguez *et al.,* 2016).

The same geological events in the region of between Conchos and Nazas (Mapimi Bolson desertification and particularly alkaline basalt activity in the south of Hidalgo del Parral) was also the cause of the disconnection by a vicariant post-dispersal event that isolate the proto-Nazas/Piaxtla (Lineage III) and Mezquital basin population (Lineage IV) from the proto-Fuerte/Conchos occurred *ca.* 3.3 Mya, in the Pliocene and consistent with the observed in other ichthyofauna as *Codoma ornata* (Schönhuth *et al.,* 2015). The vicariant event was at the same time of two range expansion events of the complex from Nazas to Piaxtla and Mezquital basins also associated alkaline basalt events (Smith and Miller, 1986; Ferrari *et al.,* 2007; Pérez-Rodríguez *et al.,* 2016) allowing the fauna interchange as *Campostoma ornatum* (Domínguez-Domínguez *et al.,* 2011) and a species of the semi-aquatic snake of the genus *Thamnophis* (Conant, 1963).

It is difficult to identify specific geologic events involved in the cladogenesis of Lineages III and IV date in *ca.* 2.6 Mya, due to physiographic complexity and the occurrence of several overlapping geologic events. However, the isolation of the proto-Nazas/Piaxtla from the Mezquital populations possibly had its source in a geologic event, the alkaline volcanic activity occurring in the region during the Pliocene, especially south of Laguna Santiaguillo in the Nazas basin, in the southern region of the SMOC (Ferrari *et al.,* 2007) (Fig. 4) separating the tributary El Tunal (Mezquital Basin) from the Nazas basin, probably occurring in the Late Miocene and Pliocene (Damon and Nieto-Obregón, 1979; Luhr *et al.,* 2001; Aguirre-Díaz *et al.,* 2008).

The consequences of the isolation on the genetic variation of *Catostomus plebeius-nebuliferus* populations are: high genetic differentiation between Lineages III and IV, a high number of MS between lineages, high haplotype diversity, and no shared haplotypes in both molecular markers (Table 2). This is similar to the genetic divergence observed in co-distributed fishes such as *Codoma ornata* (Schönhuth *et al.,* 2015), *Cyprinodon* spp. (Echelle and Echelle, 1998), *Gila* spp (Schönhuth *et al.,* 2014) and *Dioda* spp. (Schönhuth *et al.,* 2012a).

Pleistocene isolations

The origin and evolution of the three basins within the Guzman system date from the Late Cretaceous and Tertiary to the Late Pleistocene, and its complexity could explain the species diversity and genetic divergence of fishes in the region. During the Early Pleistocene, the activity of the Rio Grande Rift and the glacial stage (Kansan, Middle Pleistocene) (Strain, 1966) were likely involved in the fragmentation of the Cabeza de Vaca paleolake and the formation of the Palomas pluvial paleolake, a relict of the Cabeza de Vaca that was fragmented into several endorheic rivers, lagoons, and springs. Among these, the Casas Grandes and Upper Rio Grande basins were first isolated from Del Carmen and Santa Maria (*ca.* 1.8 Mya) by the formation of the Los Muertos Bolson (Reeves, 1969; Wood and Mayden, 2002; Domínguez-Acosta and Gill 2007; Echelle, 2008) (Fig. 5). The isolation of the Upper North Rio Grande, including its flow to the Gulf of Mexico and capture of the Conchos River, has been dated to the Early Pleistocene (Gustavson, 1991; Galloway *et al.,* 2000; Rosenthal and Forstner, 2004; Leeder and Mack, 2009), but the tectonic subsidence in the Rio Grande Rift and desertification of Los Muertos Bolson has continued to during the Middle Pleistocene and the present (Mack and James*,* 1993; Berggren *et al.,* 1995; Mack, Love and Seager 1997) allowing the fragmentation of the Guzman system during as is confirmed in this study (1.1 Mya for the isolation of Palomas from Casas Grandes and 1.2 Mya for the isolation of Del Carmen from

Santa Maria) (Fig. 4). This is also consistent with previous studies that found low genetic divergence between populations of *Catostomus plebeius* of the Mimbres River (type locality of the species) and *C. plebeius* in the Palomas River (Upper North Rio Grande) (McPhee *et al.,* 2008).

The three basins within the Guzman system were isolated and reconnected several times during the Pre-Illinoian, Illinoian, and Pre-Holocene periods (Ortega-Ramirez *et al.,* 1998; Castiglia and Fawcett, 2006; Roy *et al.,* 2012). The high interbasin genetic distances and the lack of shared haplotypes for *cytb* and the low genetic differentiation and haplotypes shared among basins for *GHI* is evidence that genetic remixing among these basins. Ancestral isolation of the Del Carmen and Casas Grandes populations has also been reported in *Campostoma ornatum* (D_P = 2.9%; Φ_{ST} = 0.935) (Domínguez-Domínguez *et al.,* 2011), and considerable genetic differentiation (F_{ST} = 0.380) has been reported in *Cyprinella formosa* (Girard, 1856) (Mayden and Hillis, 1990; Broughton and Gold, 2000; Wood and Mayden, 2002; Schönhuth *et al.,* 2010). The isolation event separating the populations within *Campostoma ornatum* and *Cyprinella formosa* was also dated at the Pleistocene. Remixing gene pool and low genetic structure has been observed in other co-distributed fishes along the Guzman basin, including *Gila* spp. (Schönhuth *et al.,* 2012b; 2014) and *Cyprinodon* spp. (Echelle *et al.,* 2005; Echelle, 2008).

The vicariant event involved in the isolation of the Nazas from the Piaxtla populations and the Conchos from the Fuerte populations were estimated to have occurred 1.4 Myr and 0.6 Myr in the Pleistocene, respectively (Fig. 2 and 4), and is associated to the most recent volcanic eruptions and tectonic movements in the SMOC (Burr, 1976). This confirms results obtained for *Campostoma ornatum* (Domínguez-Domínguez *et al.,* 2011), *Gila pulchra* (Conchos River and Fuerte basin) (Schönhuth *et al.,* 2014), and *Codoma ornata* (upper Conchos River and Fuerte basin) (Schönhuth *et al.,* 2015) that show low genetic differentiation as well as shared haplotypes among basins as we observed in nuclear marker (Fig. 3).

Although the Piaxtla population has been recognized as *Catostomus plebeius* (Miller *et al.,* 2005) and suggested as an independent lineage (Ferris *et al.,* 1982), our results show that this population is closely related to *C. nebuliferus*, as has also been demonstrated by Unmack *et al.,* (2014). This distribution could be associated with ancient river capture from geographically close tributaries of the Nazas and Piaxtla basins.

Intra and Within-basin differentiation and taxonomic implications

High number of mutation steps and genetic distances where estimated for *cytb* in the *Catostomus plebeius-nebuliferus.* Surprisingly, higher between-haplotype MS values (MS = 4-7) were found within the Mezquital, Nazas, Fuerte, Casas Grandes, and Santa Maria basin populations than between the Nazas *vs.* Piaxtla (MS = 3) and Conchos *vs.* Fuerte populations (MS = 3), reflecting considerable genetic differentiation in *cytb* (Fig. 3 and Tables 3-5). The intrapopulation genetic structure in those drainages could be explained in part by life history. Migration episodes associated with environmental fluctuations on spawning sites and food availability have been extensively described in Catostomidae species (Raney and Webster, 1942; Geen, 1958; Hohler, 1981; Tyus and Karp, 1990; Rinne, 1995; Bunt and Cooke, 2001; Schmetterling and McFee, 2006; Cole *et al.,* 2008). Accordingly, the historic climate fluctuations in the area (Matcalfe *et al.,* 2000; 2006; Mastretta-Yanes *et al.,* 2015) could promote similar migration episodes in *C. plebeius-nebuliferus* populations, which could be responsible for the isolation of migrants in separate regions within the basins. However, the nuclear *GHI* is in disagreement because low number of mutations steps but significant structure was detected. The discordance between molecular markers coukd be result of the incomplete lineage sorting and an unclear species boundary limits resources for nuclear marker.

First, the *Catostomus plebeius-nebuliferus* complex was revealed to comprise a set of distinct genetic groups, suggesting that it is configured by at least four independent evolutionary lineages. Cladogenesis was found to be associated with the high tecto-volcanic activity and climate fluctuations from the Pliocene to the present in Northwest Mexico. This is supported by the phylogenetic relationships and the genetic differences between lineages for *cytb*, which are higher than those reported among sibling vertebrate groups for *cytb* (~2%) (Brown, George and Wilson, 1979; Avise, 1998; Bradley and Baker, 2001) and between sister species within the family (Berendzen, Simons and Wood, 2003). They are similar to those observed among species of other fish genera partially or completely distributed in basins of northwest Mexico and the southwestern USA, such as *Cyprinodon* spp (Echelle *et al.,* 2005; Echelle, 2008), *Gila s*pp (Schönhuth *et al.,* 2012b; 2014), *Cyprinella* spp (Broughton and Gold, 2000; Wood and Mayden, 2002; Schönhuth *et al.,* 2010), *Campostoma* spp (Domínguez-Domínguez *et al., 2011*; Schönhuth *et al.,* 2011), and *Moxotoma* spp (Pérez-Rodíguez *et al.,* 2016). However, the genetic distances estimated between species of Catostomidae are 3.33% (Bart *et al.,* 2010) taking into account different genera of the family, suggesting that an exaustive review of if is possible the delimitation species of catostomids and the Catostomus genus with *GHI.* Thus, we recommend a taxonomic revision of the *Catostomus plebeius-nebuliferus* due to the high divergences and genetic structure reported in this work and not clear species limits between populations of the complex.

CONSERVATION CONSIDERATIONS

This research provides important information applicable to the conservation of *Catostomus nebuliferus* and *C. plebeius*. Defining conservation units is of particular interest, especially in species considered threatened, as is the case of the investigated fish, which are listed as in danger of extinction in the Norm-059-SEMARNAT-2010 (SEMARNAT, 2010) and are distributed in semiarid regions with increasing issues of water use (Alcocer and Bernal-Brooks, 2010).

Currently, the evolutionary significance unit (ESU) is recognized as a population or group of populations warranting separate management or priority for conservation according to the following factors: reproductive isolation and adaptation, reciprocal monophyly, and exchangeability of populations (Allendorf and Luikart, 2007).

Based on the high genetic divergence, number of mutation steps, phylogenetic relationships, shared haplotypes, reproductive isolation, and the subsequent null exchangeability of populations, we suggest 9 ESUs corresponding to each basin population of the *Catostomus plebeus-nebuliferus* complex: Lineage I = (1) Upper North Grande River, (2) Casas Grandes, (3) Santa Maria and (4) Del Carmen basins; Lineage II = (5) Fuerte and (6) Upper South Grande River basin (Conchos River); Lineage III = populations distributed in (7) Piaxtla and (8) Nazas basins; and Lineage IV = the population in (9) Mezquital basin. The definition of these nine ESUs should aid in conservation of the species complex and of the evolutionary lineages of which they are configured.

Acknowledgements

We thank Rogelio Rosas Valdez and Hugo Mejia Madrid for field support in obtaining fish samples. Thanks to Javier Pérez Tris and Silvia Perea for valuable comments, and to the reviewers from The Lucidus Consultancy, Alan Pike and Kathleen Hills, for English corrections and suggestions. Thanks to the partial funding provided by CIC-UMSNH 2014 and the Ministerio de Economía y Competitividad of Spain (Project Reference: CGL2013-41375-P), and CONACYT for sabbatical grant to ODD (Number: 260596).

References

Aguirre-Díaz GJ, Labarthe-Hernández G, Tristán-González M, Nieto-Obregón J, Gutiérrez-Palomares I. 2008. The Ignimbrite Flare-Up and Graben Calderas of the Sierra Madre Occidental, Mexico. In: Gottsmann J, Martí J. Eds. *Caldera Volcanism Analysis, Modelling and Response*. Developments in Volcanology, Elsevier, Oxford, UK. 10: 143–180.

Alcocer J, Bernal-Brooks FW. 2010. Limnology in Mexico. *Hydrobiologia.* 644: 1–54.

Allendorf FW, Luikart G. 2007. *Conservation and the genetics of populations*. Blackwell Publishing Ltd., Oxford, UK.

Aranda-Gómez JJ, Luhr JF, Housh TB, Valdez-Moreno G, Chávez-Cabello G. 2005. El vulcanismo tipo intraplaca del Cenozoico tardío en el centro y norte de México: una revisión. *Boletín de la Sociedad Geológica Mexicana, Volumen Conmemorativo del Centenario Temas Selectos de la Geología Mexicana.* 3: 187-225.

Avise JC. 1998. The history and purview of phylogeography: A personal reflection. *Molecular Ecology.* 7: 371–379.

Baird SF, Girard C. 1854. Descriptions of new species of fish collected in Texas, New Mexico and Sonora, by Mr. John H. Clark on the U.S. and Mexican boundary survey, and in Texas by Capt. Stewart Van Vliet, USA*. Proceedings of Academy of Natural Sciences.* 8: 24–29.

Bandelt HJ, Forster P, Röhl A. 2000. Median-Joining networks for inferring intraspecific phylogenies *Molecular biology and evolution.* 16: 37–48.

Bart HL, Reneau PC, Doosey MH, Bell CD. 2010. Evolutionary divergence of duplicate copies of the growth hormone gene in suckers (Actinopterygii: Catostomidae). *International journal of molecular sciences.* 11: 1090–1102.

Beaulieu JM, Tank DC, Donoghue MJ. 2013. A Southern Hemisphere origin for campanulid angiosperms, with traces of the break-up of Gondwana. *BMC Evolutionary Biology.* 13: 80.

Berendzen PB, Simons AM, Wood RM. 2003. Phylogeography of the northern hogsucker, *Hypentelium nigricans* (Teleostei: Cypriniformes): Genetic evidence for the existence of the ancient Teays River. *Journal of Biogeography.* 30: 1139–1152.

Berggren WA, Hilgen FJ, Langereis CG, Kent DV, Obradovich JD, Raffi I, Raymo ME, Shackleton NJ. 1995. Late Neogene chronology: New pespectives in high-resolution stratigraphy. *Geological Society of America Bulletin.* 107, 1272-1287.

Bradley RD, Baker RJ. 2001. A test of the genetic species concept: cytochrome-b sequences and mammals. *Journal of Mammalogy.* 82: 960–973.

Broughton RE, Gold JR. 2000. Phylogenetic relationships in the North American cyprinid genus *Cyprinella* (Actinopterygii: Cyprinidae) based on mitochondrial ND2 and ND4L gene sequences. *Copeia.* 2000: 1–10.

Brown WM, George MJr, Wilson AC. 1979. Rapid evolution of animal mitochondrial DNA. *Proceedings of the National Academy of Sciences.* 76: 1967–1971.

Bunt CM, Cooke SJ. 2001. Post-spawn movements and habitat use by greater redhorse, *Moxostoma valenciennesi*. *Ecology of Freshwater Fish.* 10: 57–60.

Burr BM. 1976. A review of the Mexican Stoneroller, *Campostoma ornatum* Girard (Pisces: Cyprinidae). *Transactions of the San Diego Society of Natural History.* 18: 127–143.

Buth DG, Crabtree CB. 1985. A comparison of the genetic characteristics of Zuni sucker (*Catostomus discobolus yarrow*), the bluehead sucker (*Catostomus discobolus*), and the Rio Grande sucker (*Catostomus plebeius*). *Report of the New Mexico Department of Game and fish*, Santa Fe, New Mexico, USA, 34pp.

Castiglia PJ, Fawcett PJ. 2006. Large Holocene lakes and climatic change in the Chihuahua Desert. *Geology.* 34: 113–116.

Ceballos G, Arroyo-Cabrales J, Ponce E. 2010. Effects of Pleistocene environmental changes on the distribution and community structure of mammalian fauna of Mexico. *Quaternary Research.* 73: 464–473.

Clements MD, Bart HLJr, Hurley DL. 2012. A different perspective on the phylogenetic relationships of the Moxostomatini (Cyprinoformes: Catostomidae) based on cytochrome-b and Growth Hormone intron sequences. *Molecular Phylogenetics and Evolution.* 63: 159–167.

Cole DD, Mock KE, Cardall BL, Crowl TA. 2008. Morphological and genetic structuring in the Utah Lake sucker complex. *Molecular Ecology.* 17: 5189–5204.

Crabtree CB, Buth DG. 1987. Biochemical systematics of the catostomid genus *Catostomus*: assessment of *C. clarki*, *C. plebeius*, and *C. discobolus* including the Zuni sucker, *C. d. yarrow*. *Copeia.* 1987: 843–854.

Damon PE, Nieto-Obregón J. 1979. Un plegamiento Neogénico en Nayarit y Jalisco y evolución geomorfológica del Rio Grande de Santiago. *Asociación Ingenieros de Minas, Metalurgistas y Geólogos de México, Memoria, XII Convención Nacional*. 156–191.

Domínguez-Acosta M, Gill TE. 2007. PIXE based geochemical characterization of the pluvial lake Palomas system, Chihuahua, Mexico. *Procceedings of the 11th International Congerence on PIXE and its Analytical Applications*, Puebla, Mexico.

Domínguez-Domínguez O, Alda F, Pérez-Ponce de León G, García-Garitagoitia JL, Doadrio I. 2008. Evolutionary history of the endangered fish Zoogoneticus quitzoensis (Bean, 1898) (Cyprinodontiformes: Goodeidae) using a sequential approach to phylogeography based on mitochondrial and nuclear DNA data. *BMC Evolutionary Biology.* 8: 161 DOI: 10.1186/1471-2148-8-161.

Domínguez-Domínguez O, Pérez-Ponce de León G. 2009. ¿La mesa central de México es una provincia biogeográfica? Análisis descriptivo basado en componentes bióticos dulceacuícolas. *Revista Mexicana de Biodiversidad.* 80: 835–852.

Domínguez-Domínguez O, Vila M, Pérez-Rodríguez R, Remón N, Doadrio I. 2011. Complex evolutionary history of the Mexican stoneroller *Campostoma ornatum* Girard, 1856 (Actinopterygii: Cyprinidae). *BMC Evolutionary Biology.* 11: 153.

Doosey MH, Bart HLJr, Saitoh K, Miya M. 2010. Phylogenetic relationships of Catostomid fishes (Actinopterygii: Cypriniformes) based on mitochondrial ND4/ND5 gene sequences. *Molecular Phylogenetics and Evolution.* 54: 1028–1034.

Drummond AJ, Ho SYW, Rawlence N, Rambaut A. 2007. *A rough guide to BEAST* 1.4. University of Edinburgh, United Kingdom.

Dupanloup I, Schneider S, Excoffier L. 2002. A simulated annealing approach to define the genetic structure of populations. *Molecular Ecology.* 11: 2571–2581.

Echelle AA, Echelle AF. 1998. Evolutionary relashionships of Pupfishes in the *Cyprinodon eximius* complex (Atherinomorpha: Cyprinodontiformes). *Copeia.* 1998: 852–865.

Echelle AA, Carson EW, Echelle AF, Van Den Bussche RA, Dowling TE, Meyer A. 2005. Historical biogeography of the New-World Pupfish genus *Cyprinodon* (Teleostei: Cyprinodontidae). *Copeia.* 2005: 320–339.

Echelle AA. 2008. The western Pupfish clade (Cyprinodontidae: *Cyprinodon*): mtDNA divergence times and drainage history. *Geological Society of America Special Papers.* 439: 27–38.

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

Ferrari L, Valencia-Moreno M, Bryan S. 2007. Magmatism and tectonics of the Sierra Madre Occidental and its relation with the evolution of the western margin of North America. *Geological Society of America Special Papers.* 422: 1–39.

Ferris SD, Buth DG, Whitt GS. 1982. Substantial genetic differentiation among populations of *Catostomus plebeius*. *Copeia*. 1982: 444–449.

Galloway WE, Ganey-Curry PE, Li X, Buffler RT. 2000. Cenozoic depositional history of the Gulf of Mexico Basin. *American Association of Preloleum Geologists Bulletin.* 84: 1743-1774.

Garman S. 1881. New and little-known reptiles and fishes in the museum collections. *Bulletin of the Museum of Comparative Zoology.* 8: 85–93.

Geen GH. 1958. Reproduction of three species of suckers (Catostomidae) in British Columbia. Unpublished D. Phil. Thesis. University of British Columbia.

Girard C F. 1856 Researches upon the cyprinoid fishes inhabiting the fresh waters of the United States, west of the Mississippi Valley, from specimens in the museum of the Smithsonian Institution. *Proceedings of the Academy of Natural Sciences of Philadelphia.* 8: 165–213.

Gustavson TC. 1991. Arin basin depositional system and paleosol: Fort Hancock and Camp Rice Formations (Pliocene-Pleistocene), Hueco Bolson, west Texas and adjacent Mexico. *Bureau of Economic Geology, Report Investigation, University of Texas.* 198.

Hawley JW. 1969. Notes on the geomorphology and late Cenozoic geology of northwestern Chihuahua. In: Cordoba DA, Wengerd SA, Shomaker JW EDS. *The border region (Chihuahua, Mexico and USA)*. New Mexico Geological Society 20th Annual Fall Field Conference Guidebook. 228 p.

Henry CD, Aranda-Gómez JJ. 2000. Plate interactions control middle-late Miocene proto-Gulf and Basin and Tange extension in the southern Basin and Range. *Tectonophysics.* 318: 1–16.

Hohler DB. 1981. A dwarfed population of *Catostomus rimiculus* (Catostomidae: Pisces) in Jenny Creek, Jackson County, Oregon. Unpublished D. Phil. Thesis Oregon State University.

Laitnen NJ. 2012. The phylogeogrtaphy of the Mountain Sucker [*Catostomus* (*Pantosteus*) *platyrhynchus*]. Unpublished D. Phil. Thesis. Brigham Young University.

Lanfear R, Calcott B, Ho SY, Guindon S. 2012. Partitionfinder: Combined selection of partitioning schemes and substitutions models for phylogenetic analyses. *Molecular Biology and Evolution.* 29: 1695-1701.

Leeder MR, Mack GH. 2009. Basin-fill incision, Rio Grande and Gulf of Corinth rifts: convergent response to climatic and tectonic drivers. In: Nichols G, Williams E, Paola C, Wiler J, Sons Eds. *Sedimentary processes, environments and basins: A tribute to Peter Friend (Special Publication 38 of the IAS),* 9-27.

Le Sueur, CA. 1817. A new genus of fishes, of the order Abdominales, proposed, under the name of *Catostomus*; and the characters of this genus, with those of its species, indicated. *Journal of the Academy of Natural Sciences*. 1: 88-96, 102-111.

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

Luhr JF, Henry CD, Housh TB, Aranda-Gómez JJ, McIntosh WC. 2001. Early extension and associated mafic alkali volcanism from the southern Basin and Range Province: geology and petrology of the Rodeo and Nazas volvanic fields, Durango (Mexico). *Geological Society of America Bulletin.* 113: 760–773.

Machordom A, Doadrio I. 2001. Evidence of a Cenozoic Betic-Kabilian connection based on freshwater fish phylogeography (Luciobarbus, Cyprinidae). *Molecular Phylogenetics and Evolution.* 18: 252–263.

Mack GH, James WC. 1993. Control of basin symmetry on fluvial lithofacies. Camp Rice and Palomas Formations (Plio-Pleistocene) in the southern Rio Grande Rift, USA. In: Marzo M, Pulgdefabregas C. Eds. *Alluvial sedimentation.* Special Publication of the International Association of Sedimentologists. 17, 439-449.

Mack GH, Love DW, Seager WR. 1997. Spillover model for axial rivers in regions of continental extension: the Rio Mimbres and Rio Grande in the southern Rio Grande rift, USA. *Sedimentology.* 44, 637-652.

Mastretta-Yanes A, Moreno-Letelier A, Piñero D, Jorgensen TH, Emerson BC. 2015. Biodiversity in the Mexican highlands and the interaction of geology, geography and climate within the Trans-Mexican Volcanic Belt. *Journal of Biogeography.* 42: 1586-1600.

Mayden RL, Hillis DM. 1990. Natural history and systematics of the largemouth shiner, *Cyprinella bocagrande* (Teleostei: Cypriniformes), with comments on conservation status. *Copeia.* 1990: 004–1011.

Mayden RL, Matson RH, Hillis DM. 1992. Speciation in the North American genus *Dionda* (Teleostei: Cypriniformes). In: Mayden RL Eds. *Systematics, historical ecology and North American freshwater fishes.* Stanford University Press, Stanford, California, 710-746.

McPhee MV, Osborne MJ, Turner TF. 2008. Genetic diversity, population structure, and demographic history of the Rio Grande Sucker, *Catostomus* (*Pantosteus*) plebeius, in New Mexico. *Copeia.* 2008: 191–199.

Miller RR, Minckley WL, Norris SM. 2005. *Freshwater Fishes of México*. University Of Chicago Press, Chicago, USA.

Near TJ, Benard MF. 2004. Rapid allopatric speciation in logperch darters (Percidae: Percina). *Evolution*. .58: 2798–2808.

Ortega-Ramírez J. 1990. *Le sommet du remplissage Quaternaire de la Laguna Babícora (État de Chihuahua, Nord-ouest du Mexique); reconstitution des paléoenvironnements à partir de la sédimentologie et de la stratigraphie*. Thesis Dissestation, Louis Pasteur University, Strasbourg, France.

Ortega-Ramírez JR, Valiente-Banuet A, Urrutia-Fugauchi J, Mortera-Gutiérrez CA, Alvarado-Valdéz G. 1998. Paleoclimatic changes during the Late Pleistocene-Holocene in Laguna Babícora, near the Chihuahua Desert, Mexico*. Canadian Journal of Earth Sciences.* 35: 1168–1179.

Page LM, Burr BM. 2011. Peterson field guide to freshwater fishes of North America north of Mexico. *Houghton Mifflin Harcourt*, Boston, Massachusetts, USA.

Palumbi SR, Martin A, Romano S, McMillan MO, Stice L, Grabowski G. 1991. *Simple fool's guide to PCR.* University of Hawaii, Honolulu.

Perdices AA, Carmona JA, Fernández-Delgado C, Doadrio I. 2001. Nuclear and mitochondrial data reveal high genetic divergence among Atlantic and Mediterranean populations of the Iberian killifish *Aphanius iberus* (Teleostei: Cyprinodontidae). *Heredity.* 87: 314–324.

Perdices A, Bermingham E, Montilla A, Doadrio I. 2002. Evolutionary history of the genus *Rhamdia* (Teleostei: Pimelodidae) in Central America. *Molecular Phylogenetics and Evolution.* 25: 172–189.

Pérez-Rodríguez R, Domínguez-Domínguez O, Mar-Silva AF, Doadrio I, Pérez-Ponce de León, G. 2016. The historical biogeography of the southern group of the suker genus *Moxostoma* (Teleostei: Catostomidae) and the colonization of central Mexico. *Zoological Journal de Linnean Society.* DOI: 10.1111/zoj.12383

Rambaut A, Suchard MA, Xie D, Drummond AJ. 2014. Tracer v1.6, Available at http://beast.bio.ed.ac.uk/Tracer

Raney EC, Webster DA. 1942. The spring migration of the common White Sucker, *Catostomus c. commersonnii* (Lacépède), in Skaneateles Lake Inlet, New York. *Copeia.* 1942: 139–148.

Reeves CC. 1969. Pluvial Lake Palomas, northwestern Chihuahua, Mexico. *New Mexico Geological Society Guidebook.* 20: 143–155.

Rice WR. 1989. Analysing tables of statistical tests. *Evolution.* 43: 223–225.

Riddle BR, Hafner DJ. 2006. A step-wise approach to integrating phyloeographic and phylogenetic niogeographic perspectives on the history of a core North American warm deserts biota. *Journal of Arid Environments.* 66: 435–461.

Rinne JN. 1995. Reproductive biology of the Rio Grande sucker, *Catostomus plebeius* (Cypriniformes), in a montane stream, New Mexico. *Southwestern Naturalist.* 40: 237–241.

Rinne JN, Hughes RM, Calamusso B. 2005. Historical changes in large river fish assemblages of the Americas. *American Fisheries Society Symposium 45, Bethesda, Maryland*.

Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics.* 19: 1572–1574.

Rosenthal J, Forstner MRJ. 2004. Effects of Plio-Pleistocene barrier on Chihuahua Desert herpetofauna. In: Hoyt CA, Karges J Eds. *Proceedings of the sixth symposium on the natural resources of the Chihuahua Desert region*. 269-282.

Roy PD, Jonathan MP, Pérez-Cruz LL, Sánchez-Córdova MM, Quiroz-Jiménez JD, Romero FM. 2012. A millennial-scale Late Pleistocene-Holocene palaeoclimatic record from the western Chihuahua Desert, Mexico. *Boreas.* 41: 707–717.

Schmetterling DA, McFee JA. 2006. Migrations by fluvial largescale suckers (Catostomus macrocheilus) after transport upstream of Milltown Dam, Montana. *Northwest Science.* 80: 18–25.

Schönhuth S, Mayden RL. 2010. Phylogenetic relationships in the genus *Cyprinella* (Actinopterygii: Cyprinidae) based on mitochondrial and nuclear gene sequences. *Molecular Phylogenetics and Evolution.* 55: 77–98.

Schönhuth S, Blum MJ, Lozano-Vilano L, Neely DA, Varela-Romero A, Espinosa H, Perdices A, Mayden RL. 2011. Inter-basin exchange and repeated headwater capture across the Sierra Madre Occidental inferred from the phylogeography of Mexican stonerollers. *Journal of Biogeography.* 38: 1406–1421.

Schönhuth S, Hills DM, Neely DA, Lozano-Vilano L, Perdices A, Mayden RL. 2012a. Phylogeny, diversity, and species delimitation of the North American Round-Nosed Minnows (Teleostei: *Dionda*), as inferred from mitochondrial and nuclear DNA sequences. *Molecular Phylogenetics and Evolution.* 62: 427–446.

Schönhuth S, Shiozawa DK, Dowling TE, Mayden RL. 2012b. Molecular systematics of western North American cyprinids (Cyprinodontiformes: Cyprinidae). *Zootaxa.* 3586: 281–303.

Schönhuth S, Perdices A, Lozano-Vilano L, García-de-León FJ, Espinosa H, Mayden RL. 2014. Phylonetic relationship of North American western chubs of the genus *Gila* (Cyprinidae, Teleostei), with emphasis on southern species. *Molecular Phylogenetics and Evolution.* 70: 210–230.

Schönhuth S, Lozano-Vilano L, Perdices A, Espinosa H, Mayden RL. 2015. Phylogeny, genetic diversity and phylogeogrpahy of the genus *Codoma* (Teleostei, Cyprinidae). *Zoologica Scripta.* 44: 11–28.

SEMARNAT. 2010. *Norma Oficial Mexicana 059: Protección Ambiental-Especies Nativas de México de Flora y Fauna Silvestres-Categorías de Riesgo y Especificaciones para su Inclusión, Exclusión o Cambio-Lista de Especies en Riesgo*. Diario Oficial de la Federación, México.

Silvestro D, Michalak I. 2012. raxmlGUI: a graphical front-end for RAxML. *Organisms Diversity and Evolution.* 12: 335–337.

Smith GR. 1966. Distribution and evolution of the North American catostomid fishes of the subgenus *Pantosteus*, genus *Catostomus*. *Miscelaneous Publications, Museum of Zoology, University of Michigan.* 129: 1–132.

Smith GR. 1992. Phylogeny and biogeography of the Catostomidae, freshwater fishes of North America and Asia, In: Mayden RL. Ed. *Systematics, historical ecology and North American freshwater fishes*. Stanford University Press, Stanford.

Smith GR, Stewart JD, Carpenter NE. 2013. Fossil and recent mountain suckers *Pantosteus,* and significance of introgression in catostomin fishes of Wester United States. *Occasional Papers of The Museum of Zoology University of Michigan.* 743: 1-59.

Smith ML, Miller RR. 1986. The evolution of the Rio Grande basin as inferred from its fish fauna. In: Hocutt CH, Wiley EO. Eds. *The zoogeography of North American freshwater fishes*. New York, John Wiley and Sons, 457–485.

Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics.* 22: 2688–2690.

Strain WS. 1966. Blancan mammalian fauna and Pleistocene formations. Hudspeth County, Texas. *Texas Memorial Museum, Bulletin 10*. Austin, Texas.

Sublette JE, Hatch MD, Subblete M. 1990. The fishes of New Mexico. *University of New Mexico Press*. Albusquerque, New Mexico, USA.

Sun YH, Xie CX, Wang WM, Liu SY, Treer T, Change MM. 2007. The genetic variation and biogeography of the catostomid fishes based on mitochondrial and nucleic DNA sequences. *Journal of Fish Biology.* 70: 291–309.

Tamura K, Nei M. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution.* 10: 512-526**.**

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution.* 28: 2731–2739.

Tavaré S. 1986. Some probabilistic and statistical problems in the analysis of DNA sequences. *Lectures on mathematics in the life sciences.* 17: 57–86.

Tyus HM, Karp CA. 1990. Spawning and movements of Razorback Sucker, *Xyrauchen texanus*, in the Green River basin of Colorado and Utah. *Southwestern Naturalist.* 35: 427–433.

Unmack PJ, Dowling TE, Laitinen NJ, Secor CL, Mayden RL, Sciozawa DK, Smith GR. 2014. Influence of introgression and geological processes on phylogenetic relationships of western North American Mountain Suckers (*Pantostues*, Catostomidae). *PloS ONE.* 9, e90061.

Wei-Jen C, Mayden RL. 2012. Phylogeny of suckers (Teleostei: Cypriniformes: Catostomidae): further evidence of relationships provided by the single-copy nuclear gene IRBP2. *Zootaxa.* 3586: 195–210.

Wood RM, Mayden RL. 2002. Speciation and anagenesis in the genus Cyprinella of Mexico (Teleostei: Cyprinidae): a case study of Model III allopatric speciation. *Reviews in Fish Biology and Fisheries.* 12: 253–271.

Yu Y, Harris AJ, He XJ. 2010. S-DIVA (Statistical Disperal-Vicariance Analysis): A tool for inferring biogeographic histories. *Molecular Phylogenetics and Evolution.* 56: 848-850.

Yu Y, Harrid AJ, Blair C, He XJ. 2015. RASP (Reconstruct Ancestral State in Phylogenies): A tool for historical biogeography. *Molecular Phylogenetics and Evolution.* 87: 46-49.

Zardoya R, Doadrio I. 1999. Molecular evidence on the evolutionary and biogeographical patterns of European Cyprinids. *Journal of Molecular Evolution.* 49: 227–237.

Zhang J, Kapli P, Pavlidis P, Stamatakis A. 2013. A general species delimitation method with applications to phylogenetic placements. *Bioinformatics.* 29: 2869–2876.

Figures

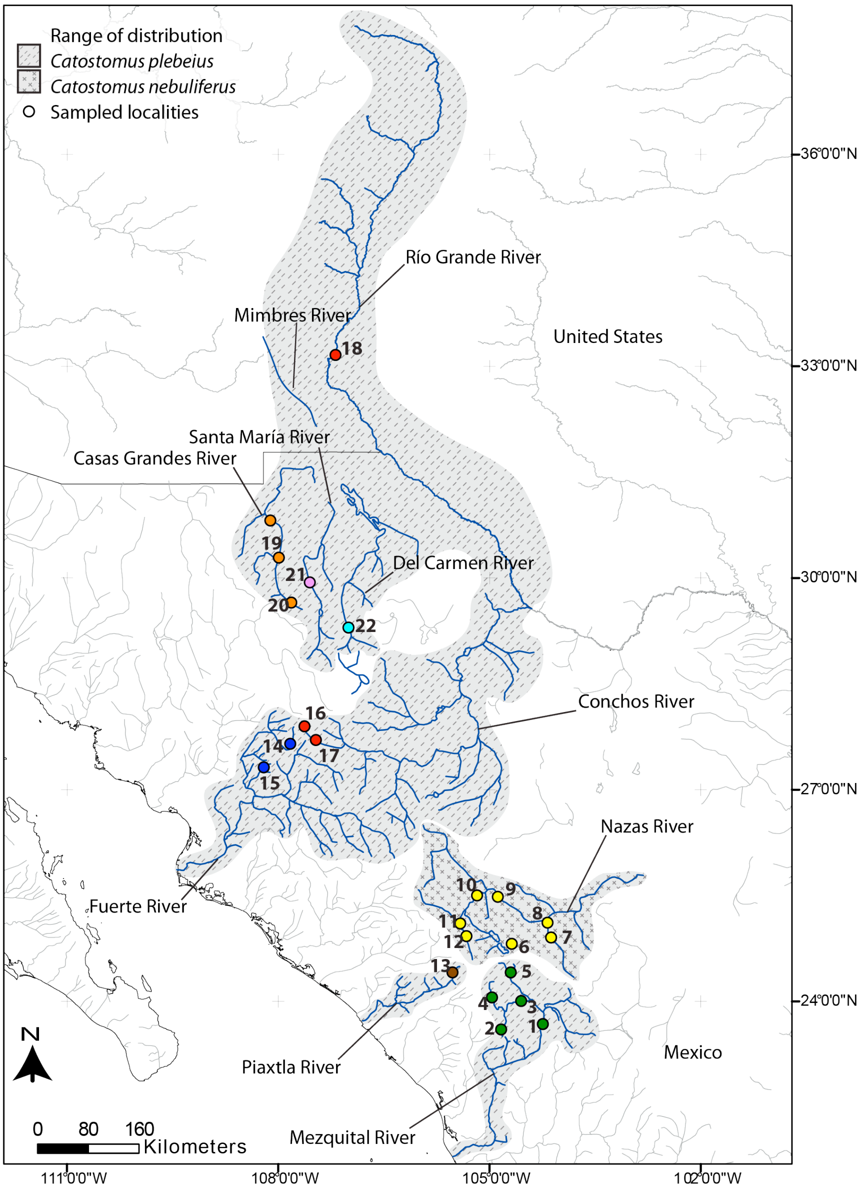


Figure 1. Distribution of *Catostomus plebeius-nebuliferus* complex and sampling localities. the colors correspond to the eight main basins and numbers to the localities: (1) Santa Gertrudis, (2) Arroyo Las Bayas, (3) Pino Suarez, (4) La Barranca, (5) Puente Mimbres, (6) El Cuarto, (7) Peñón Blanco, (8) Covadonga, (9) Arroyo Torreones, (10) El Peñasco, (11) El Olote, (12) Atotonilco, (13) Las Vegas, (14) Fuerte, (15) Oteros, (16) Bocoyna, (17) Conchos, (18) South Fork Palomas Creek, (19) Casas Grandes, (20) Ignacio Zaragoza, (21) Buenaventura and (19) Santa Clara.

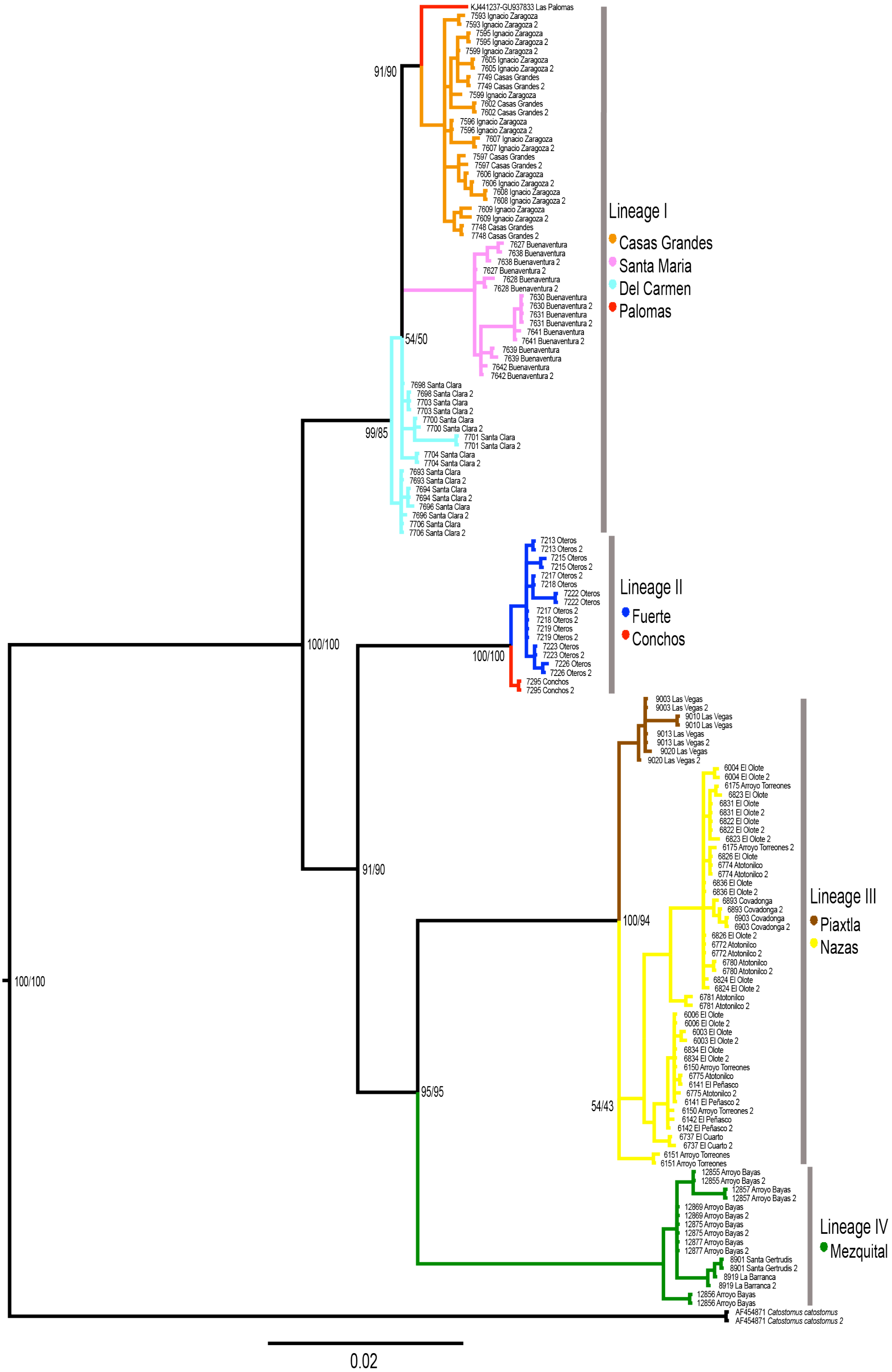


Figure 2. Phylogenetic reconstruction based on the concatenated data of *cytb* gene and *GHI* region (Bayesian inference/maximum likelihood).

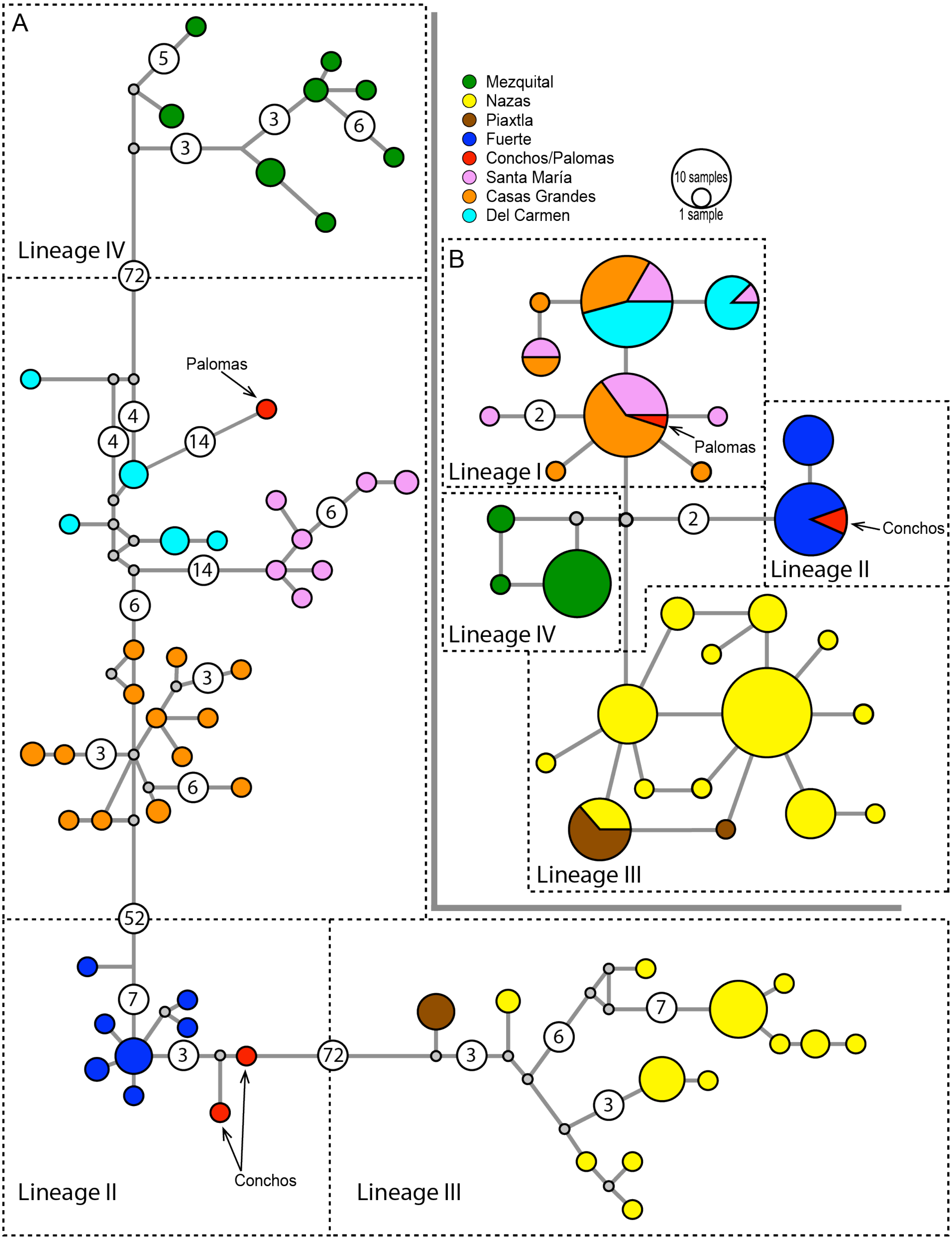


Figure 3. Haplotype network for *Catostomus plebeius-nebuliferus* complex based on *cytb* (A) and *GHI* (B).

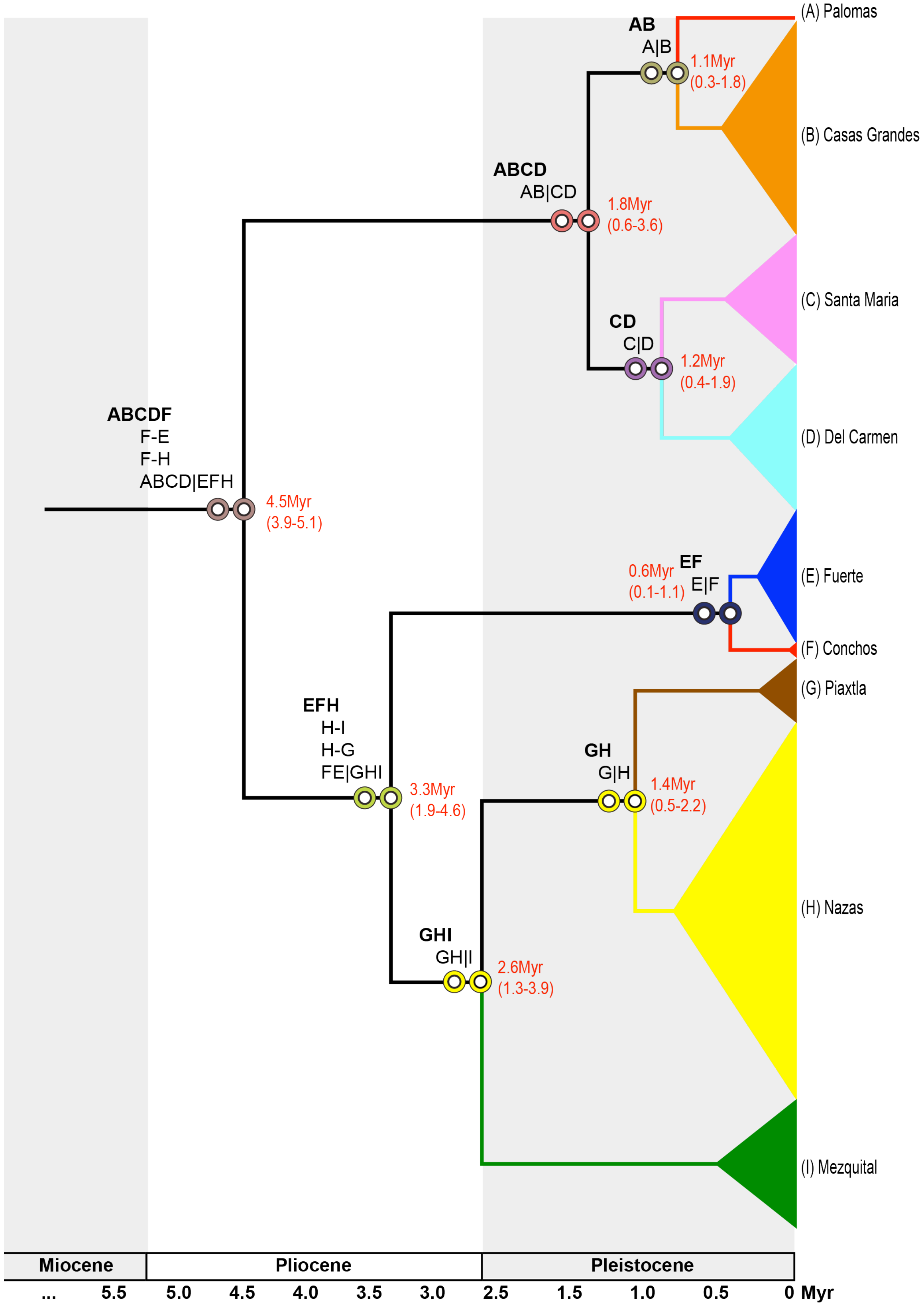


Figure 4. Divergence times estimation, S-DIVA and S-DEC analyses of the *Catostomus plebeius-nebuliferus* complex. Red labels represent the divergence times millions years ago. The bold names represent the ancestral area estimated. Dispersal events and vicariant events are represented with - and | respectively.

	Biogeographic event	Date	Probable event
(1)	Range expansion from Conchos to Fuerte and Nazas	4.5 Mya (3.9-5.1) Pliocene	Basin catchment by alkaline basalt activity and Mapimi Bolson evolution
(2)	Isolation of Proto Guzmán system/Palomas		Formation of Babícora-Bustillos subsection
(3)	Range expansion from Nazas to Mezquital and Piaxtla	3.3 Mya (1.9-4.6) Pliocene-Early Pleistocene	Alkaline basalt activity during evolution of SMOC
(4)	Isolation of Proto Nazas/Piaxtla - Mezquital		Alkaline basalt activity in the south of Hidalgo del Parral
(5)	Isolation of Mezquital	2.6 Mya (1.3-3.9) Pliocene-Early Pleistocene	Alkaline volcanic activity in the south of Laguna Santiaguillo
	Within lineage isolation		
(6)	Río Grande/Casas Grande	1.8 Mya (0.6-3.6) Early Pleistocene	Fragmentation of Cabeza de Vaca paleolake, formation of Palomas paleolake system and Los Muertos Bolson, and connection of Upper Río Grande with Conchos River
(7)	Nazas - Piaxtla isolation	1.4Mya (0.5-2.2) Middle Pleistocene	Fault formation and alkaline basalts during SMOC evolution
(8,9)	Fragmentation of Guzman system	1.1-1.2 Mya Middle Pleistocene	Tectonic subsidence in the Río Grande Rift and desiccation processes in Los Muertos Bolson
(10)	Fuerte - Conchos isolation	0.6 Mya (0.1-1.1) Middle Pleistocene	Fault formation and alkaline basalts during SMOC evolution

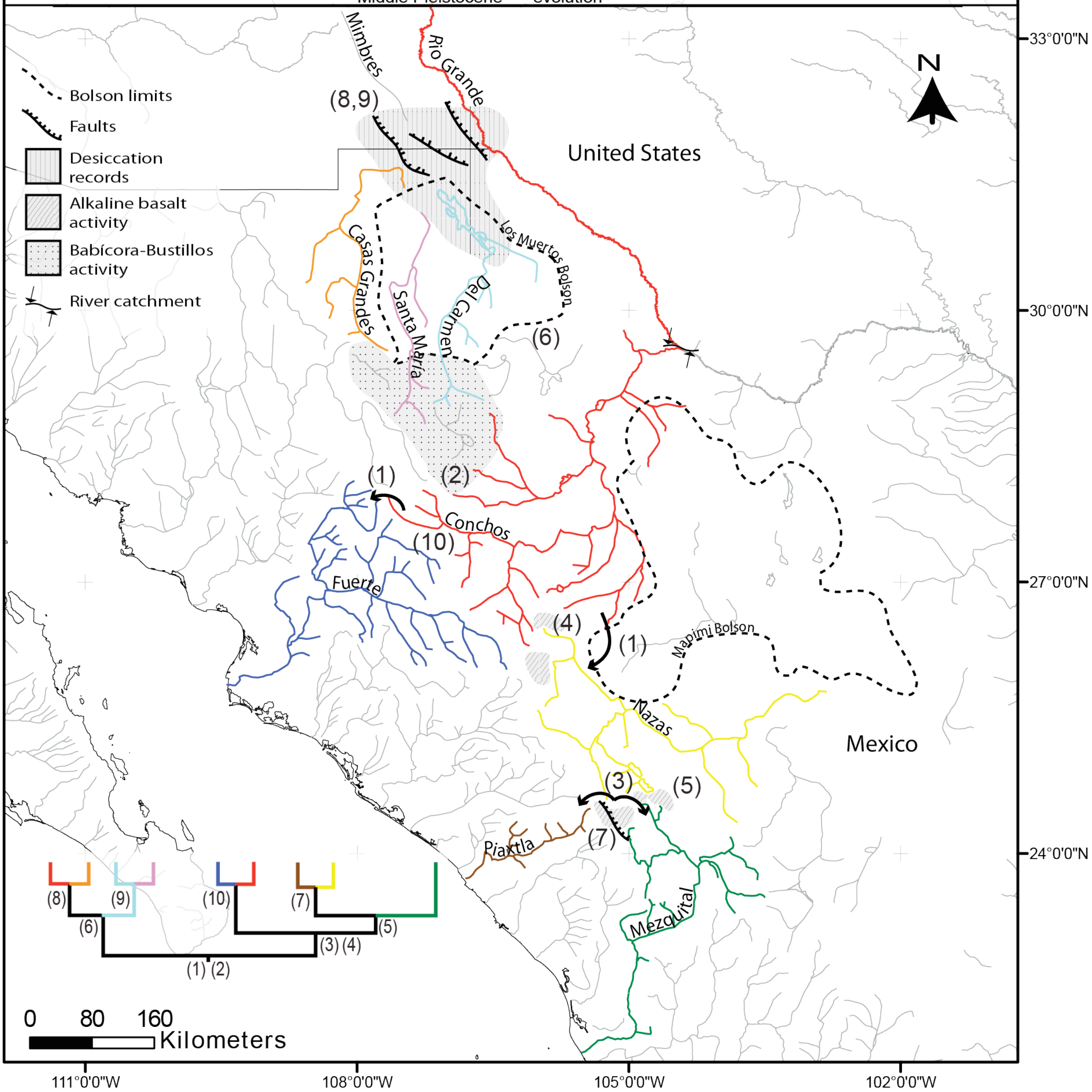


Figure 5. Biogeographical scenario for the *Catostomus plebeius-nebuliferus* complex in the Northwestern Mexico.

Tables

Table 1. Sampled populations of the *Catostomus plebeius-nebuliferus* complex.

Species	**Basin**	***cytb* individuals**	***GHI* individuals**	**GenBank Accession**
Catostomus nebuliferus	Nazas	29	29	
Catostomus plebeius	Mezquital	12	8	
	Piaxtla	4	4	
	Fuerte	10	9	*cytb*: KJ441240 (Unmack *et al.,* 2014)
	Uppers North R. Grande (Palomas)	1	1	*cytb*: KJ441237 (Unmack *et al.,* 2014) *GHI*: GU937833 (Clements *et al.,* 2012)
	Upper South R. Grande (Conchos)	2	1	*cytb*: KJ441239 (Unmack *et al.,* 2014)
	Del Carmen	10	9	*cytb*: KJ441238 (Unmack *et al.,* 2014)
	Santa Maria	8	8	
	Casas Grandes	15	13	*cytb*: KJ441235 (Unmack *et al.,* 2014)

Table 2. Genetic diversity for *cytb|GHI* markers in *Catostomus plebeius-nebuliferus* populations.

Basins	**Hn**	**π**	**h**	**Θ_S**
Nazas	12|18	0.010|0.003	0.850|0.829	0.009|0.003
Mezquital	8|3	0.008|0.001	0.924|0.342	0.010|0.001
Piaxtla	1|3	--|0.008	--|0.464	--|0.001
Fuerte	7|3	0.003|0.001	0.873|0.542	0.006|0.0005
Upper South R. Grande (Conchos)	2|1	0.003|--	--	--
Upper North R. Grande (Palomas)	1|1	--	--	--
Del Carmen	5|2	0.005|0.001	0.800|0.503	0.006|0.0005
Santa Maria	7|7	0.005|0.006	0.964|0.792	0.005|0.003
Casas Grandes	13|6	0.006|0.002	0.981|0.683	0.008|0.002

n = sample size, Hn = number of haplotype, π = nucleotide diversity, h = haplotype diverisity, Θ_S = segregative site proportion per site.

Table 3. Absolute pairwise un-corrected *p*-distances D_P for *cytb* (under the diagonal) and *GHI* (above the diagonal using pairwise-deletion) of populations of *Catostomus plebeius-nebuliferus* complex.

	Mezquital	**Fuerte**	**USR Grande (Conchos)**	**Casas Grandes**	**Santa Maria**	**Del Carmen**	**Piaxtla**	**Nazas**	**UNR Grande (Palomas)**
Mezquital	--	0.9	0.8	0.8	0.8	0.9	0.7	0.8	0.6
Fuerte	8.7	--	0.1	0.7	0.7	0.8	0.9	0.7	0.5
UNR Grande (Conchos)	8.8	0.6	--	0.6	0.6	0.7	0.8	0.7	0.5
Casas Grandes	8.6	5.7	5.9	--	0.2	0.2	0.7	0.6	0.1
Santa Maria	8.1	6.6	6.6	2.6	--	0.2	0.8	0.7	0.1
Del Carmen	7.9	5.7	5.7	1.3	2.0	--	0.9	0.7	0.2
Piaxtla	8.9	7.6	7.2	7.6	7.9	7.5	--	0.4	0.6
Nazas	9.3	8.1	7.7	7.8	7.9	7.8	1.4	--	0.5
USR Grande (Palomas)	8.5	6.3	6.3	1.7	2.6	1.6	7.4	8.0	

UNR = Upper North River, USR= Upper South River

Table 4. Spatial Analysis of Molecular Variance (SAMOVA) of the *Catostomus plebeius-nebuliferus* complex baes on *cytb*|*GHI*. Only the results for K = 2, 4, and 8 are shown, as they are meaningful in phylogenetic, systematic, and biogeographic terms.

Correspondence	**Groups**	Φ_{CT}	Φ_{SC}	Φ_{ST}
One genetic pool (via AMOVA)	--	--	0.9028*|0.6904*	--
Catostomus plebeius/*Catostomus nebuliferus*	(Palomas/Casas Grandes/Santa Maria/Del Carmen/Conchos/Fuerte/Mezquital) (Piaxtla/Nazas)	0.3545*|0.1707	0.8742*|0.6606*	0.9188*|0.7185*
Phylogenetic inference (Lineages)	(I) (II) (III) (IV)	0.7685*|0.6166*	0.6324*|0.2715*	0.9149*|0.7207*
Best genetic structure identified without geographic information	(Pamolas) (Casas Grandes) (Santa Maria) (Del Carmen) (Conchos/Fuerte) (Piaxtla) (Nazas) (Mezquital)	0.8784*|0.7460*	0.2051*|0.2156	0.9033*|0.6912*

* $p < 0.05$

Table 5. Genetic differentiation using pairwise Φ_{ST} for *cytb* (under the diagonal) and *GHI* (above the diagonal) among basins.

	Nazas	**Piaxtla**	**Mezquital**	**Fuerte**	**Conchos**	**Casas Grandes**	**Del Carmen**	**Santa Maria**	**Palomas**
Nazas	--	**0.532**	**0.742**	**0.722**	0.656	**0.640**	**0.723**	**0.572**	0.619
Piaxtla	**0.461**	--	**0.875**	**0.897**	0.914	**0.783**	**0.907**	**0.596**	0.900
Mezquital	**0.896**	**0.931**	--	**0.896**	0.899	**0.801**	**0.903**	**0.657**	0.890
Fuerte	**0.899**	**0.968**	**0.936**	--	0.050	**0.777**	**0.887**	**0.635**	0.868
Conchos	0.874	0.990	0.917	0.473	--	0.710	0.894	0.412	1.000
Casas Grandes	**0.889**	**0.940**	**0.923**	**0.921**	0.911	--	**0.314**	0.013	0.327
Del Carmen	**0.888**	**0.960**	**0.922**	**0.938**	0.934	**0.649**	--	**0.178**	0.789
Santa Maria	**0.884**	**0.951**	**0.914**	**0.936**	0.921	**0.795**	**0.777**	--	0.200
Palomas	0.872	1.000	0.908	0.951	0.956	0.655	0.763	0.780	--

Bold format = $p < 0.05$ after Bonferroni correction.

Supplementary information

Table 1S. Evolutionary substitution model and estimated parameters for *cytb* and *GHI* by Akaike Information Criterion.

Gene	Evolutionary model	lnL	Gamma
cytb	GTR+G	-638.643	0.17
GHI	TrN+G	-1453.496	0.33

Note: lnL = log likelihood

Capítulo 3: New insight in the evolutionary history and populations of the recent discovered hybrid complex of the fish genus *Squalius* Bonaparte, 1837 (Actinopterygii: Cyprinidae)

DIUSHI KERI CORONA-SANTIAGO and IGNACIO DOADRIO

Departamento de Biodiversidad y Biología Evolutiva, Museo Nacional de Ciencias Naturales. CSIC. c/ José Gutiérrez Abascal, 2. 28006 Madrid, Spain.

For submission in: Molecular Phylogenetics and Evolution

Abstract

Polyplidization in animals including fish can arise by failure of cell division during meiosis, failure of cell division after mitotic doubling, production of unreduced eggs and hybridization between species. In some cases, the hybridization has given rise to complete fish families. In 2009 was reported a morphotype that could be consist in a new unisexual hybrid complex in sympatry with *Squalius alburnoides* in the Guadiana basin, specifically in the Estena River (Spain), and with reproductive modes seems to *S. alburnoides*. This study includes an exhaustive sampling effort to determine the distribution of the new complex using mitochondrial and nuclear markers and to estimate its genetic structure and divergences with the aim of understand its evolutionary history. We corroborate the *Squalius* sp as a new hybrid complex conformed by polyploidy males and females with a genome R that represent and unknown ancestral species actually an with high genetic divergences respect to its sister groups in the south of Portugal. The complex *Squalius* sp is distributed in just 4 populations in the middle Guadiana. Its evolutionary history is associates with the historical events (mainly geological process) since Late Miocene and Pliocene. Our results suggest that the complex could suffer at least one event of hybridization in the last 700 000 year by the presence of the mitochondrial genome P of *S. alburnoides* in the new complex. Low genetic structure was observed in the genome R of the complex and models with low suitable habitat where predicted for the future. The genetic information obtained in this work reflects the necessity to apply efficient programs for conservation, as well as a formal description of the new complex hybrid.

Introduction

The duplication of genes and genomes was postulated to be an important process for the evolution of functional and organismal diversity long before we entered in the genome sequence era (Qian and Zhang, 2014). The rising of polyploidy groups implies evolutionary mechanism as the pseudogenization, subfuntionalization, neofunctionalization and the interplay between these (Force *et al.,* 1999; He and Zhang *et al.,* 2005; Näsvall *et al.,* 2012), however the one of the most important remarks of the implications of the polyploidization is increase the potential of evolutionary lines to produce organismic diversity (Ohno, 1970). In fact, the evolution by whole genome duplication has been related with some of the principal radiation of vertebrates as occur for the Teleost fishes (Glasauer and Neuhauss, 2014). Previous studies suggest that the more basal lineages of the Actinopterygii revealed that their divergences from the stem lineage of fishes preceded the event of duplication called 3R (Kuraku and Meyer, 2010). However, the first evidence of the evolution by genome duplication was the identification of higher number of genes for some gene families in teleost fishes than tetrapods (Wittbrodt *et al.,* 1998), and the intragenome redundancy was later revealed for teleost as well (Meyer and Van de Peer, 2005). Polyplidization in animals including fish can arise by failure of cell division during meiosis, failure of cell division after mitotic doubling, production of unreduced eggs and hybridization between species (Otto & Whitton, 2000).

The hybridization is extended in most of the species groups in the world (Cunha *et al.,* 2011), however, is the process with less attention in the ways for the origin of new species, being the allopatric and sympatric speciation the most common process took into account to explain the species formation (Baack and Rieseberg, 2007; Abbott *et al.,* 2013). The hybridization is a mechanism less rare in the nature and is possible to detect thanks to the innovation and development of the molecular techniques. In fishes, the hybridization is a common process but resulting in sterile hybrids probably by the high number of species with external fertilization (Mallet, 2008). In some cases, the hybridization has given rise to complete fish families (i.e. Salmonidae and

Catostomidae in fish) by mechanism as allopolyploidization as singles species as *Poecilia formosa* (Girard, 1859) in America or *Squalius alburnoides* (Steindachner, 1866) in the Iberian Peninsula involving the evolution by whole genome duplication (Leggatt and Iwama, 2003). In the family Cyprinidae, *Squalius alburnoides* is probably the best know hybrid complex. This complex is endemic of the Iberian Peninsula that consists in diploids (2n = 50), triploids (3n = 75) and tetraploids (4n = 100) individuals with a high prevalence of females (3n) over males (Cunha *et al.,* 2011). Use several reproduction modes as parthenogenesis and gynogenesis depending of the gametic configuration (haploids or not reduced diploids) resulted of the polyploidy condition of the females (di- tri- or tretaploids) (Alves *et al.,* 2001), and in all the cases using the meiotic hybridogenesis (Pala and Coehlo, 2005; Crespo-López *et al.,* 2006). The use of hibridogenesis maintains intact one of the genomes (Genome A), which correspond to an extinct species evolutionary related with *Anaecypris hispanica,* another endemism from the Iberian peninsula (Crespo-López *et al.,* 2006; Robalo *et al.,* 2006). However, no diploid hybrid individuals with the ancestral genomes can be recovered through the fecundation haploid gametes that in general are males.

In 2009 was reported a morphotype that could be consist in a new hybrid complex in sympatry with *Squalius alburnoides* in the Guadiana basin, specifically in the Estena River (Spain) (Cunha and Doadrio, 2006), and characterized as a unisexual complex with the absence of males, distributed in medium course of the Guadiana basin. This hybrid fish present a characteristic morphological traits exhibit as big and deciduous scales, and coloration differences as gold color, which has given it the name of Gold Calandino (Calandino Dorado in spanish). The previous information based on mitochondrial *cytb* and the nuclear intron *S7* showed that the new complex consist in a group that use reproductive strategies as *S. alburnoides*, with polyploidy individuals (triploids with high rate and diploid, all females) and possibly using the same modes of reproduction. The presence of the mitochondrial genome related with *S. pyrenaicus* suggested an actual reproduction of this new complex that we will call in all the work as *Squalius* sp with *S. pyrenaicus* and with *S. alburnoides* by the presence of nuclear alleles of this species in the new hybrid complex. In addition, the species *Squalius* sp posses a genome that we call R that was related with *S. torgalensis* and *S. aradensis* and could be represent an extinct member of the genus *Squalius* in the peninsula. However, not clear divergences were established between the genome R respect the sister species.

The distribution of *Squalius* sp describe the population of the new complex as reduced during the last decade and contrary to the extended distribution of the complex *S. alburnoides* (Doadrio 2002; Doadrio *et al.,* 2011). As occur in *S. alburnoides*, the complex *Squalius* sp is a complete field of study and for begin the first research about it, and represent a challenge for the future evolutionary researches by the difficult to identify the complex in field due to the diverse morphotypes of *S. alburnoides*. Thus, this study includes an exhaustive sampling effort to determine the distribution of the new complex using mitochondrial and nuclear markers and to estimate its genetic structure and divergences with the aim of understand its evolutionary history. We aim was also to know the conservation status of the populations of *Squalius* sp, at the present and predictions of habitat suitable toward the future. For the last, we want establish the criteria for a formal description of the new polyploid hybrid complex. Taking into account the complexity of the strategies for the study we test different method as design primers for orthologous and paralogous genes, and also it has been tried to perform *in-vitro* crosses to determine the modes of reproduction but without success. Therefore, in this work we show the field results obtained and the genetic information collected.

Material and Methods

Sampling and DNA extraction

The complete distribution of the new complex *Squalius* sp was sampled for a total of 52 tributaries of the middle course of Guadiana Basin (Fig. 1). Because the morphology was not determined consistently in previous work, we take samples of different morphotypes of *Squalius alburnoides* of different locations of the middle Guadiana*,* and those in accordance with the original description of *Squalius* sp (Cunha and Doadrio, 2011). In all the cases was obtained a tissue samples of pectoral fin and stored in absolute alcohol, while some complete specimens where stored for posterior morphological studies in the fish collection of the Museo Natural de Ciencias Naturales, Madrid, Spain.

The extraction of the DNA extraction was implemented using the QIAGEN Blood extraction kit and performing the methodology from manufacturer. Two molecular markers where obtained for this work, the complete mitochondrial Cytochrome b *cytb* and the nuclear intron *S7*. In the case of *cytb*, primers designed in previous works; GLuDG (Palumbi *et al.,* 1991) and H16460 (Perdices *et al.,* 2002), were used for the amplification using the following protocol: 2 min at 94ºC followed by 35 cycles of 45 seg at 94ºC for DNA denaturation, 1 min at 46ºC for primer alignment, 1.5 min at 72ºC for synthesis, and a final extension of 5 min at 72ºC.

The amplification of the intron *S7* was the strategy to identify between *Squalius alburnoides* and the complex, designing specific primers for ortologous and paralogous alleles. Two methods where designed since cloning with the method described in previous work (Sambrook, Fritsch and Maniatis, 1989): (1) primers to obtain the ortologous alleles corresponding for the genome R of *Squalius* sp, and (2) primers to obtain the paralogous genomes resulted of the hybridization (genome A of *S. alburnoides* and P of *S. pyrenaicus*). The primers used to amplify the *S7* of R were *S7F2tor* with sequence For-5'-GTACAAAATAAGTAACATTTTCCTCC-3' and *S7R3tor* with the sequence Rev-5'-CCCTCACTTTGTTCCAAACCTGAA-3' with the follow protocol: 2 min at 95ºC followed by 35 cycles of 30 seg at 94ºC, 1.5 min at 52.6ºC, 2 min at 72ºC, and 7 min at 72ºC for final extension. . When PCR amplify for S7 with primers, the individuals could be *Squalius* sp (genome R) or individuals *S. alburnoides* AP or AAP amplifying only the genome P, not A. For amplify the *S7* in the genome A were used the primers called *S7F1* (Chow and Hazama, 1998) and S7albR designed in this work with the sequence Rev-5'-CATTTACTAACACTCACTTTGTTCCA-3' using the next protocol: 2 min at 95ºC followed by 35 cycles of 30 seg at 94ºC, 1.5 min at 54.1ºC, 2 min at 72ºC, and 7 min at 72ºC for final extension.

Total PCR product where sent to Sanger sequencing in Macrogen Inc. (Netherlands) for their posterior analyses.

Genetic diversity and structure

The genetic variability was estimated using DNAsp v 5.0 (Librado and Rozas, 2009) for all the populations collected for the complex *Squalius* sp in both molecular markers: haplotype diversity *h*, nucleotide diversity *π* and numbers of haplotypes H. For nuclear region *S7* haplotypes by InDels (insertion or deletion) was determined and a test of recombination was estimated using DNAsp v 5.0 with coalescent algorithm with 1000 replicates. In addition, uncorrected genetic *p*-distances were estimated for both molecular markers using pairwise-deletions method for not remove the gaps. Genetic structure within the complex was determined using Analysis of Molecular Variance AMOVA with the software ARLEQUIN v 3.5.1.3 (Excoffier and Lischer, 2010).

Phylogenetic inference and haplotype network

Tree reconstruction was performed using two analyses for both molecular markers independently: Bayesian Inference (ML) and Maximum Likelihood (ML). For this, the estimation of substitution model of evolution using JModeltest v v2.1.6 (Darriba *et al.,* 2012) that was Generalized time-reversible (Tavaré, 1986) + Invariable sites proportion GTR+I for *cytb* and Generalized time-reversible + Gamma distribution GTR+GAMMA for *S7*. The ML analysis was implemented in RAxMLGUI v. 1.3.1 (Stamatakis, 2006; Silvestro and Michalak, 2012) applying 10000 bootstraping replicates. For the BI analysis, we use the software Mrbayes v v. 3.2.6 (Ronquist and Huelsenbeck, 2003) performing 4 MCMC for 5 000 000 generations and sampling each 100 generations. In this case, a phylogeny was constructed for the genus *Squalius* including sequences from the GenBank of species members of the Leuciscinae subfamily (Table A1, of the Supplementary information) to identify the position and the phylogenetic relationships of the nuclear and mitochondrial genomes of the complex *Squalius* sp within the genus.

Haplotype networks were constructed using TCS v 1.21 (Clement, Posada and Crandall; 2000) in both cases to take into account the variation that gaps add to the genetic variation. In the case of the *cytb* gene, samples of *Squalius alburnoides* and *S. pyrenaicus* were included to observe the shared haplotypes between species, and mutations steps between the *Squalius* sp complex and these species.

Divergence time estimation between genomes

The calculation of the time of divergence between the genomes that possess *Squalius* sp respect to the sister species was performed using the software BEAST v 1.8.0 (Drummond *et al.,* 2007) in two independents analyses for mitochondrial and nuclear DNA sequences. The sequences included where the same used for phylogenetic analysis to include fossil record for the molecular clock calibration. In this case, two fossil records where included, one that correspond to the main node of the genus *Squalius* date on *ca.* 13MYa in the Miocene stage (for both molecular markers), and the second one was established in the main node for *Scardinius* genus dated in *ca.* 10MYa, also for the Miocene period (Machordom and Doadrio, 2001; Cunha *et al.,* 2002; Doadrio and Carmona, 2003). In addition, for calibration of molecular clock for the mitochondrial gene, the mutation rate estimated for teleost fishes (Zardoya and Doadrio, 1999) was included: 0.76% - 2.2% of genetic divergence per million year between lineages. For *S7* and *cytb* region, 80 000 000 and 60 000 000 millions of generations were applied respectively, and sampling each 1000 trees. In both analyses, Tracer v 1.5 was implemented to evaluate the effective sample size (ESS) for posterior parameters. The Maximum Clade Credibility Tree was constructed and discarding the 10% percent of states in each analysis.

Predicting ecological niches for Squalius sp in different periods

The modeling of suitable habitat for the hybrid complex *Squalius* sp was implemented using the maximum-entropy approach with MaxEnt v 3.3.3 (Phillips, Anderson and Schapire, 2006; Phillips and Dudík, 2008). The predictions were obtained for 1 period in the past (Middle Holocene; ca. 6000 years ago), for the present, and two period in the future (2041-2060 and 2061-2080) using just bioclimatic variables available (http://www.worldclim.org/bioclim) in all the cases used. The prediction for the past is in accordance of the Holocene climatic optimum characterized as a warm period but with high wet in some regions as the Iberian Peninsula and the North Africa increasing the number and level of the body water (Bartlein *et al.,* 2010). For the future projections the 6.0 Representative Concentration Pathways (RCP6.0) (Fujino *et al.,* 2006) was selected from downscaling Intergovernmental Panel on Climate Change 5 scenario (IPPC5) (Coupled Model Intercomparison Project 5) (Taylor, Stouffer and Meehl, 2012) using the model Community Climate System Model 4 (CCSM4) for both periods (Moss *et al.,* 2010). Before the modeling, the selection of the not correlated bioclimatic variables was conducted for present condition in the Iberian Peninsula performing a correlation analysis in R packages v 3.3.3 (R Core Team, 2013), and discarding one of the variables which Pearson coefficient of correlation between pairs was higher than 0.8 (Rissler and Apodaca, 2007). Then, the collinearly correlated variables were discarded conducting a Variance inflation factor with the package VIF included in the package HH: Statistical Analysis and Data Display (Heiberger and Holland, 2004) of the R package.

The MaxEnt analyses were conducted performing with runs of 5000 iterations with the 6 bioclimatic variables selected, bootstrapping of 100 replicates to obtain 95% confidence intervals for probability in each model, regularization multiplier values from 0.01 to 1, and evaluating area under to ROC curve (AUC) values and using the omission rate to discard identical AUC in different RM values for each projection.

Results

Populations sampled and Squalius sp determination

In the 52 localities of the middle Guadiana basin sampled, only in 4 was present *Squalius* sp (Fig. 1): Jola river (header tributary of Gévora River), Higuera Stream (header tributary of Matachel River), in the header of Guadarranque River (that lead directly in the Guadiana River) and in the header of the Estena River (out in the Cijara reservoir of the Guadiana R.). A total of 233 samples were sampling which 91 (Table 1) correspond to *Squalius* sp. The population of Squalius sp with the low number of samples obtained was Higuera Stream with 2 specimens observed, while the population with the high number of specimens observed was the Jola River, followed of the population of Guadarranque River. Males individuals could be determined during the sampling activity by rub each individual. When the PCR products were sequenced, the genome P was not isolated in any individual with the genome R. Therefore, we can determine that the sequences of the genome P obtained in all the localities represent individuals of *S. alburnoides*, and the genome P is absent in *Squalius* sp.

When PCR was conducted to amplify the genome A in *Squalius* sp, some individual do not have this genome, in fact some individuals of *Squalius* sp possess two alleles of R. In addition, polyploid individuals of *Squalius sp* with an allele A were determined when heterozygous RR could be determined and determining the presence of genome A of *S. alburnoides* in *Squalius* sp. However it is not possible to determine the polyploidy in some cases with this method, i.e. when homozygous RR is present and an allele A is obtained.

Phylogenetic relationships and genetic divergences

No significant recombination was observed in the *S7* sequences. The BI and the ML were congruence respect to the relationships observed (Fig. 2 and Fig. 3). In the case of the phylogenetic tree based on S7 (Fig. 3), was observed that the genome R of *Squalius* sp is more related with the species *S. aradensis* and *S. torgalensis* while the genome A found in the *Squalius* sp is related with the ancestral genome A of *S. alburnoides* which sister species is *Anaecypris* *hispanica*. Any sequences of *S7* intron obtained with each pair of primers were related with *S. pyrenaicus*. The genetic divergences estimated to S7 between the genome R respect to *S. aradensis* and *S. torgalesis* was 2.8% and higher than 9% respect to the Genome A. The Genome A of *Squalius* sp has a genetic divergence of 1.3% respect the genome A of *S. alburnoides*.

The phylogenetic relationships based on *cytb* suggest that the genome P *of Squalius sp* is related to *S. pyrenaicus* and *S. alburnoides*, whit a genetic divergence of 0.8% and 0.4% respectively. The genetic distance between the mitochondrial region of *Squalius* sp respect to *S. aradensis* and *S. torgalensis* is 10.1% and 11.2% respectively.

Divergence times for the different genomes

Large ranges were estimated for both molecular markers in the divergence time estimation. The analysis shows that the mediterranean lineage (sensu Sanjur *et al.,* 2003) of the Iberian Peninsula was isolated during the Late Miocene and the early Pliocene (10.1- 3.7Ma) (Highest posterior density HPD ≥ 95%) (Fig. 2). The separation between the genome R respect to the clade which is located *S. alburnoides* and *S. pyrenaicus* occurred 8.4-2.1Ma during the Late Miocene and Pliocene. The genome R of *Squalius* sp was isolate during the last 5.9Ma to 0.5Ma during the Late Miocene to Pleistocene from the species *S. aradensis* and *S. torgalensis*. Moreover, The divergence time between genome A of *Squalius* sp respect with the *S. alburnoides* was estimated for Pliocene and Pleistocene, however could be underestimated due to the shared haplotypes observed in the haplotype networks or the highly divergence proportionated by gaps, difficult to estimated and the models of evolution of this polymorphism.

In the case of divergence times based *cytb*, the genome P was isolated during the Pleistocene (HPD ≥ 95%) (Fig. 3) *ca.* 2.0 to 0.2Ma from the *S. pyrenaicus* or *Squalius alburnoides*, while the Portuguese endemism could be isolated before this date, during the Late Miocene, 10.5-5.6Ma.

Genetic diversity and structure in populations

As we could expect the genetic diversity estimated in the four populations of *Squalius* sp was lower in the region *S7* than in *cytb* (Table 1). The number of haplotypes ranges from 1 to 20 for the mitochondrial *cytb*, while 1 to 9 in the *S7*. For *cytb*, the population with more than two individuals with more haplotypes was Guadarranque with 20 haplotypes in 27 individuals obtained. Only ten haplotypes of *cytb* were observed in the population of Jola when we sampled 40 individuals, being the population with less haplotype diversity without takes into account the population of Higuera Stream. In the case of the sequences of *S7* intron, the population with less haplotype diversity and with high number of individuals studied was Jola with h = 0.766, while the number of haplotypes observed in the Guadarranque and the Estena river was the same (H = 9) in n = 20 and n = 14 respectively. In general, the nucleotide diversity was low in all the populations ranging from 0.001 to 0.004 in *cytb* and from 0.002 to 0.004 in *S7*. In addition, polymorphisms between sequences of S7 due insertion or deletion (InDels) were observed in all the populations of the *Squalius* sp.

Haplotype netowork based on *S7* for the genome R of *Squalius* sp not show a geographic congruence of the haplotypes with haplotypes of the four populations shared between these but with the rising of unique haplotypes in each population (Fig. 4). The high number of mutation steps was observed between one haplotype from Guadarranque and one Jola haplotype. The rest of haplotypes were separated by 1 to 2 mutation steps each other. In contrast, the haplotype network based on *cytb* shows haplotypes shared between genome P of *S. alburnoides* but not with *S. pyrenaicus*. As in S7, the network not showed a geographic congruence but only one haplotype is shared between populations from Jola and Guadarranque the most geographically distant populations. The haplotypes of *S. alburnoides* are not shared in any case with the *Squalius* sp population from Guadarranque, as occur in the rest of populations. Between 1 to 4 mutations steps separated the mitochondrial haplotype of the genome P of *Squalius* sp, including the haplotypes of *S. alburnoides*.

Low percent of variation but significant (Φ_{ST} = 0.050 and 0.312 for *S7* and *cytb* respectively; p < 0.05) was observed between populations of *Squalius* sp in both markers (Table 3). This suggest that the best cluster to explain the genetic variation in the hybrid complex is when only one genetic pool is took into account, because the percentage of variation was higher within populations than between them. This significant genetic differentiation was estimated between the populations Jola respect to Estena and Guadarranque for both markers (Table 4). Although the Φ_{ST} values were relatively high between the populations from La Higuera Stream respect to Estena and Jola Rivers, they were not significant. In addition, not significant genetic structure was observed between the populations from the Estena and Guadarranque rivers.

Prediction of suitable habitat for Squalius sp

Evaluating with the AUC values, the four projections was estimated using regularization multiplier value of 0.01. In general, all the projections show model of ecological niches with low probability (Fig. 5). The prediction for the present conditions suggest a model where the suitable habitat is located in headers of the south part of the middle Guadiana, around to the known population in La Higuera Stream but with low probability. In fact, no suitable conditions were predicted to the present in the rest of the Guadiana River including the actual populations as those form Jola, Guadarranque and Estena rivers. The projection for 6000 years ago suggest a more wide ecological niches including for the regions around from Jola River and La Higuera Stream in the middle of Guadiana River but with low probability. As occur in the prediction of the present, the Guadarranque and Estena Rivers were regions with null suitable habitats in the past. The projections in the future periods showed that the possible ecological niches could increase during the 2041-2060 but decreasing in the period of 2071-2080, but in all the cases the probability was low. Again, as was observed in all the rest of projections, the Rivers as Jola, Guadarranque, Estena, the Upper Guadiana and the Lower Guadiana could not be suitable areas to habit in the Future.

Discussion

We corroborate the *Squalius* sp as a new hybrid complex similar to *S. alburnoides*, but with some different evolutionary traits. We describe these characteristics in this work and will open new ways to the research in *Squalius* hybrid complexes. Formal description of *Squalius* sp as a new species of hybrid origin due take account the results of our research. The difficult to describe as a independent evolutionary lineage or species to *Squalius* sp is linked to the hybrid, morphotypes and polyploidy conditions. In this study, we recognize characteristics that could be taken into account to recognize to *Squalius* sp as a new species among them the possession of an exclusive ancestral genome that we call R in all individuals and that is highly divergent respect to their phylogenetycally closer related species.

Evolutionary history of the hybrid complex Squalius sp

The Pliocene has been considered an important period of diversification for Iberian cyprinids in general, as well as other freshwater fauna (Zardoya and Doadrio, 1998; Ribera and Vogler, 2004; Robalo *et al.,* 2006; Sousa-Santos, Collares-Pereira and Almada, 2007, 2014; Almada and Sousa-Santos, 2010: Perea, Cobo-Simon and Doadrio, 2016);

The isolation of the Mediterranean clade of *Squalius* species in the Late Miocene and early Pliocene (Fig. 2) in the Iberian peninsula is in according to previous works that placed the separation of mediterranean an the northern lineage in the late Miocene (Sanjur *et al.,* 2003, Perea *et al.,* 2010) (Fig. 2). This separation has been postulated as consequence of the end of Alpine Orogeny (Perea *et al.,* 2011). The isolating of *Squalius* sp lineage in the Guadiana basin during the Pliocene and Pleistocene of *Squalius* sp resulted of the disconnection of the Guadiana River from Proto-Mira/Arade, reflected in high genetic distances in both molecular markers (i.e. 2.8% *S7* with to *S. aradensis* and *S. torgalesis* respectively; Table 2) and at the same time, the isolation of *S. pyrenaicus* from his sister group *S. valentinus* which is consistent with the geological history of the region as the transformation of the large Iberian Miocene endorheic lagoons into the current exoreic basins (Calvo *et al.,* 1993;

Pais *et al.,* 2012). In addition the isolation of the Guadiana has been supported in previous works of the genus (Coelho *et al.,* 1995; Carmona *et al.,* 2000; Doadrio and Carmona, 2003; Sanjur, Carmona and Doadrio, 2003, Mesquita *et al.,* 2005; Perea, Cobo-Simón and Doadrio, 2016). The last suggest that the *S. pyrenaicus* and the *Squalius* sp not coexisted during enough time to speciation in the Guadiana. Our data placed the origin of *Squalius* sp. complex during the Late Miocene when endorheic basins are present in the Iberian Peninsula. The current distribution of *Squalius sp*. complex is in agrees with the tertiary deposits than placed the main endorheic Guadiana basin (Friend & Dabrio, 1996; Tejero & Garzón-Heydt, 2014; Tejero *et al.,* 2017). (De Vicente *et al.,* 2011; Tejero *et al.,* 2017). The formation of the basin could have commenced with the first Alpine movements in the peninsular interior under N-S stresses that were to act from the Oligocene to Upper Miocene (Tejero & Garzón-Heydt, 2014). Subsequently, the stress field was to change to NNW and NW from the end Miocene to the present and give place to the actual configuration of the Guadiana basin (De Vicente *et al.,* 2011). This distribution of *Squalius* sp associated to tertiary Guadiana basin is also shared by *Anaecypris* (Steindachner, 1866) a small cyprinid fish with ancient origin in the Guadiana basin (Zardoya & Doadrio, 1998).

The discrepancy between isolation events of genomes (nuclear genome R and mitochondrial P) of *Squalius* sp is indicative that hybridization by *S. alburnoides* and/or *S. pyrenaicus* occur after the isolation of the ancestral *Squalius* sp in the Guadiana basin. The origin and the speciation event of *S. pyrenaicus* in the Pliocene, suggest that the presence of *S. pyrenaicus* in the Guadiana Basin it dates back to the origin of the basin in simpatry with *Squalius* sp, allowing the hybridization. However, we could expect that *Squalius* sp have a complete nuclear genome of *S. pyrenaicus* as occur in *S. alburnoides* (Cunha *et al.,* 2011) and not only mitochondrial haplotypes of the genome P shared with *S. alburnoides*, but the nuclear genome P is absent in *Squalius* sp. Thus, we can suggest two hyphotesis of hybridization and origin of the complex *Squalius* sp with *S. pyrenaicus*: (1) the hybridization occur since the cohabitation of both species which Squalius sp acquired the mitochondrial genome P and probably the nuclear genome P since the Late Miocene and Pliocene, which was substituted subsequently when hybridization with *S. alburnoides* occur in the

Pleistocene, involving a reproductive isolation of *Squalius* sp with *S. pyrenaicus*; (2) the hybridization between *Squalius* sp and *S. pyrenaicus* never occurs, and the genome P present in *Squalius* sp is result of the hybridization in the Pleistocene (as was estimated in the present work 0.2 to 2.0MYa, Fig. 3 with higher the 10% of divergence with respect the Portuguese sister groups; Table 2) with *S. alburnoides* which obtained the genome P from *S. pyrenaicus* in a hybridization event dated for the last 700 000 years in the Tagus-Guadiana (Sousa-Santos *et al.,* 2007; Cunha *et al.,* 2011), with the subsequent dispersion for the north basins of the peninsula. The recent connection between the Tagus and the Guadiana is not clear; there is evidence that the southern Tagus and the Sado drainages have been connected with a common estuary since earlier Pleistocene until present (Azevedo and Cabral, 1986; Brito *et al.,* 1997), this last river considered an affluent of the Guadiana River until the middle to late Pleistocene (Sousa-Santos *et al.,* 2007). In addition, a study suggests that the Upper Tagus and the Upper Guadiana were connected since Pliocene, and in the Pleistocene (*ca.* 0.53 MYa), a secondary contact occur in the lower sections of both drainages (Moya-Palomares *et al.,* 2002; Sousa-Santos *et al.,* 2007).

In conclusion, the new hybrid complex *Squalius* sp could be originated during the Late Miocene by a first hybridization suggestion that if it is possible we are facing the oldest fish hybrid complex of the Iberian Peninsula, with a second hybridization in the Pleistocene involving several unknown biological process that promote the reproductive isolation with *S. pyrenaicus* sometime in the past. On the other hand, *Squalius* sp could represent a second complex originated at the same time of *S. alburnoides* arise too, during the Pleistocene.

In the last decade has been considered the existence of a causal relationship between Teleost whole genome duplication event 3R TSGD and the phenotypic diversification and the speciation of teleost fishes (Kuraku and Meyer 2010). However has been observed that the early-branching lineages after the TSGD (Osteoglossomorpha and Elophomorpha) show relative low species diversity been a case against a link between TSGD and the massive species diversification. Therefore, if the speciation and organismic diversification is not a common process involved with the evolution by whole genome duplication, which process are they? First, the gene duplication implies benefits and cost. Shorter-term benefits include advantages of increased gene dosage and thus increased gene expression (Wagner, 2010). Such advantages may exist both for gene products that are in extremely high demand in a cell, and for genes that are expressed at very low levels when in single copy (Bar-Even *et al.,* 2006). Another short-term benefit arises in cases where the new duplicated gene differ in function and sequence and it is beneficial to the cell, its carrier may rise in frequency through natural selection (Vinckenbosch *et al.,* 2006; Wagner, 2010). In this case, the evolution of *Squalius* sp and *Squalius alburnoides* could be associated to historical process that led into to acquire strategies of gene duplication for a rapid adaptation to changing environments. The region where is situated the Guadiana River has been characterized by extremely climatic changes which cause the desiccation of the Guadiana by the increment of the temperature higher than 40ºC, isolation the fish populations in temporally pounds until the pluvial periods that reconnect the tributaries (Zamora, 1987). This characteristic of the river including the interglacial period ca. 700 000, (Prieto, 2012) and the geological process that isolate the Guadiana basin mentioned above, could be the process associated for the adaptation of new strategies for adaptation of *Squalius*. However, a lot of questions will be arising for this, as why *Squalius* sp not have the same “success” in the wild with a reduced distribution and populations? One answer for this is because *S. alburnoides* move the populations of the ancient *Squalius* sp, not by the high dispersal capacity of *S. alburnoides* (Cunha *et al.,* 2011) but by lower possibility of reproduction with individuals of the same species associated to the high density of S. alburnoides in the basin and the external fertilization used in both cases.

Characterization of populations of Squalius sp and conservation implications

The complex *Squalius* sp consist, unlinked to found previously (Cunha and Doadrio, 2009), in males and females with an genome R is inherited hemiclonally using hybridogenesis as occur in *S. alburnoides* with the genome A (Crespo-López *et al.,* 2007) and present in all the nuclear sequences obtained in this work (Fig. 4). Taking into account the polyploid ratio obtained previously (Cunha and Doadrio, 2009) and our results, we can reconstruct a model of the strategies of reproduction of the new complex. The *Squalius* sp complex have a genome A by the interaction and hybridization event with *S.*

alburnoides but some individuals not have the genome A, only the R*,* suggesting that the new complex use reproductive strategies as *S. alburnoides* (Cunha *et al.,* 2011), where haploid gametes (R and A gametes) in triploid females are produced by hibridogenesis allowing the restoration of non-hybrid individuals. Due to the absence of the genome P (Fig. 4), the number of polyploidy configurations is reduced drastically, suggesting that the only triploids and diploid reported in 2009 could be: RR, RA, RRA and RAA. We do not found triploid with configuration RRR and in any case, and is consistent with the absent of AAA *S. alburnoides* in the Guadiana River, and the absence of tetraploid in this basin, and the high rate of triploid (Cunha *et al.,* 2011). In summary, as occur in *S. alburnoides*, the rate of males of *Squalius* sp is very low in the field and we were able to determine them by rub each individual collected (6 males in 91 individual sampled in the Guadarranque River), localized in the shady and fast areas of the river headwaters, always in absent or low presence of *S. alburnoides*.

The distribution of the complex is very reduced, and in this work four populations were determinate with different abundance, and separated by distance provably significantly. This is supported by the low but significant fixation index between Jola respect to Guadarranque and Estena estimated ($\mathbf{\Phi_{ST}}$ = 0.410 and 0.382 respectively, Table 4) in the genome R. The last suggest connectivity between populations in the past because the percentage of variation between populations is very low and lower than the percentage explained within populations (until 95% in the *S7*; Table 3), and the current structure could be influenced by the actual configuration of the Guadiana basin. In general, the Guadiana River have high transformation in the last century (Zamora 1987), and all the populations of the complex are situated in the header of the rivers, where the presence of exotic species is lower except the population in the Matachel. The population in La Higuera Stream in the Matachel River, is very low and in a high contamination habitat with high exotic and invasive observed in this work as *Lepomis gibbosis*, *Gambusia holbrooki* or *Exos lucius* in the most of the sampled points in the Guadiana. The construction of the dams in the entire basin allows the rapid colonization of the exotic species being one of the most important factors of risks for the aquatic ecosystem (Pérez-Bote and Pula, 2002).

According with our presence data taking into account only bioclimatic variables, our models of ecological niches suggest that suitable habitat with high probability currently are reduced or null, indicative of the vulnerability to climate fluctuations, unlinked to the relative more extensive distribution predicted for the past (Fig. 5). The same pattern was observed for the future conditions, being the suitable habitat reduced and in placed with high actual contamination and presence of exotics species. Is for this, that we recommend that the populations of the hybrid complex *Squalius* sp must be considered as an highly endangered group and Operational Conservation Units (Doadrio *et al.,* 1996) with the necessity of a formal description of the complex (Collares-Pereira and Coelho, 2010) including morphological information, sex ratio and polyploidy frequency, and model of reproduction, to apply efficient programs of natural resources management in the basin for conservation.

Finally, the taxonomic description of the complex is complicated for the fact of parental species belong to distinct lineages, being and nomenclatural problem still unresolved due to their hybrid condition considered as *S. alburnoides* (Gregory and Mable, 2005) from now the second and one of the most complicated polyploidy systems discovered in vertebrates.

Acknowledgments

Thanks to Daniela Galván, Hamid Ghanavi, Elena González, Rosa Gabriela Beltrán and Carla Sousa for the support in the field. Thanks to Silvia Perea for the comments and suggestions for the work. This work was funded by the Red de Parques Nacionales 2012 and the Ministerio de Ciencia e Innovación (CGL2010-15231-BOS).

References

Abbott R, Albach D, Ansell S, Arntzen JW, Baird SJ, Bierne N *et al.* 2013. Hybridization and speciation. *Journal of Evolutionary Biology*. 26(2), 229-246.

Almada V, Sousa-Santos C. 2010. Comparisons of the genetic structure of *Squalius* populations (Teleostei, Cyprinidae) from rivers with contrasting histories, drainage areas and climatic conditions based on two molecular markers. *Molecular Phylogenetics and Evolution*. 57(2), 924-931.

Alves MJ, Coelho MM, Collares-Pereira MJ. 2001. Evolution in action through hybridisation and polyploidy in an Iberian freshwater fish: a genetic review. *Genetica*. 111(1-3), 375-385.

Azevedo MT, Cabral J. 1986. Estatigrafia e TectoÂnica da penõÂnsula de SetuÂbal. *Actas do. 2*.

Baack EJ, Rieseberg LH. 2007. A genomic view of introgression and hybrid speciation. *Current Opinion in Genetics Development*. 17(6), 513-518.

Bar-Even A, Paulsson J, Maheshri N, Carmi M, O'Shea E, Pilpel Y, Barkai N. 2006. Noise in protein expression scales with natural protein abundance. *Nature genetics*. 38(6), 636-643.

Bartlein PJ, Harrison SP, Brewer S, Connor S, Davis BAS, Gajewski K, Guiot J, Harrison-Prentice TI, Henderson A, *et al.* 2010. Pollen-based continental climate reconstructions at 6 and 21 ka: a global synthesis. Climate Dynamics, DOI: 10.1007/s00382-010-0904-1

Brito RM, Briolay J, Galtier N, Bouvet Y, Coelho MM. 1997. Phylogenetic Relationships within genus *Leuciscus* (Pisces, Cyprinidae) in portuguese fresh waters, based on mitochondrial DNA cytochrome b Sequences. *Molecular Phylogenetics and Evolution*. 8(3), 435-442.

Calvo Sorando JP, Daams R, Morales J, López Martínez N, Agustí J, Anadón P, *et al.* 1993. Up-to-date Spanish continental Neogene synthesis and paleoclimatic interpretation. *Revista de la Sociedad Geológica de España*. 6(3-4), 29-40.

Carmona JA, DomÍnguez J, Doadrio I. 2000. Congruence between allozyme and cytochrome b gene sequence data in assessing genetic differentiation within the Iberian endemic Chondrostoma lemmingii (Pisces: Cyprinidae). *Heredity*. 84(6), 721-732.

Chow S, Hazama K. 1998. Universal PCR primers for S7 ribosomal protein gene introns in fish. *Molecular Ecology*. 7(9), 1255-1256.

Clement M, Posada D, Crandall K. 2000. TCS: a computer program to estimate gene genealogies. *Molecular Ecology.* 9: 1657–1660.

Coelho MM, Brito RM, Pacheco TR, Figueiredo D, Pires AM. 1995. Genetic variation and divergence of *Leuciscus pyrenaicus* and *L. carolitertii* (Pisces, Cyprinidae). *Journal of Fish Biology*. 47(sA), 243-258.

Collares‐Pereira MJ, Coelho MM. 2010. Reconfirming the hybrid origin and generic status of the Iberian cyprinid complex *Squalius alburnoides*. *Journal of Fish Biology.* 76(3), 707-715.

Crespo-López ME, Duarte T, Dowling T, Coelho MM. 2006. Modes of reproduction of the hybridogenetic fish *Squalius alburnoides* in the Tejo and Guadiana rivers: an approach with microsatellites. *Zoology.* 109(4), 277-286.

Crespo‐López ME, Pala I, Duarte TL, Dowling TE, Coelho MM. 2007. Genetic structure of the diploid–polyploid fish *Squalius alburnoides* in southern Iberian basins Tejo and Guadiana, based on microsatellites. *Journal of Fish Biology*. 71(sc), 423-436.

Cunha C, Doadrio I, Abrantes J, Coelho MM. 2011. The evolutionary history of the allopolyploid *Squalius alburnoides* (Cyprinidae) complex in the northern Iberian Peninsula. *Heredity*. 106(1), 100-112.

Cunha C, Doadrio I. 2009. Evidencias de un nuevo complejo híbrido unisexuado del género Squalius (Teleostei, Cyprinidae) en el río Estena (Parque Nacional de Cabañeros): una aproximación filogenética. *Proyectos de Investigación en Parques Nacionales*. 357-375.

Cunha C, Mesquita N, Dowling TE, Gilles A, Coelho MM. 2002. Phylogenetic relationships of Eurasian and American cyprinids using cytochrome b sequences. *Journal of Fish Biology*. 61(4), 929-944.

Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nature methods*. 9(8), 772-772.

De Vicente G, Cloetingh SAPL, Van Wees JD, Cunha PP. 2011. Tectonic classification of Cenozoic Iberian foreland basins. *Tectonophysics*. 502(1), 38-61.

Doadrio I, Carmona JA. 2003. Testing freshwater Lago Mare dispersal theory on the phylogeny relationships of Iberian cyprinid genera Chondrostoma and *Squalius* (Cypriniformes, Cyprinidae). *Graellsia*. 59(2-3), 457-473.

Doadrio I, Carmona JA. 2003. Testing freshwater Lago Mare dispersal theory on the phylogeny relationships of Iberian cyprinid genera *Chondrostoma* and *Squalius* (Cypriniformes, Cyprinidae). *Graellsia*. *59*(2-3), 457-473.

Doadrio I, Carmona JA. 2004. Phylogenetic relationships and biogeography of the genus Chondrostoma inferred from mitochondrial DNA sequences. *Molecular Phylogenetics and Evolution*. 33(3), 802-815.

Doadrio I, Perdices A, Machordom A. 1996. Allozymic variation of the endangered killifish *Aphanius iberus* and its application to conservation. *Environmental Biology of Fishes.* 45(3), 259-271.

Doadrio I. 2002. Atlas y libro rojo de los peces continentales de Espana; Ambiente MdM, editor. *Madrid, Spain*.

Doadrio I. 2011. *Ictiofauna continental española: Bases para su seguimiento*. Ministerio de Medio Ambiente y Medio Rural y Marino, Centro de Publicaciones.

Drummond AJ, Ho SYW, Rawlence N, Rambaut A. 2007. *A rough guide to BEAST* 1.4. University of Edinburgh, United Kingdom.

Excoffier L, Lischer HE. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular ecology resources*. 10(3), 564-567.

Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J. 1999. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics*. 151(4), 1531-1545.

Friend PF, Dabrio CJ. 1996. *Tertiary basins of Spain: the stratigraphic record of crustal kinematics* (Vol. 6). Cambridge University Press.

Fujino J, Nair R, Kainuma M, Masui T, Matsuoka Y. 2006. Multi-gas mitigation analysis on stabilization scenarios using AIM global model. *Multigas Mitigation and Climate Policy*. *The Energy Journal*. 3 (Special Issue).

Gregory TR, Mable BK. 2005. Polyploidy in animals. *The Evolution of The Genome*. 171, 427-517.

Glasauer SM, Neuhauss SC. 2014. Whole-genome duplication in teleost fishes and its evolutionary consequences. *Molecular genetics and genomics*. 289(6), 1045-1060.

He X, Zhang J. 2005. Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. *Genetics*. 169(2), 1157-1164.

Heiberger RM, Holland B. 2004. Statistical Analysis and Data Display: An Intermediate Course: Accompanying Online Files.

Kuraku S, Meyer A. 2010. Whole genome duplications and the radiation of vertebrates, in evolution after gene duplication (eds) K. Dittmar and D. Liberles), John Wiley Sons, Inc, Hoboken, NJ, USA. doi: 10.1002/9780470619902.ch16

Leggatt RA, Iwama GK. 2003. Occurrence of polyploidy in the fishes. *Reviews in Fish Biology and Fisheries.* 13(3), 237-246.

Librado P, Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*. 25(11), 1451-1452.

López RT, Heydt GG. 2014. The Guadiana River: the role of tectonics on drainage configuration.

Machordom A, Doadrio I. 2001. Evidence of a Cenozoic Betic–Kabilian connection based on freshwater fish phylogeography (Luciobarbus, Cyprinidae). *Molecular Phylogenetics and Evolution*. 18(2), 252-263.

Mallet J. 2008. Hybridization, ecological races and the nature of species: empirical evidence for the ease of speciation. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 363(1506), 2971-2986.

Mesquita N, Hänfling B, Carvalho GR, Coelho MM. 2005. Phylogeography of the cyprinid *Squalius aradensis* and implications for conservation of the endemic freshwater fauna of southern Portugal. *Molecular Ecology*. 14(7), 1939-1954.

Meyer A, Van de Peer Y. 2005. From 2R to 3R: evidence for a fish‐specific genome duplication (FSGD). *Bioessays*, 27(9), 937-945.

Moss RH, Edmonds JA, Hibbard KA, Manning MR, Rose SK, van Vuuren DP *et al.* 2010. The next generation of scenarios for climate change research and assessment. *Nature*. 463. 747–756.

Moya-Palomares ME. 2002. *Evolución sedimentológica y geomorfológica de las Vegas Bajas del Guadiana entre Mérida y Badajoz (Espana)* (Doctoral dissertation, Tesis Doctoral. Universidade Complutense de Madrid, Madrid).

Näsvall J, Sun L, Roth JR, Andersson DI. 2012. Real-time evolution of new genes by innovation, amplification, and divergence. *Science*. 338(6105), 384-387.

Ohno S. 2013. *Evolution by gene duplication*. Springer Science Business Media.

Otto SP, Whitton J. 2000. Polyploid incidence and evolution. *Annual Review of Genetics*. 34(1), 401-437.

Pais J, Cunha PP, Pereira D, Legoinha P, Dias R, Moura D, *et al.* 2012. The Paleogene and Neogene of Western Iberia (Portugal): A Cenozoic Record in the European Atlantic Domain. In *The Paleogene and Neogene of Western Iberia (Portugal)* (pp. 1-138). Springer Berlin Heidelberg.

Pala I, Coelho MM. 2005. Contrasting views over a hybrid complex: between speciation and evolutionary “dead-end”. *Gene*. 347(2), 283-294.

Perea S, Böhme M, Zupančič P, Freyhof J, Šanda R, Özuluğ M, *et al.* 2010. Phylogenetic relationships and biogeographical patterns in Circum-Mediterranean subfamily Leuciscinae (Teleostei, Cyprinidae) inferred from both mitochondrial and nuclear data. *BMC Evolutionary Biology*. 10(1), 265.

Perea S, Cobo-Simon M, Doadrio I. 2016. Cenozoic tectonic and climatic events in southern Iberian Peninsula: implications for the evolutionary history of freshwater fish of the genus Squalius (Actinopterygii, Cyprinidae). *Molecular phylogenetics and evolution*. 97, 155-169.

Perea S, Garzón P, González JL, Almada VC, Pereira AM, Doadrio I. 2011. New distribution data on Spanish autochthonous species of freshwater fish. *Graellsia*. 91-102.

Pérez-Bote JL, Pula HJ. 2002. Usos del agua y problemática ambiental en la cuenca media del Guadiana. *Revista de Estudios Extremeños*. 58(2), 722-729.

Phillips SJ, Anderson RP, Schapire RE. 2006. Maximum entropy modeling of species geographic distributions. *Ecological modeling*. 190(3), 231-259.

Phillips SJ, Dudík M. 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography*. 31(2), 161-175.

Prieto I, Armas C, Pugnaire FI. 2012. Water release through plant roots: new insights into its consequences at the plant and ecosystem level. *New Phytologist*. 193(4), 830-841.

Qian W, Zhang J. 2014. Genomic evidence for adaptation by gene duplication. *Genome research*. 24(8), 1356-1362.

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Ribera I, Vogler AP. 2004. Speciation of Iberian diving beetles in Pleistocene refugia (Coleoptera, Dytiscidae). *Molecular Ecology*. 13(1), 179-193.

Rissler LJ, Apodaca JJ. 2007. Adding more ecology into species delimitation: ecological niche models and phylogeography help define cryptic species in the black salamander (*Aneides flavipunctatus*). *Systematic Biology*. 56(6), 924-942.

Robalo JI, Santos CS, Levy A, Almada VC. 2006. Molecular insights on the taxonomic position of the paternal ancestor of the *Squalius alburnoides* hybridogenetic complex. *Molecular Phylogenetics and Evolution.* 39(1), 276-281.

Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics.* 19: 1572–1574.

Sambrook J, Fritsch EF, Maniatis T. 1989. *Molecular cloning: a laboratory manual* (No. Ed. 2). Cold spring harbor laboratory press.

Sanjur OI, Carmona JA, Doadrio I. 2003. Evolutionary and biogeographical patterns within Iberian populations of the genus *Squalius* inferred from molecular data. *Molecular Phylogenetics and Evolution*. 29(1), 20-30.

Silvestro D Michalak I. 2012. raxmlGUI: a graphical front-end for RAxML. *Organisms Diversity and Evolution.* 12: 335–337.

Sousa-Santos C, Collares-Pereira MJ, Almada V. 2007. Reading the history of a hybrid fish complex from its molecular record. *Molecular phylogenetics and evolution*. 45(3), 981-996.

Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics.* 22: 2688–2690.

Tavaré S. 1986. Some probabilistic and statistical problems in the analysis of DNA sequences. *Lectures on mathematics in the life sciences.* 17: 57–86.

Taylor KE, Stouffer RJ, Meehl GA. 2012. An overview of CMIP5 and the experiment design. *Bulletin of the american Meteorological Society*. 93, 485-498.

Tejero R, Gomez-Ortiz D, Heydt GG, Toledo FM, Martínez CMC, Rodriguez MDMS, Suarez JJQ. 2017. Electrical resistivity imaging of the shallow structures of an intraplate basin: The Guadiana Basin (SW Spain). *Journal of Applied Geophysics*. 139, 54-64.

Vinckenbosch N, Dupanloup I, Kaessmann H. 2006. Evolutionary fate of retroposed gene copies in the human genome. *Proceedings of the National Academy of Sciences of the United States of America*. 103(9), 3220-3225.

Wagner A. 2010. On the energy and material cost of gene Duplication. *Evolution after Gene Duplication*. 207-214.

Wittbrodt J, Meyer A, Schartl M. 1998. More genes in fish?. *BioEssays*. 20(6), 511-515.

Zamora J F. 1987. El río Guadiana. *Fisiografía, Geoquímica y Contaminación. Departamento de Publicaciones de la Excma. Diputación, Badajoz*.

Zardoya R, Doadrio I. 1998. Phylogenetic relationships of Iberian cyprinids: systematic and biogeographical implications. *Proceedings of the Royal Society of London B: Biological Sciences*. 265(1403), 1365-1372.

Zardoya R, Doadrio I. 1999. Molecular evidence on the evolutionary and biogeographical patterns of European cyprinids. *Journal of molecular evolution*. 49(2), 227-237.

Figures

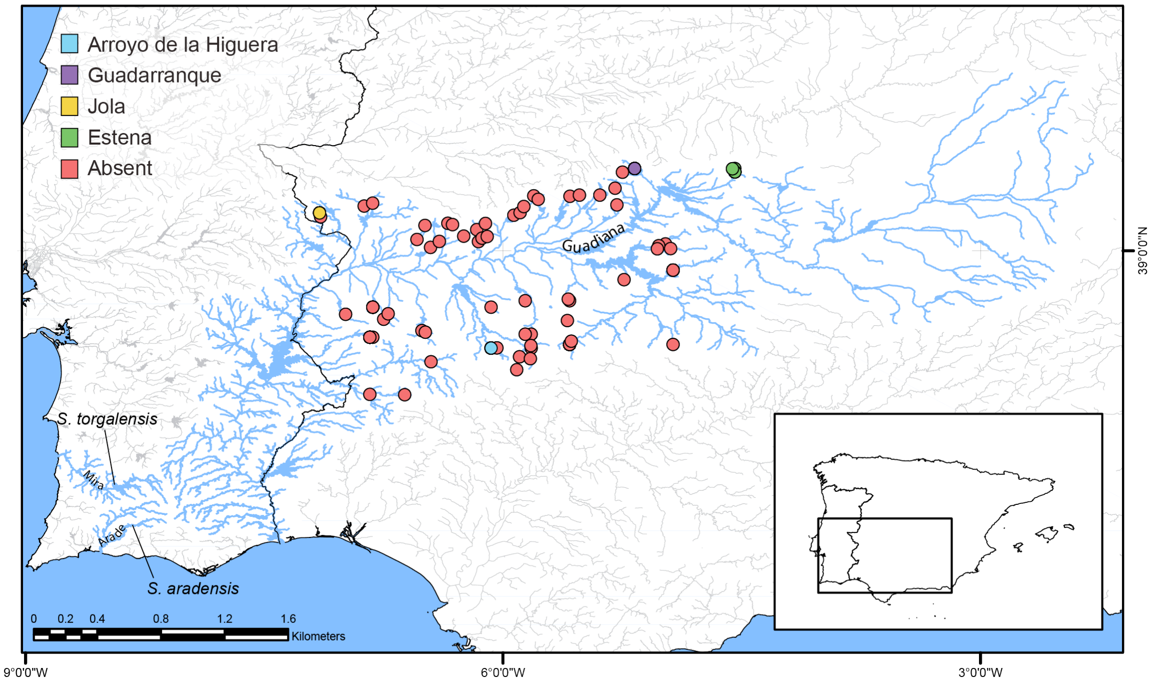


Figure 1. Sampling locations and current distribution of the hybrid complex *Squalius* sp in the Guadiana River.

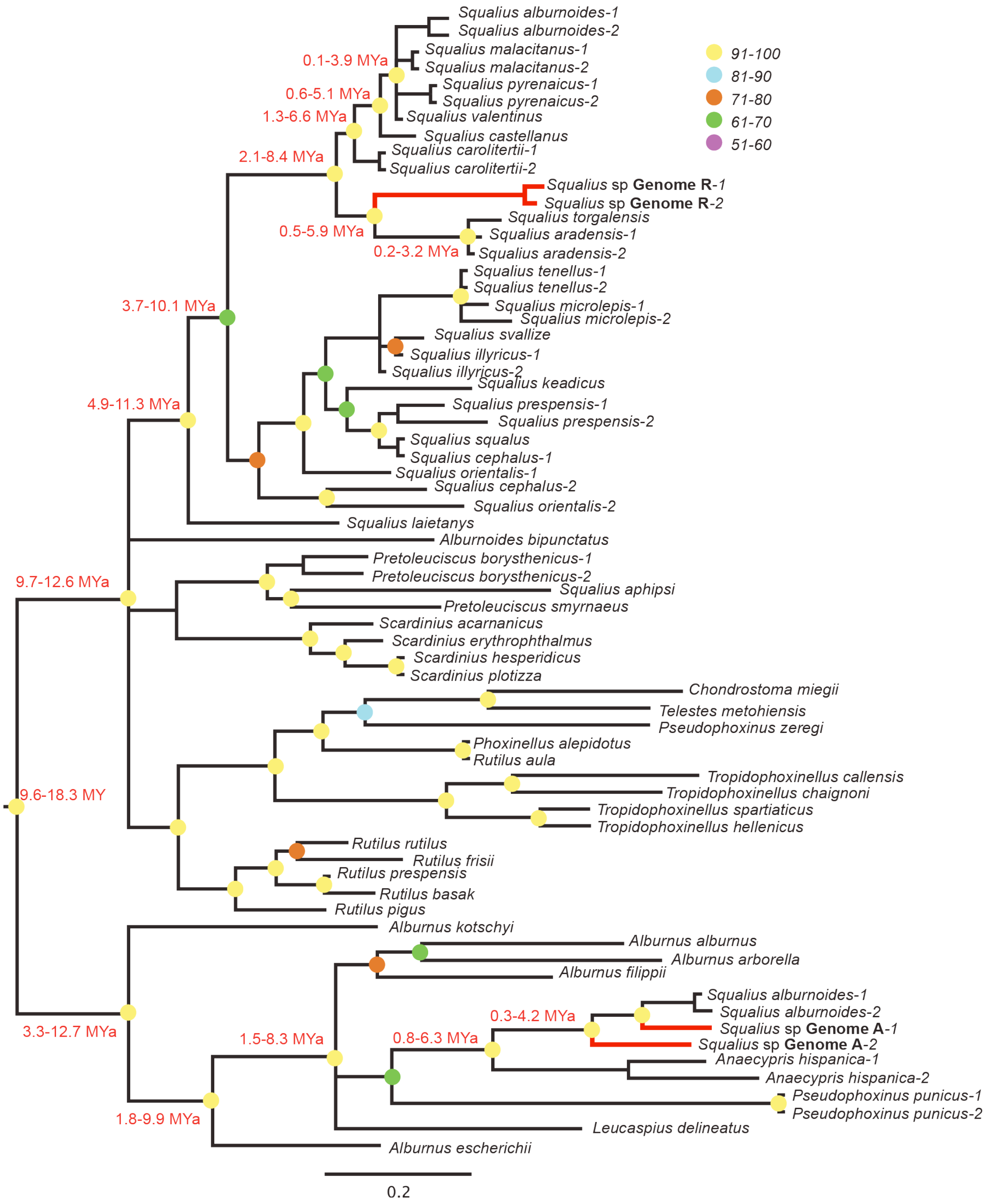


Figure 2. Bayesian inference based on the nuclear intron *S7* in part of the Leuciscinae subfamily. The circles in the nodes represent the posterior probability for Bayesian analysis and the bootstrap value for Maximum Likelihood analysis. Red branches indicate the position of the two different genomes (bold letters) of *Squalius* sp. The numbers in red are the divergence times estimates in millions years ago MYa.

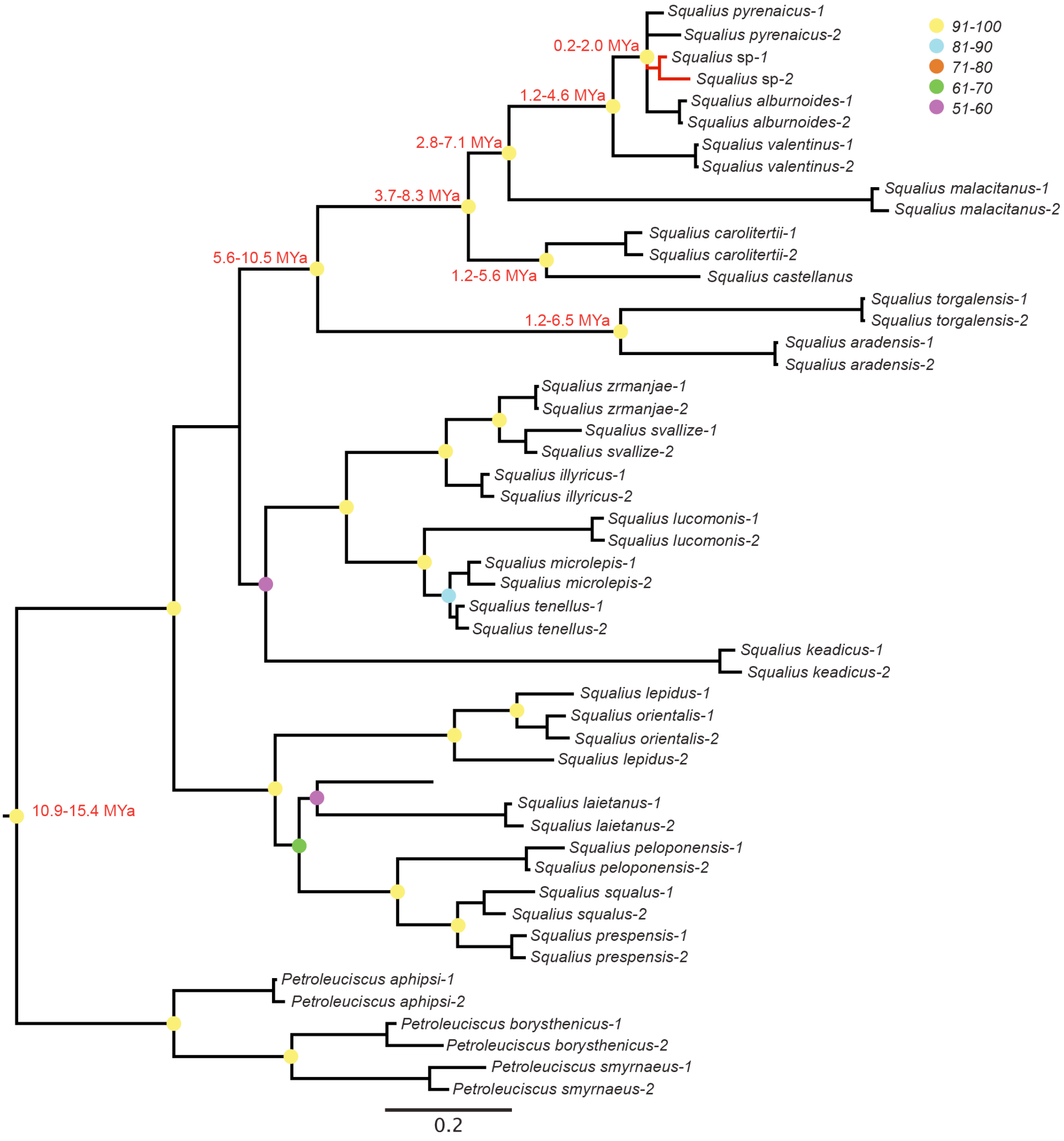


Figure 3. Bayesian inference based on the mitochondrial gene *cytb* for the genus *Squalius*. The number in the nodes represents the posterior probability for Bayesian analysis (left) and the bootstrap value for Maximum Likelihood analysis. The numbers in red are the divergence times estimates in millions years ago MYa.

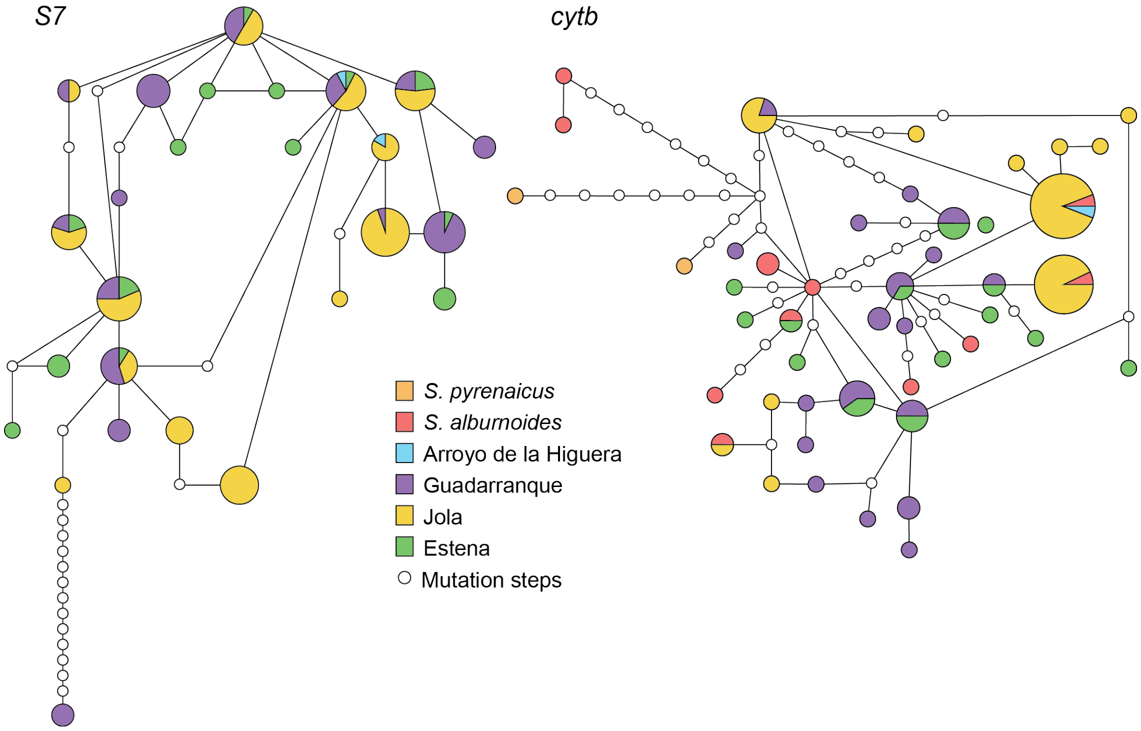
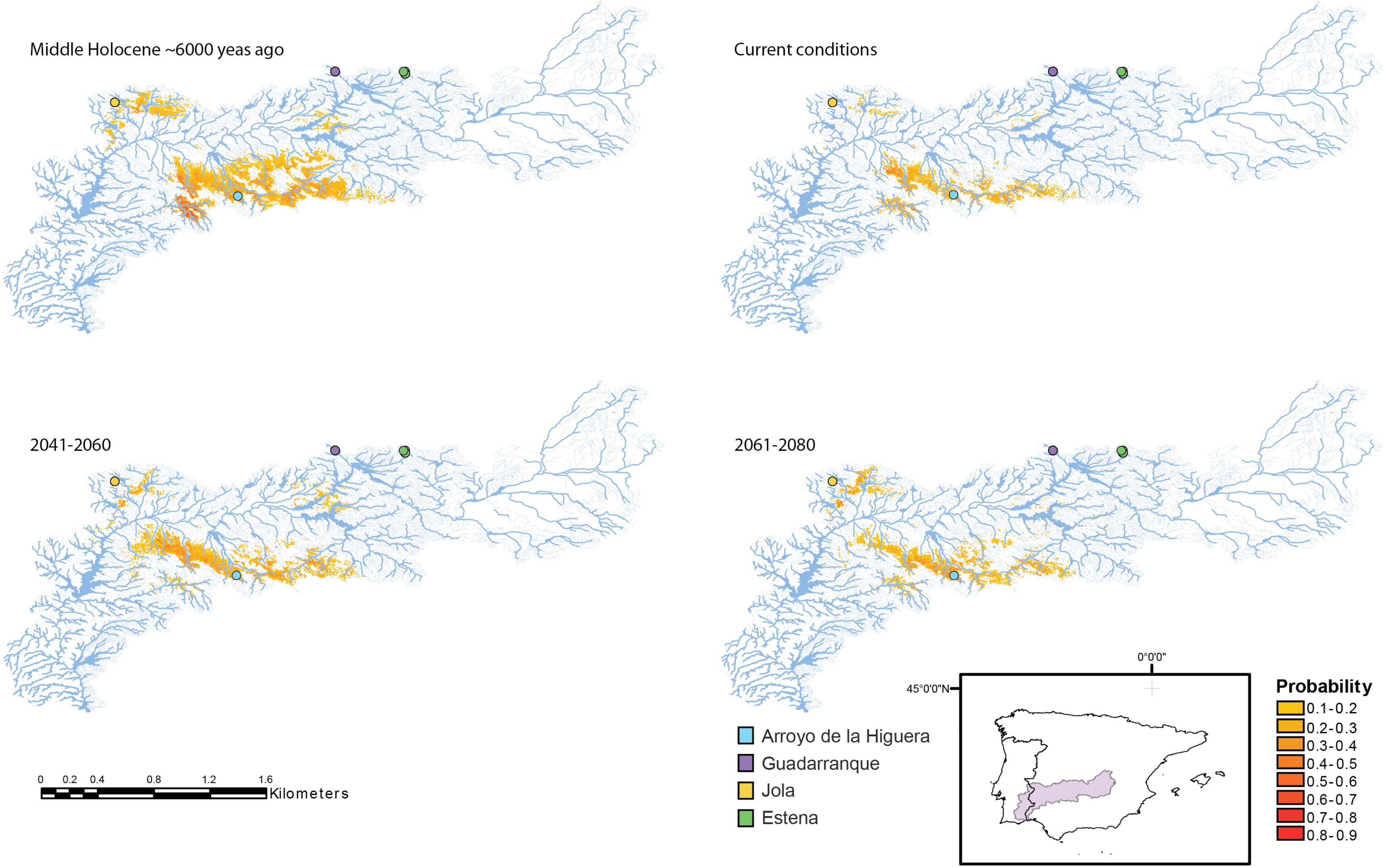


Figure 4. Haplotype network for the nuclear intron *S7* (left) and the mitochondrial gene *cytb* (right) for the hybrid complex *Squalius* sp.


Middle Holocene ~6000 yeas ago

Current conditions

2041-2060

2061-2080

Figure 5. Prediction of suitable habitat for *Squalius* sp in the Guadiana River in the past, the present and for the future conditions. The points represent the actual known populations. The values of probability of suitable ecological niche were determined in an interval confidence of 95%.

Tables

Table 1. Genetic variation in the comples *Squalius* sp in the Guadiana River for *cytb*|*S7*

Population	n	H	*h*	*π*	hInDel
Estena	17|20	14|9	0.978|0.889	0.004|0.003	2
Guadarranque	27|28	20|9	0.977|0.838	0.002|0.004	4
Arroyo de la Higuera	2|2	1|2	--	--	2
Jola	40|41	10|7	0.746|0.835	0.001|0.002	3

n = sample size, H = haplotype number by single mutations, *h* = haplotype diversity, *π* = nucleotide diversity, hInDel= haplotypes by insertions or deletions for *S7*.

Table 2. Absolute genetic distances (%)between species of the hybrid complex in the Guadiana River and the sister groups with *cytb* (above de diagonal) and *S7* (under the diagonal).

Species	***Squalius* sp**	***S. aradensis***	***S. torgalensis***	***S. alburnoides***	***S. pyrenaicus***	**Genome A of *Squalius* sp**	**Genome A of *S. alburnoides***
***Squalius* sp**		10.1	11.2	0.4	0.8	--	--
S. aradensis	2.8		6.1	10.1	9.9	--	--
S. torgalensis	2.8	0.3		11.2	11.0	--	--
S. alburnoides	3.2	2.2	2.2		0.8	--	--
S. pyrenaicus	3.2	2.2	2.2	0.8		--	--
Genome A of *Squalius* sp	9.0	8.3	8.3	7.1	7.0		--
Genome A of *S. alburnoides*	9.5	8.8	8.8	7.6	7.5	1.3	

Table 3. Analysis of molecular variance AMOVA based on nuclear *S7* intron for genome R in populations of *Squalius sp*.

Genetic configuration	**Φ_{ST}**	**Variation between populations**	**Variation within populations**
***S7* One genetic pool**	0.050*	5.0%	95.0%
***cytb* One genetic pool**	0.312*	31.2%	68.8%

* *p* value < 0.05

Table 4. Fixation index Φ_{ST} based on *cytb* (above de diagonal) and *S7* (under the diagonal) between populations of *Squalius* sp.

Population	**Estena**	**Guadarranque**	**Arroyo de La Higuera**	**Jola**
Estena		0.007	0.385	0.410*
Guadarranque	0.005		0.393	0.382*
Arroyo de La Higuera	0.099	0.029		0.498
Jola	0.055*	0.063*	0.083	

* *p* values < 0.05 after Bonferroni correction

Capítulo 4: Evolution of viviparity in Cyprinodontiformes: Mitogenomes or a few genes for resolving ancient phylogenies?

DIUSHI K. CORONA-SANTIAGO[1], FRANCISCO J. GARCÍA-DE LEÓN[2], OMAR DOMÍNGUEZ-DOMÍNGUEZ[3] and IGNACIO DOADRIO[1]

[1]Departamento de Biodiversidad y Biología Evolutiva, Museo Nacional de Ciencias Naturales, Madrid, España
[2]Laboratorio de Genética para la Conservación, Centro de Investigaciones Biológicas del Noroeste, S.C. La Paz, México.
[3]Laboratorio de Biología Acuática, Facultad de Biología, Universidad Michoacana de San Nicolás de Hidalgo, Morelia, México.

For Submission in: Systematic Biology

Abstract

The evolution of viviparity implies several evolutionary processes and plays a variety of fetal-maternal relationships, which we can assume that the mitogenome has been involved. The availability of mitogenomes for evolutionary studies is growing, however the phylogenomics approaches are not without limitations. Given the putative role of the mitogenome on viviparity evolution and the importance of the identification and selection of new molecular markers, in this study we analyze if changes in the mitochondrial genome of Cyprinodontiformes have been occur and if are related with the origin of viviparity. Also, we explore and compare the phylogenetic resolution in different regions of mitogenome. We found that the viviparity and ovoviviparity have as common ancestral to oviparous, arising independently. Our analyses reveal that during the evolution of Cyprinodontiformes differential number of genes under significant positive selection (8 to 11 in viviparous groups and 1 to 3 in oviparous) occurred, suggesting that this pattern could be linked to the origin of the viviparous and ovoviviparous. The phylogenetics hypotheses shows that the combination of three gene (*cox1*, *nd4l* and *cytb*) allow a reconstruction of phylogenetic trees with high support, resolution and same topology that if we use the complete mitogenome in the Cyprinodontiformes order, relevant to optimize economical and bioinformatics/analytical resources, and we determine that the saturation combined with other factors (alignment, informative sites or length) must be considered in the identification and selection of molecular markers for phylogenetics.

Keywords: viviparity, Cyprinodontiformes, mitogenome, positive selection,
phylogenetic tree.

Introduction

The origin of viviparity and ovoviviparity in fishes dates from Carboniferous (Lund, 1980). In the ostheichthyans have occurred 29 independently evolutionary origins, which one of the most recent viviparity origin was dated for the Middle Miocene to the Goodeidae family in the Cyprinodontiformes order (Blackburn, 2005). Ovoviviparity is assumed to be the ancestral state from which live-bearing has evolved. In fishes, initial steps in the evolution of live-bearing from ovoviviparity are believed to have involved a shift from external to internal fertilization, followed by an increased retention of eggs and developing embryos within the female reproductive system (coinciding with morphological and physiological adaptations of the ovary or oviduct) and a reduction in the thickness of the egg envelope to facilitate female-embryo interactions (Wourms *et al.,* 1981; Wourms and Lombardi, 1992; Pollux *et al.,* 2009).

Viviparity might be a non-reversible condition, and does not define a complete teleost fish group including the large order Cyprinodontiformes, which viviparity (including ovoviviparity) appears repeatedly where the phylogenetic relationships among these are not clearly established based on molecular (Meyers and Lydeard, 1993; Webb, 1998; Setiamarga *et al.,* 2008; Pohl *et al.,* 2015) or morphological data (Parenti, 1981; Costa, 1998; 2011; Hertwig, 2008).

The acquisition of viviparity brings into play variety of fetal-maternal relationships including: 1) developmental, 2) morphological, 3) trophic, 4) osmoregulatory, 6) respiratory, 6) endocrinological, and 7) inmmunological (Wourms; 1981; Kawaguchi *et al.,* 2014). Hence, the evolution of viviparity implies several and complex evolutionary processes, which we can assume that the mitochondrial genome has been involved. Mitochondrial are responsible for 95% of the eukaryotic cell’s energy through oxidative phosphorylation of ADP (adenosine diphosphate to form ATP, adenosine triphosphate) (Moreno-Sánchez, Hogue and Hansford, 1990; Bai, Shakeley and Attardi, 2000). Further, mitochondrial genes could be influenced by environmental conditions, historical events, and life history process affecting the fitness, with consequence in metabolic processes and may be taxon-specific (Wourms, 1981; Meiklejohn, Montooth and Rand, 2007; Consuegra *et al.,* 2015). The genetic draft could have a great impact in the mitochondrial genome evolution (Bazin *et al.,* 2006a).

The availability of mitochondrial genome for evolutionary approaches in several animal groups including fish (oviparous, ovoviviparous and viviparous) is growing as a result of recent technological advances (McCormacka *et al.,* 2013; Pons *et al.,* 2014). Thus giving the opportunity to compare the mitogenome between reproduction modes in fishes, and providing information about the evolution of viviparity in a molecular level. It is evident that the advantages of the inclusion of multiple genes approaches, versus those based on single genes, is addressed in favor to construct a true common phylogenetic signal, showing the evolution of the biota and not or a particular DNA region (Comas *et al.,* 2007). However, the phylogenomics approaches are not without limitations (Delsuc, Brinkmann and Philippe, 2005; Feschotte and Pritham, 2007; Philippe *et al.,* 2011; Chan and Ragan, 2013). One of the most important is the cost involved and the desire to obtain comparable data among studies, accordingly most analyses have focused on sequencing a relatively small portion of the genome, resulting in some cases trees with low phylogenetic resolution of contradictory topologies due the low number and resolution of mitochondrial markers (Willerslev *et al.,* 2009; Pacheco *et al.,* 2011). Regions responsible for the higher resolution obtained from complete mitogenomes can be identified for particular taxonomic groups by analyzing results from all mitochondrial markers, providing information of how to select and which genes could be used for determinate groups, and insights into incongruence in results among studies based on single genes (Zardoya and Meyer, 1996; Duchêne *et al.,* 2011).

Hence, given the putative role of the mitochondrial genome on viviparity evolution and the importance of the identification and selection of new molecular markers for phylogenetic studies, in this study we analyze if important changes in the mitochondrial genome of Cyprinodontiformes have been occur. The aim of this study is to describe the viviparity evolution using analyses to infer in the phylogenetic relationships between Cyprinodontiformes and the common ancestor for viviparism, detecting selective pressures related to viviparity origin, and identifying changes in gene arrangement, duplications and composition in the order. In addition, we performed multiple phylogenetic inferences for the order Cyprinodontiformes using mitogenome, single mtDNA locus and combining loci, whit the aim of explore and compare the phylogenetic resolution in different regions of the mitochondrial genome, and determine which gene or combination of genes are suitable for phylogenetics purpose in the order.

Materials and Methods

Mitogenome data

We include the available mitogenomes of ovoviviparous, oviparous and viviparous species of the Cyprinodontiformes families (Table 1): Aplocheilidae (*Aplocheilus panchax*) (Setiamarga *et al.,* 2008), Nothobranchiidae (*Nothobranchius furzeri*) (Reichwald *et al.,* 2009) and Rivulidae (*Kryptolebias marmoratus*) (Lee *et al.,* 2000) of Aplocheiloidei suborder; Goodeidae (*Xenotoca eiseni*; Setiamarga *et al.,* 2008, and *Xenotoca variata* (Corona-Santiago *et al.,* 2017); Poeciliidae (*Gambusia affinis*, *Poecilia reticulata*, *Xiphophorus hellerii* and *X. maculatus*) (Miya *et al.,* 2001; 2003; Kong *et al.,* 2013; Setiamarga *et al.,* 2008; Bai *et al.,* 2009), Cyprinodontidae (*Cyprinodon rubrofluviatilis* and *Jordanella floridae*) (Crowl *et al.,* 2007; Setiamarga *et al.,* 2008) and Fundulidae (*Funfulus diaphanous, F. grandis, F. heteroclitus, F. olivaceus*) (Setiamarga *et al.,* 2008; Whitehead, 2009) of the Cyprinodontoidei suborder.

Cyprinodontiforms mitogenomes alignment, composition and variation

Alignments were performed for each mitochondrial gene which in all the cases were aligned using MAFFT v7.222 (Katoh *et al.,* 2002). The DNA sequences of rRNA genes and tRNA were aligned considering secondary structure of RNA, and changed manually to correct obvious misalignments.

Composition skew analysis was carried out to describe the base composition in each strand of the mitogenomes, which measures the relative number of As to Ts (AT skew = [A-T]/[A+T]), Gs to Cs (GC skew = [G-C][G+C]) (Perna and Kocher, 1995), and gene duplication were included. In addition variation in the encoding genes and *d-loop* were analyzed estimating the complete variable sites, rate of non-synonymous (dN) and synonymous substitutions (dS), indels events, nucleotide diversity and proportion of conservative sites. The ratio (ω) of the dN/dS was estimated using the PAL2NAL v14 (Suyama *et al.,* 2006) removing gaps and inframe stop codons. Significant saturation in codon position was estimated with Dambe v6 (Xia, 2013).

For estimation of the genetic variation among families and between reproduction modes, the nucleotide divergences (Dxy) were obtained using Sliding window analysis. The analyses were conducted on the full alignment combinations performing a window length of 300bp with a step size of 10bp in each combination.

Phylogenetic inference and gene tree variation

Protein-coding gene sequences were aligned using MAFFT module to obtain a protein alignments; and excluding gaps, poorly aligned positions and divergent regions (including saturated positions) using GBlocks v0.91b module (Castresana, 2000; Talavera and Castresana, 2007) in TranlatorX (Abascal *et al.,* 2010), which after allow to obtain a DNA alignment.

Using a master mitogenomic data combining the cleaned independent DNA alignment, a concatenated phylogeny reconstruction based on Bayesian inference was inferred. For this, the best-fitting partitions (four partitions p1: rRNA genes; p2: tRNA genes; p3: encoding genes; p4: *d-loop*) and model of nucleotide substitution (General Time Reversible [GTR] + Invariable sites [I] + Gamma [G]) were estimated using the Akaike information criterion as implemented in PartitionFinder v1.1.0 (Lanfear *et al.,* 2012). Although the same model was performed in each partition, the shape distribution, the proportion of invariable sites and other parameters were applied as independent variable parameters.

In order to determine which and how genes are necessary to obtain the most similar phylogenetic trees to that based on the complete mitochondrial genome, independent phylogenetic reconstruction analyses were implemented for each encoding gene based on Bayesian inference too. Also, a phylogenetic tree was implemented based on only *cox1* barcode region (628bp). The best-fitting model for each encoding gene was estimated using jModeltest v2.1.6 (Darriba *et al.,* 2012) (Table 1S of the Supplementary information).

The variation between genes trees with mitogenome tree was obtained and plotted using multi-dimensional scaling (MDS) in the module TreeSetViz v3.0 (Amenta and Klingner, 2002; Huang and Li, 2013) of Mesquite system for phylogenetic computing v3.04 (Maddison and Madisson 2017), estimating the unweighted Robinson-Foulds (RF: number of splits in each tree but no in both topologies, Robinson and Foulds, 1981), the weighted RF distances (RF_W: RF taking into account support values and penalizing, to a lesser extent, incongruent subtrees, Robinson and Foulds, 1979) and the Patristic correlation (PC: correlation coefficient of the sum of branch lengths between all terminal nodes in two different trees). For MDS plot, 90% of the trees in each phylogenetical analysis were including for the analysis for estimation of variation between different gene trees/concatenated trees and the mitogenome tree.

In addition, four phylogenetic trees were obtained from concatenated combinations of genes taking as criterion the variation between gene trees and mitogenome tree: the (1) three and (2) the five genes with the highest PC values, (3) the three genes with the lowest RF_W values, and the (4) two genes with the lowest RF and PC values. The criterion to the gene combination selection was taking into account; first if is possible to minimize the number of genes to obtain the same topology based on mitogenomes; and second, taking the gene trees with high posterior probabilities (low RF_W and RF with $RF_W \approx$ RF). PartitionFinder was used to infer if a model must be applied for each genes or one model for all or part of the data combined as required.

In all the phylogenetic inferences explained above (including those based on complete mitogenomes, encoding genes or concatenating some genes) were conducted in MrBayes v3.2.6 (Huelsenbeck and Ronquist, 2001), performing two runs with four MCMC process, each of those with 5 millions of generations sampling each 100 generations. Chain convergence was evaluated using Tracer v1.6 (Rambaut *et al.,* 2014) and 10% of the trees were discarded as burn-in to consensus tree reconstruction (σ <0.001).

Ancestral states for reproductive mode and mitochondrial adaptive evolution

In order to infer in the evolutionary history of traits about the different reproductive modes including in this work, an ancestral states reconstruction was performed using Beast v1.8.2 (Drummond *et al.,* 2012) based in complete mitogenomes cleaned data set. This analysis was performed using 100 millions of generations sampling each 1000 generations. Chain convergence was evaluated using Tracer v1.6 and based on the effective sample size (> 1000), discarding 10% of the total of generations to Maximum Clade Credibility tree construction.

The significance of ω variation between groups with different reproductive modes in a evolutionary context was estimated performing different McDonald-Kreitman test (MKT) (McDonald and Kreitman, 1991) in DnaSP v5, which allow to detect significance deviation from neutrality, indicative of historical selection pressure resulting in changes of functionality or expression in mitochondrial genes associated to the origin of viviparity, ovoviviparity and oviparity in Cyprinodontiformes order. Of there are changes by positive selection by adaptive mutations associated to modes of reproduction; we can assume patterns as 1) variation in the number of genes under selection between viviparous and ovoviviparous, or by 2) variation in the kind of gene or mitochondrial protein complex under selective pressure between modes of reproductions. The MKT was performed using MKT server (Egea *et al.,* 2008) selecting the synonymous substitutions as neutral and applying the Jukes&Cantor correction (Jukes and Cantor, 1969). The significance of the neutrality deviation was evaluated with the p values of X^2 test.

Results

Cyprinodontiformes mitogenome composition

The mitogenomes size in Cyprinodontiformes range between 16 177bp to 19 527bp, being *Nothobranchius furzeri* (Nothobranchiidae family) the species with the largest mitochondrial genome, while *Jordanella floridae* whit the smallest (Table 1), both species are oviparous. Gene duplication was observed for the *d-loop* region in oviparous (*Nothobranchius furzeri* and *Kryptolebias marmoratus*), while only gene duplication in viviparous was observed for tRNA-Met (*Xenotoca eiseni*). Asymmetrical strand composition was observed in three families (Nothobranchiidae, Aplocheilidae and Poeciliidae) showing positive AT skew and negative GC skew, while in other oviparous (Cyprinodontidae, Rivulidae, Goodeidae and Fundulidae) and viviparous (Goodeidae) was found richness of T over A (negative AT skew) and C over G (negative GC skew) (Table 1).

Nucleotide divergences and encoding genes variation

Comparing among groups with different reproductive modes show in general the same nucleotide divergence (Fig. 1). The highest divergences were observed between oviparous and ovoviviparous in most of the regions. However the nucleotide divergence between all groups in general are the same, except in the *d-loop* region for comparison between oviparous with ovoviviparous. As we expected, the divergences between species of the same family are lower than between families (we show the viviparous case Fig.1). The mitogenome regions showing the greatest divergences and diversity between families and reproductive modes (independently of its length), was *d-loop,* followed by *atpase8* and *nd6* genes (Fig. 1, Table 2). In contrast, the lowest values of genetic divergence were observed in rRNAs sequences.

Encoding genes diversity shows that the region with most indels events in the order is the complete *cox1* gene (12 events) (Table 2). The non-coding region *d-loop* shows 94 events. The ω ratio in all the genes for each reproductive mode analyzed was lower than 1 (Table 2, Fig. 2). All the genes show significant saturation in the third codon position except the genes *cox1* and *nd4l* with null saturation, while also the *atpase8* shows saturation in the first position.

Phylogenetics relationship and topology variation

Bayesian inference based on the complete mitogenome shows well-supported relationships between all clades and internal nodes (posterior probabilities > 0.99 σ = 0.0003) (Fig. 3). All families included were resolved as monophyletic, with the oviparous Aplocheilidae, Rivulidae and Nothobranchiidae as basal members. The ovoviviparous family Poeciliidae is the sister group of Cyprinodontidae, Fundulidae and Goodeidae families. The viviparous subfamily Goodeinae is the sister group of the oviparous family Fundulidae, with respect to oviparous Cyprindontidae (Fig. 3).

The most similar gene trees using MDS were the based on the single encoding genes *cox1*, *nd4l*, *cytb*, *atpase6* and *nd5* (Fig. 4). The gene trees based on *cox1* and *nd4l* showed the topologies with the best relation RF ≈ RF_W, with the lowest values of splits (RF = 1), less number of branch with high support were penalized (RF_W = 2.7160-2.9025) and the topology with the high similarity (PC = 0.9768-0.9786) with respect to the complete mitogenomes.

The combination of genes that allow to obtain the most similar phylogenetic tree with respect to the tree based on the mitochondrial genome were *cox1*, *nd4l* and *cytb* (PC = 1; RF_W = 0.9968; RF = 0), followed by the tree reconstructed with the five genes with the highest PC values (*cox1*, *nd4l*, *cytb*, *atpase6* and *nd5*). Also, in the other two concatenated trees lower patristic correlation, RF_W and RF distances than the trees based in single genes were observed: with the minimum genes combined (*cox1* and *nd4l*) and the trees obtained with 3 genes (*cox1*, *nd4l* and *nd5*) selected in accordance with the lowest values of RF_W.

The gene trees with the highest RF_W and RF distances in the PC, were those based on *atpase8* and *cox2* (Fig. 4), while those based on the barcode region (*cox1*) shows topologies with high distance respect to the trees based on the complete mitogenome and incongruous with the gene trees based on the complete *cox1*. As expected, the gene trees constructed based on the most variable region *d-loop* has high RF_W and RF distances respect to those based on the complete mitogenome.

Ancestral states and MKT estimation

The MKT analysis shows significant deviation form neutrality (NI > 1) in encoding mitochondrial genes during the evolution of Cyprinodontiformes (Fig. 3, Table 2S). High proportion of adaptive substitutions (α) was observed in most of the cases of positive selection. The largest number of genes under selection was observed when the viviparous and ovoviviparous have arisen, ranging from 9 in ovoviviparous (8 under positive selection and 1 under negative selection) to 11 in viviparous (all under positive selection) genes. Meanwhile, low (between 2 to 5) number of mitochondrial genes was observed under significant deviation from neutrality when two oviparous lineages have been emerged.

Discussion

In the comparison of the mitochondrial genomes for species of oviparous ovoviviparous and viviparous Cyprinidontiforms, we found gene duplications in the order and significant changes in bases composition. These changes in encoding genes have significant deviation of neutrality during the order evolution and that currently are under purify selection. We found that the viviparity and ovoviviparity have as common ancestral to oviparous, arising independently. The MKT analyses reveal that during the evolution of Cyprinodontiformes significant structural changes in most of the encoding genes are associated to adaptive substitutions, resulting in alterations of the gene functionalization or expression via positive selective pressures. Observing the differential number of genes under positive selection between reproductive modes, we suggest that these changes are linked to the origin of the viviparous and ovoviviparous (higher number of genes under positive selection in viviparous and ovoviviparous than in oviparous). According to the literature we propose that selection by internal (genomic co-evolution) and external (environment changes) has played an important role in the evolution of mitogenome and probably of the viviparity (live-bearing and oviviparous), since has been reported that changes in mitochondrial genome has a critical role in fitness, fertility, adaptability and speciation.

In this work we present a new well-supported phylogenetic hypothesis about Cyprinodontiformes order. We found disagreements between the phylogenetic relationships observed with the data included in this work with that of the molecular phylogeny proposed for the order in previous works. However we suggest that this hypothesis should be taken with caution due nuclear data and complete taxa sampling should be included to infer a true phylogenetic hypothesis. We explore the consequence in the support and topology variation by using the complete mitogenome, single genes and combining genes, founding that the combination of three gene (*cox1*, *nd4l* and *cytb*) allow a reconstruction of phylogenetic trees with high support, resolution and same topology that if we use the complete mitogenome in the Cyprinodontiformes order, relevant to optimize economical and bioinformatics/analytical resources.

We consider that our result show important information about the evolution of viviparity in fishes and for the identification and selection of genes for supraspecific phylogenetics studies with the possibility to apply in population genetics in Cyprinodontiformes y other vertebrates.

Evolution of viviparity in Cyprinodontiformes

The phylogenetic approach is congruent with those proposed using morphological data for the families Aplocheilidae, Rivulidae, Nothobranchiidae (Costa, 1998; Hertwig, 2008), and partially congruent with the phylogenetic relationships observer using some molecular markers for these families (Meyers and Lydeard, 1993; Pohl *et al.,* 2015). The relationships between families Poeciliidae, Fundulidae, Goodeidae and Cyprinodontidae were well supported and less unambiguously resolved as in the most recent morphological and molecular hypothesis suggested (Hertwig, 2008; Pohl *et al.,* 2015). Those studies show to Cyprinodontidae as the basal family of Poeciliidae, Fundulidae and Goodeidae, in contrast with the relationships observed in this work, Poeciliidae occupy the basal position. The family Fundulidae is the sister group of Goodeidae as is observed in previous works taking into account that in this work is not included the family Profundulidae.

The ancestral state reconstruction shows that the viviparity and ovoviviparity have as common ancestral to oviparous, and the ovoviviparity is not an intermediate state for viviparity (Fig. 3). This is in agreement with the hypothesis about that viviparity arising independently in the Cyprinidontiformes order (Meyer and Lydeard, 1993; Pollux *et al.,* 2009).

Our results show that the most of coding regions have different patterns of substitution, suggesting that different genes have experienced several positive selective pressures during the origin of viviparous groups. This pattern is characterized by higher number of genes under positive selection in live-bearing and oviparous groups than in oviparous, suggesting that the differential number of genes under positive selection between oviparous, ovoviviparous an viviparous could be associated with origin of these modes of reproduction. We observed between 8 to 11 genes where under positive selection in the ovoviviparous (Poeciilidae) and live-bearing (Goodeidae) families respectively, while between 1 to 3 genes in oviparous origin (Fig. 3 and Table 2S of the Supplementary information), and being, for mitochondrial genome, cox3 and atpase8 and nd5 the genes that differ between ovoviviparous and viviparous. Because is difficult to link this pattern to the origin and evolution of viviparity, we relate these adaptive changes with their effect in processes linked to viviparity observed into mammals, reptiles, and some model species: metabolism fitness, embryo/development, fecundity, mito-nuclear interactions and mitochondrial down/upregulation.

The subunits genes *nd1-nd6* (Complex I), that has been documented that modifications and overexpression could result in increasing metabolism fitness (some mutation in *nd5*) (Ealson *et al.,* 2008; Lane, 2011; Talbert *et al.,* 2015), was under positive selection except the *nd4l* gene (under neutral selection) and the *nd5* (in negative selection). The complex IV (*cox* subunits 1-3), that is known his upregulation during the embryo development, active nutrition transport and placental modeling by the energy demands (Ballard *et al.,* 2007; Dowling *et al.,* 2008); is not more conserved than complex I, and V (*atpase* subunits) as usually occur in vertebrates (Pesole *et al.,* 1999; Meiklejohn, Montooth and Rand, 2007). Complex IV shows significant positive selection in *cox1* (in ovoviviparous and viviparous) and *cox3* (in viviparous), and no significant selection of *cox2*. In the complex V, where the ATP production ratio is potentially affected and upregulated by nuclear interaction during vertebrate embryogenesis (Bogue *et al.,* 2000; Zhang *et al.,* 2004) being a critical factor as in cell energy production as in embryo viability (Wallace and Fan, 2010; Lane *et al.,* 2011). We identify in complex V positive selection to the subunit atpase8 (in viviparous and ovoviviparous) and atpase6 (in viviparous). Other additional process identify that could affect processes involves in viviparity are the differential selective pressure associated to asymmetrical strand composition in all the families (AT Skew and GC Skew) analyzed in this work. The asymmetry could result from selection and/or transcription-coupled mutation (Francino and Ochman, 1997; Frank and Lobry, 1999; Tillier and Collins, 2000) and intensifying the emergence of new mutations (Mackiewicz *et al.,* 1999; Morton and Morton; 2007). However, is unknown the relation and the effect of the asymmetrical composition in processes related with viviparity.

Phylogenetic implications and molecular marker resolution

The selection of genes or regions for phylogenetic inferences could be significantly determined by the incorrect identification of orthologs, erroneous alignments, or underestimation of saturation events occurring in our data (Philippe *et al.,* 2011). In this case, we determine that saturation play an important role to select gene for phylogenetic studies. We identify null saturation in genes (*cox1* and *nd4l*) that obtain phylogenetic trees with the lowest PC, RF and RF_W with respect to trees based on the complete mitogenome (Table 2). In contrast, the third codon position appears saturated in all genes, inclusive in the genes that worst supported and resolved topologies (*atpase8* and *cox2*) in comparison with the phylogenetic tree obtained with the complete mitogenomes. Although we can determine a clear pattern of substitution saturation, some genes as *cytb*, *atpase6* and *nd5*, genes with low PC, RF and RF_W values, show saturation in the third codon position. This result is indicative that the saturation and its effect in phylogenetic inferences depend of its intensity and with the combination of other factors hardly to separate as gene length and informative sites (Dunn *et al.,* 2008; Philippe *et al.,* 2009; Shierwater *et al.,* 2009), and could be associated with the time between cladogenetic events (Saitou and Nei; 1986; Philippe, Chenuil and Adoutte, 1994; Zardoya and Meyer, 1996). In addition, we observed that this effect is reflected in the RF_W values, for example in the genes *atpase6* (638bp) and *nd5* (1838), both with the saturated third codon position, generate trees with different splits penalized by its posterior probabilities (RF_W = 3.4679 and RF_W = 2.7703, respectively). In the same way occur in genes as cox2 (691bp) with the same third position saturated and generating trees the worsts posterior probabilities and resolved trees. The effect of the intensity of saturation is reflected also in the atpase6 where the first and third position are saturated in Cyprinodontiformes, suggesting that the most realistic model of sequence evolution is not well estimated when high number of multiple substitutions occurring at the same position (i.e., homoplasy; Philippe *et al.,* 2011). Also, we did not identify a pattern that relates gene length/variable sites with level of phylogeny resolution. This is important because the gene length or the hypervariable/conservative sites (Fig. 1), the dN/dS ratio, nucleotide divergence/diversity or the indels events are not characteristics that help to select between genes for phylogenetic studies, factors described also in previous works (Zardoya and Meyer, 1996). Phylogenetic inference based on the complete mitogenome can be obtained with genes or regions that are not the most variables. For example, *cox1* with a length of 1551bp (618 variables sites in the order Pi = 0.1927) has more resolution (in terms of support and resolved topology) for the Cyprinodontiformes phylogeny than *nd4* with a length of 1381bp

(805 variable sites Pi = 0.2777). Small regions have more resolution than large genes, but this is not a patter maintained in all genes of small lenght. For example, the gene *nd4l* with 297bp (184 variable sites in the order Pi = 0.2625) resolve the phylogeny of the order better than *cox2* (300 variable sites Pi = 0.1887), both with around 10% between species of the total variable sites. Other examples or incongruences and low supported and resolved phylogenetic tree reconstruction are those obtained with two regions widely used for phylogenetics (Patwardhan, Ray and Roy, 2014), the *d-loop* and a fragment of *cox1* corresponding to barcode region. Both regions results in phylogeny low support and resolved and well below those obtained by the five genes that resolve better the Cyprinodontiformes phylogeny. This suggests that *cox1* barcode could be useful to identify species of many animal groups but is not a region recommended to deep phylogenetic approaches in fish due to loss of informative sites or missestimation of substitution model (Meiklejohn *et al.,* 2014; i.e. of *cox1* barcode incongruence: Tobe, Kitchener and Linacre, 2010; Chen, Jiang and Qiao, 2012; Campos-Soto, Torres-Pérez and Solari, 2015; Cong *et al.,* 2016), while the hypervariability of *d-loop*, widely used for phylogenetic and phylogeographic studies (Taberlet, 1996; Rocha-Olivares *et al.,* 2005; Arif and Khan, 2009), do not contribute to resolving phylogeny probably due to alignment ambiguity and saturation sequences (i.e. for incongruence *d-loop*: Burridgen, 1999; Rueness *et al.,* 2003; Paterson and Velazco, 2008; Verma, Singh and Kumar, 2016).

When multiple genes are concatenated and analyzed, non-phylogenetic signal is caused by the inclusion of sequences that deviate from the true species phylogeny or by the inability of the methods to correctly handle multiple substitutions (Liu, *et al.,* 2008). However, the correct selection of the genes could contribute to the reconstruction of resolved and supported phylogeny but it could depend of the phylogenetic ‘depth’. This was observed when 5 genes were combined (*cox1*, *nd4l*, *cytb*, *atpase6* and *nd5*) obtaining a phylogeny lower supported than combining 3 genes (*cox1*, *nd4l* and *cytb*), being the last the best combination in to resolve the Cyprinodontiformes phylogeny in agreement respect to those based on the complete mitochondrial genome with null differences. This results contrasts partially the idea that *atpase6* and *nd4l* are poor genes for phylogenetic recovering (Zardoya and Meyer, 1996; Meiklejohn *et al.,* 2014), suggesting that the selection of genes for phylogenetics most be done carefully and taking into account several factors.

In other context, the neutrality does not determine congruence in topology and is not a factor that is necessary to exclude or select a region for phylogenetics (Roje, 2014), and the number of sites under positive selection does not predict congruence with other markers or the topology inferred with the concatenated dataset (Larmuseau *et al.,* 2010). This is also corroborated in this work which we observe that can be obtained trees well-supported and resolved independently if currently all genes are under negative purify selection (Fig. 2) or if during evolution were under positive or negative selection. Only the gene *nd4l* was maintained under neutrality in all the Cyprinodontiformes evolution but this pattern is not observed in *cox1* or *cytb*, genes that we obtain with high support and resolved phylogenetic trees for the order. The same pattern was observed in genes as *atps8* and *cox2*.

Between one to three mitochondrial genes are necessary to well-resolved and supported phylogeny and identical if using the complete mitogenome. Its selection depends of factors of saturation intensity, ambiguity alignment due to level of taxa divergence and the model of evolution for sequences estimated linked, but more studies should be applied to determine the measure in that factors as length, variable sites, sample size and selection is involved in the difficulty to resolve phylogenies (Meiklejohn *et al.,* 2014). We suggest that the molecular markers identified are sufficient to resolve large phylogenies, but our result must be compared using nuclear data to determine if this genes are sufficient to obtain a more realist or true phylogenetic hypothesis about the evolution of the order and if our results can be applied in other animal groups.

Acknowledgements

We thanks to Mario García París for the comments and suggestions for this work. We thanks to the Supercomputing Center of Galicia (CESGA) for the informatics resources facilities.

References

Abascal F, Zardoya R, Telford MJ. 2010. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Resources*. gkq291.

Amenta N, Klingner J. 2002. Case study: Visualizing sets of evolutionary trees. In *Information Visualization, 2002. INFOVIS 2002. IEEE Symposium on* (pp. 71-74). IEEE.

Arif IA. Khan HA. 2009. Molecular markers for biodiversity analysis of wildlife animals: a brief review. *Animal Biodiversity and Conservation*. 32(1), 9-17.

Bai JJ, Liu YF, Quan YC, He XP, Ye X. 2009. The complete mitochondrial genome of the RR-B strain of swordtail (*Xiphophorus helleri*) Short Communication. *Mitochondrial DNA. 20*(4), 72-77.

Bai Y, Shakeley RM, Attardi G. 2000. Tight control of respiration by NADH dehydrogenase ND5 subunit gene expression in mouse mitochondria. *Molecular and Cellular Biology*. *20*(3), 805-815.

Ballard JWO, Melvin RG, Katewa SD, Maas K. 2007. Mitochondrial DNA variation is associated with measurable differences in life-history traits and mitochondrial metabolism in *Drosophila simulans*. *Evolution. 61*(7), 1735-1747.

Bazin E, Glémin S, Galtier N. 2006. Population size does not influence mitochondrial genetic diversity in animals. *Science*. *312*(5773), 570-572.

Blackburn D. Evolutionary origins of viviparity in fishes. In Grier HJ, Uribe MC eds. 2005. Viviparous Fishes. *New Life Publications.* Homestead, Florida. 287-301.

Bogue CW, Ganea GR, Sturm E, Ianucci R, Jacobs HC. 2000. Hex expression suggests a role in the development and function of organs derived from foregut endoderm. *Developmental Dynamics*. *219*(1), 84-89.

Boore JL. 1999. Animal mitochondrial genomes. *Nucleic Acids Resources.* 27:1767-1780.

Burridge CP. 1999. Molecular phylogeny of *Nemadactylus* and *Acantholatris* (Perciformes: Cirrhitoidea: Cheilodactylidae), with implications for taxonomy and biogeography. *Molecular Phylogenetics and Evolution*. 13(1), 93-109.

Campos-Soto R, Torres-Pérez F, Solari A. 2015. Phylogenetic incongruence inferred with two mitochondrial genes in *Mepraia* spp. and *Triatoma* *eratyrusiformis* (Hemiptera, Reduviidae). *Genetics and Molecular biology*. 38(3), 390-395.

Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution*. *17*(4), 540-552.

Chan CX, Ragan MA. 2013. Next-generation phylogenomics. *Biology Direct*. *8*(1), 3.

Chen R, Jiang LY, Qiao GX. 2012. The effectiveness of three regions in mitochondrial genome for aphid DNA barcoding: a case in Lachininae. *PloS One*. 7(10), e46190.

Comas I, Moya A, González-Candelas F. 2007. From phylogenetics to phylogenomics: the evolutionary relationships of insect endosymbiotic γ-Proteobacteria as a test case. *Systematic Biology*. *56*(1), 1-16.

Cong Q, Shen J, Borek D, Robbins RK, Otwinowski Z, Grishin NV. 2016. Complete genomes of Hairstreak butterflies, their speciation, and nucleo-mitochondrial incongruence. *Scientific Reports*. 6.

Consuegra S, John E, Verspoor E, De Leaniz CG. 2015. Patterns of natural selection acting on the mitochondrial genome of a locally adapted fish species. *Genetics Selection Evolution*. *47*(1), 58.

Corona-Santiago DK, Domínguez-Domínguez O. 2013. Goodeids phylogeny. In Kempkes M, Köck M, Stawikouski R, eds: Contributions to the biology and conservation of splitfins (Goodeidae). Germany: *Westarp Wissenschaften.* 681.

Corona-Santiago DK, Galván-Tirado C, García-De León FJ, Doadrio I, Domínguez-Domínguez O. 2017. The complete mitogenome of the live-bearing fish *Xenotoca variata* (Bean, 1887)(Actinopterygii: Goodeidae). *Mitochondrial DNA Part B*. *2*(1), 119-121.

Costa WJ 2011. Phylogenetic position and taxonomic status of *Anablepsoides*, *Atlantirivulus*, *Cynodonichthys*, *Laimosemion* and *Melanorivulus* (Cyprinodontiformes: Rivulidae). *Ichthyological Exploration of Freshwaters*. *22*(3), 233.

Costa WJ. 1998. Phylogeny and classification of the Cyprinodontiformes (Euteleostei: Atherinomorpha): A reappraisal. En: Malabarba, L.R.; Reis, R.E.; Vari, R.P.; Lucena, Z.M. Lucena, C.A.S. (Eds.), *Phylogeny and Classification of Neotropical Fishes.* Edipucrs, Porto Alegre. p. 519-536.

Crowl TM, Reneau PC, Broughton RE. 2007. *Cyprinodon rubrofluviatilis* mitochondrion, complete genome. Direct submission. Oklahoma Biological Survey, University of Oklahoma (Unpublished data, GenBank: EF442803.2).

Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods*. *9*(8), 772-772.

Delsuc F, Brinkmann H, Philippe H. 2005. Phylogenomics and the reconstruction of the tree of life. *Nature Reviews Genetics*. *6*(5), 361-375.

Doadrio I, Domínguez O. 2004. Phylogenetic relationships within the fish family Goodeidae based on Cytochrome b sequence data. *Molecular Phylogenetics and Evolution.* 31:416–430.

Domínguez-Domínguez O, Doadrio I, Perez-Ponce de Leon G. 2006. Historical biogeography of some river basins in central Mexico evidenced by their goodeine freshwater fishes: A preliminary hypothesis using secondary Brooks parsimony analysis. *Journal of Biogeography.* 33:1437–1447.

Domínguez-Domínguez O. 2008a. Filogeografía de *Zoogoneticus quitzeoensis*, *Xenotoca variata* y *Alloophorus robustus* (Cyprinodontiformes: Goodeidae) en el centro de México: implicaciones taxonómicas y de conservación. Doctoral Dissertation. Instituto de Biología, UNAM. Distrito Federal, México.

Dowling DK, Friberg U, Lindell J. 2008. Evolutionary implications of non-neutral mitochondrial genetic variation. *Trends in Ecology Evolution*. *23*(10), 546-554.

Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*. *29*(8). 1969-1973.

Duchêne S, Archer FI, Vilstrup J, Caballero S, Morin PA. 2011. Mitogenome phylogenetics: the impact of using single regions and partitioning schemes on topology, substitution rate and divergence time estimation. *PLoS One*. *6*(11), e27138.

Dunn CW, Hejnol A, Matus DQ, Pang K, Browne WE, Smith SA, Sørensen MV. 2008. Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature*. *452*(7188), 745-749.

Easlon E, Tsang F, Skinner C, Wang C, Lin SJ. 2008. The malate–aspartate NADH shuttle components are novel metabolic longevity regulators required for calorie restriction-mediated life span extension in yeast*. Genes Development*. *22*(7), 931-944.

Egea R, Casillas S, Barbadilla A. 2008. Standard and generalized McDonald–Kreitman test: a website to detect selection by comparing different classes of DNA sites. *Nucleic Acids Research*. *36* (suppl 2). W157-W162.

Feschotte C, Pritham EJ. 2007. DNA transposons and the evolution of eukaryotic genomes. *Annual Review of Genetics. 41*, 331-368.

Foulds L, Robinson R. 1981. Enumeration of binary phylogenetic trees. *Combinatorial Mathematics VIII*. 187-202.

Francino MP, Ochman H. 1997. Strand asymmetries in DNA evolution. *Trends in Genetics.13*(6), 240-245.

Frank AC, Lobry JR. 1999. Asymmetric substitution patterns: a review of possible underlying mutational or selective mechanisms. *Gene*. *238*(1), 65-77.

Hertwig ST. 2008. Phylogeny of the Cyprinodontiformes (Teleostei, Atherinomorpha): the contribution of cranial soft tissue characters. *Zoologica Scripta*. *37*(2), 141-174.

Huang H, Li Y. 2013. MASTtreedist: visualization of tree space based on maximum agreement subtree. *Journal of Computational Biology*. *20*(1), 42-49.

Huelsenbeck JP, Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*. *17*(8), 754-755.

Inoue JG, Miya M, Tsukamoto K, Nishida M. 2003. Basal actinopterygian relationships: a mitogenomic perspective on the phylogeny of the “ancient fish”. *Molecular Phylogenetics and Evolution. 26*(1), 110-120.

Jukes TH, Cantor CR. 1969. Evolution of protein molecules. *Mammalian Protein Metabolism*. *3*(21), 132.

Katoh K, Misawa K, Kuma KI, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic acids Research*. *30*(14), 3059-3066.

Kawaguchi M, Tomita K, Sano K, Kaneko T. 2015. Molecular events in adaptive evolution of the hatching strategy of ovoviviparous fishes. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*. *324*(1), 41-50.

Kong XF, Li JT, Sun XW. 2016. Complete mitochondrial genome of the guppy (*Poecilia reticulata*). *Mitochondrial DNA Part A*. *27*(1), 228-229.

Lane N. 2011. Mitonuclear match: optimizing fitness and fertility over generations drives ageing within generations. *Bioessays*. *33*(11), 860-869.

Lane N. 2011. The costs of breathing. *Science*. *334*(6053), 184-185.

Lanfear R, Calcott B, Ho SY, Guindon S. 2012. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution*. *29*(6), 1695-1701.

Larmuseau MHD, Raeymaekers JAM, Hellemans B, Van Houdt JKJ, Volckaert FAM. 2010. Mito-nuclear discordance in the degree of population differentiation in a marine goby. *Heredity*. 105(6), 532-542.

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

Liu L, Pearl DK, Brumfield RT, Edwards SV. 2008. Estimating species trees using multiple-allele DNA sequence data. *Evolution*. 62(8), 2080-2091.

Lund R. 1980. Viviparity and intrauterine feeding in a new Holocephalan fish from the Lower Carboniferous of Montana. *Science.* 209:697-699.

Luo R, Liu B, Xie Y *et al.,* 2009. Soapdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience.* 1:18.

Mackiewicz P, Gierlik A, Kowalczuk M, Dudek MR, Cebrat S. 1999. How does replication-associated mutational pressure influence amino acid composition of proteins?. *Genome Research*. *9*(5), 409-416.

Maddison WP, Maddison DR. 2017. Mesquite: a modular system for evolutionary analysis. Version 3.2 http://mesquiteproject.org.

McCormack JE, Hird SM, Zellmer, AJ, Carstens BC, Brumfield RT. 2013. Applications of next-generation sequencing to phylogeography and phylogenetics. *Molecular Phylogenetics and Evolution. 66*(2), 526-538.

McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature*. *351*(6328), 652.

Meiklejohn CD, Montooth KL, Rand DM. 2007. Positive and negative selection on the mitochondrial genome. *Trends in Genetics*. *23*(6), 259-263.

Meiklejohn KA, Danielson MJ, Faircloth BC, Glenn TC, Braun EL, Kimball RT. 2014. Incongruence among different mitochondrial regions: a case study using complete mitogenomes. *Molecular Phylogenetics and Evolution*. *78*, 314-323.

Meyer A, Lydeard C. 1993. The evolution of copulatory organs, internal fertilization, placentae and viviparity in killifishes (Cyprinodontiformes) inferred from a DNA phylogeny of the tyrosine kinase gene X-src. *Proceedings of the Royal Society of London B: Biological Sciences. 254*(1340), 153-162.

Miya M, Takeshima H, Endo H, Ishiguro NB, Inoue JG, Mukai T, Shirai SM. 2003. Major patterns of higher teleostean phylogenies: a new perspective based on 100 complete mitochondrial DNA sequences. *Molecular Phylogenetics and Evolution*. *26*(1), 121-138.

Moreno-Sánchez R, Hogue BA, Hansford RG. 1990. Influence of NAD-linked dehydrogenase activity on flux through oxidative phosphorylation. *Biochemical Journal*. *268*(2), 421-428.

Morton RA, Morton BR. 2007. Separating the effects of mutation and selection in producing DNA skew in bacterial chromosomes. *BMC Genomics*. *8*(1), 369.

Pacheco MA, Battistuzzi FU, Lentino M, Aguilar RF, Kumar S, Escalante AA. 2011. Evolution of modern birds revealed by mitogenomics: timing the radiation and origin of major orders. *Molecular Biology and Evolution*. *28*(6), 1927-1942.

Parenti LR. 1981. A taxonomic revision of the Andean killifish genus *Orestias* (Cyprinodontiformes, Cyprinodontidae). *American Museum of Natural History*.

Patwardhan A, Ray S, Roy A. 2014. Molecular markers in phylogenetic studies-A Review. *Journal of Phylogenetics Evolutionary Biology*. *2014*.

Perna NT, Kocher TD. 1995. Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. *Journal of Molecular Evolution*. *41*(3), 353-358.

Pesole G, Gissi C, De Chirico A, Saccone C. 1999. Nucleotide substitution rate of mammalian mitochondrial genomes. *Journal of Molecular Evolution*. *48*(4), 427-434.

Philippe H, Brinkmann H, Lavrov DV, Littlewood DTJ, Manuel M, Wörheide G, Baurain D. 2011. Resolving difficult phylogenetic questions: why more sequences are not enough. *PLoS Biol*. *9*(3), e1000602.

Philippe H, Chenuil A, Adoutte A. 1994. Can the Cambrian explosion be inferred through molecular phylogeny?. *Development*, *1994* (Supplement). 15-25.

Philippe H, Derelle R, Lope P, Pick K, Borchiellini C, Boury-Esnault N, Da Silva C. 2009. Phylogenomics revives traditional views on deep animal relationships. *Current Biology. 19*(8), 706-712.

Pohl M, Milvertz FC, Meyer A, Vences M. 2015. Multigene phylogeny of cyprinodontiform fishes suggests continental radiations and a rogue taxon position of *Pantanodon*. *Vertebrate Zoology*. *65*(1), 37-44.

Pollux BJA, Pires MN, Banet AI, Reznick DN. 2009. Evolution of placentas in the fish family Poeciliidae: an empirical study of macroevolution. *Annual Review of Ecology, Evolution and Systematics. 40*, 271-289.

Pons J, Bauzà-Ribot MM, Jaume D, Juan C. 2014. Next-generation sequencing, phylogenetic signal and comparative mitogenomic analyses in Metacrangonyctidae (Amphipoda: Crustacea). *BMC Genomics. 15*(1), 566.

Rambaut A, Suchard MA, Xie D, Drummond AJ. 2014. Tracer v1.6, Available from: http://beast.bio.ed.ac.uk/Tracer.

Reichwald K, Lauber C, Nanda I, Kirschner J, Hartmann N, Schories S, Glöckner G. 2009. High tandem repeat content in the genome of the short-lived annual fish *Nothobranchius furzeri*: a new vertebrate model for aging research. *Genome Biology*. *10*(2), R16.

Ritchie MG, Hamil RM, Graves JA *et al.,* 2007. Sexual differentiation: population genetic divergence and sexual dimorphism in mexican goodeid fish. *Journal of Evolutionary Biology.* 20:2048-2055.

Robinson DF, Foulds LR. 1979. Comparison of weighted labelled trees. In *Combinatorial mathematics VI* (pp. 119-126). Springer Berlin Heidelberg.

Rocha-Olivares A, Garber NM, Garber AF, Stuck KC. 2005. Structure of the mitochondrial control region and flanking tRNA genes of *Mugil cephalus. Hidrobiológica*. 15(2), 139-149.

Roje DM. 2014. Evaluating the effects of non-neutral molecular markers on phylogeny inference. *PloS One*, 9(2), e87428.

Rueness EK, Stenseth NC, O'donoghue M, Boutin S, Ellegren H, Jakobsen KS. 2003. Ecological and genetic spatial structuring in the Canadian lynx. *Nature*. 425(6953), 69-72.

Saitou N, Nei M. 1986. The number of nucleotides required to determine the branching order of three species, with special reference to the human-chimpanzee-gorilla divergence. *Journal of Molecular Evolution*. *24*(1-2), 189-204.

Schierwater B, Eitel M, Jakob W, Osigus HJ, Hadrys H, Dellaporta SL, DeSalle R. 2009. Concatenated analysis sheds light on early metazoan evolution and fuels a modern “urmetazoon” hypothesis. *PLoS Biol*. *7*(1), e1000020.

Setiamarga DH, Miya M, Yamanoue Y, Mabuchi K, Satoh TP, Inoue JG, Nishida M. 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), 598-605.

Suyama M, Torrents D, Bork P. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Research*. *34*(suppl 2). W609-W612.

Taberlet, P. 1996. The use of mitochondrial DNA control region sequencing in conservation genetics. *Molecular Genetic Approaches in Conservation*. 125-142.

Talavera G, Castresana J. 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systematic Biology*. *56*(4), 564-577.

Talbert ME, Barnett B, Hoff R, Amella M, Kuczynski K, Lavington E, Eanes WF. 2015. Genetic perturbation of key central metabolic genes extends lifespan in Drosophila and affects response to dietary restriction. In *Proccedings of the Royasl Society of London B*. 282,1815, 20151646.

Tillier ER, Collins RA. 2000. The contributions of replication orientation, gene direction, and signal sequences to base-composition asymmetries in bacterial genomes. *Journal of Molecular Evolution*. *50*(3), 249-257.

Tobe SS, Kitchener AC, Linacre AM. 2010. Reconstructing mammalian phylogenies: a detailed comparison of the cytochrome b and cytochrome oxidase subunit I mitochondrial genes. *PloS One*. 5(11), e14156.

Velazco PM, Patterson BD. 2008. Phylogenetics and biogeography of the broad-nosed bats, genus *Platyrrhinus* (Chiroptera: Phyllostomidae). *Molecular Phylogenetics and Evolution*. 49(3), 749-759.

Verma R, Singh M, Kumar S. 2016. Unraveling the Limits of Mitochondrial Control Region to Estimate the Fine Scale Population Genetic Differentiation in Anadromous Fish *Tenualosa ilisha*. *Scientifica*. 2016.

Wallace DC, Fan W. 2010. Energetics, epigenetics, mitochondrial genetics. *Mitochondrion*. *10*(1), 12-31.

Webb SA. 1998. *A phylogenetic analysis of the Goodeidae (Teleostei: Cyprinodontiformes)* (Doctoral dissertation).

Whitehead A. 2009. Comparative mitochondrial genomics within and among species of killifish. *BMC Evolutionary Biology*. *9*(1), 11.

Willerslev E, Gilbert MTP, Binladen JHoSY. Campos PF, Ratan A, Nowak-Kemp M. 2009. Analysis of complete mitochondrial genomes from extinct and extant rhinoceroses reveals lack of phylogenetic resolution. *BMC Evolutionary Biology*. *9*(1), 95.

Wourms JP, Grove BD, Lombardi J. 1981. 1 The Maternal-Embryonic Relationship in Viviparous Fishes. *Fish physiology. 11*, 1-134.

Wourms JP, Lombardi J. 1992. Reflections on the evolution of piscine viviparity. American Zoologist. *32*(2), 276-293.

Xia X. 2013. DAMBE5: a comprehensive software package for data analysis in molecular biology and evolution. *Molecular Biology and Evolution*. *30*(7), 1720-1728.

Zardoya R, Meyer A. 1996. Phylogenetic performance of mitochondrial protein-coding genes in resolving relationships among vertebrates. *Molecular Biology and Evolution*. *13*(7), 933-942.

Zhang W, Yatskievych TA, Baker RK, Antin PB. 2004. Regulation of Hex gene expression and initial stages of avian hepatogenesis by Bmp and Fgf signaling. *Developmental biology*. *268*(2), 312-326.

Figures

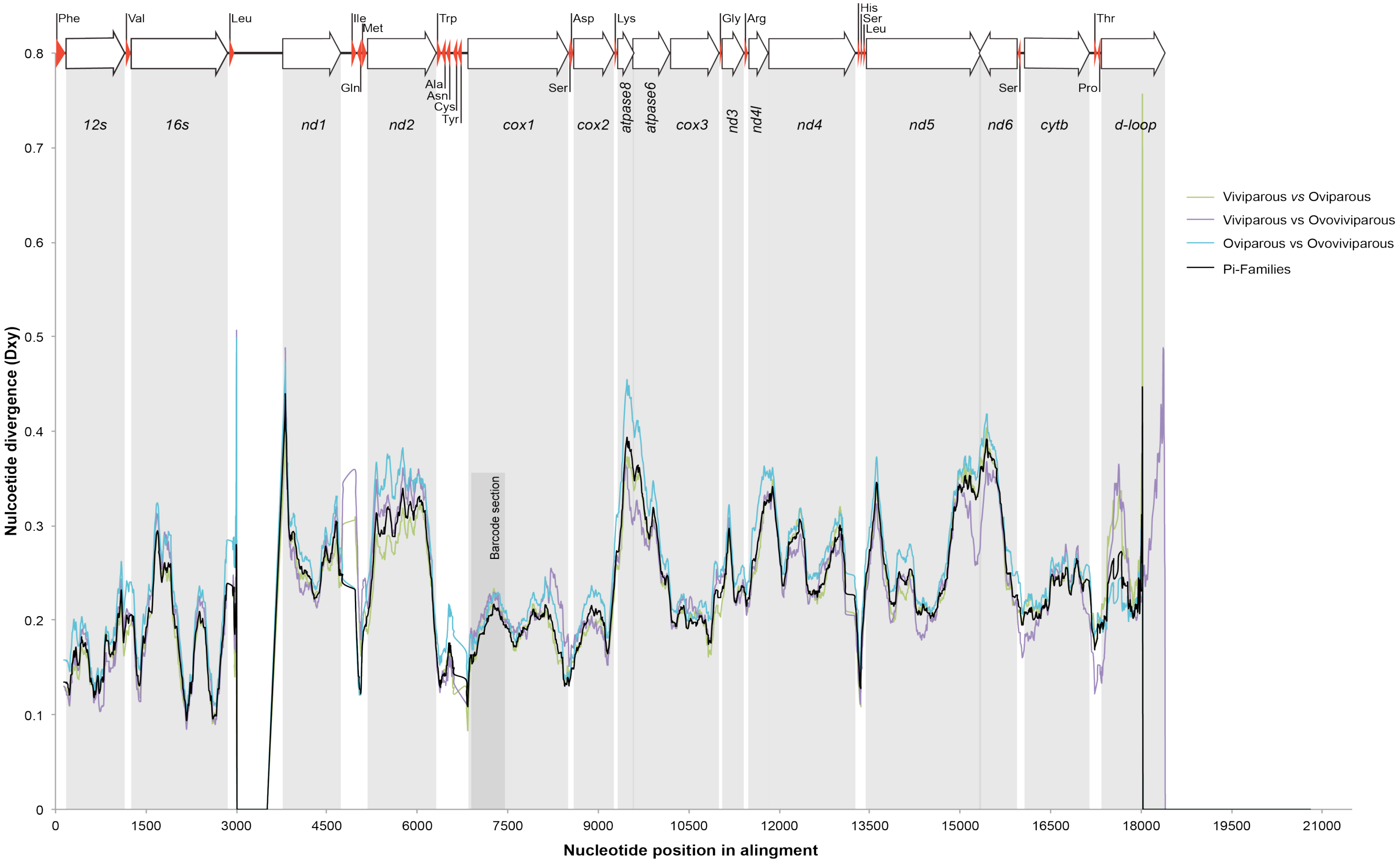


Figure 1. Sliding window of the nucleotide divergence for the complete mitogenome different reproductive modes in Cyprinodontiformes. Pi values in family represent the nucleotide diversity.

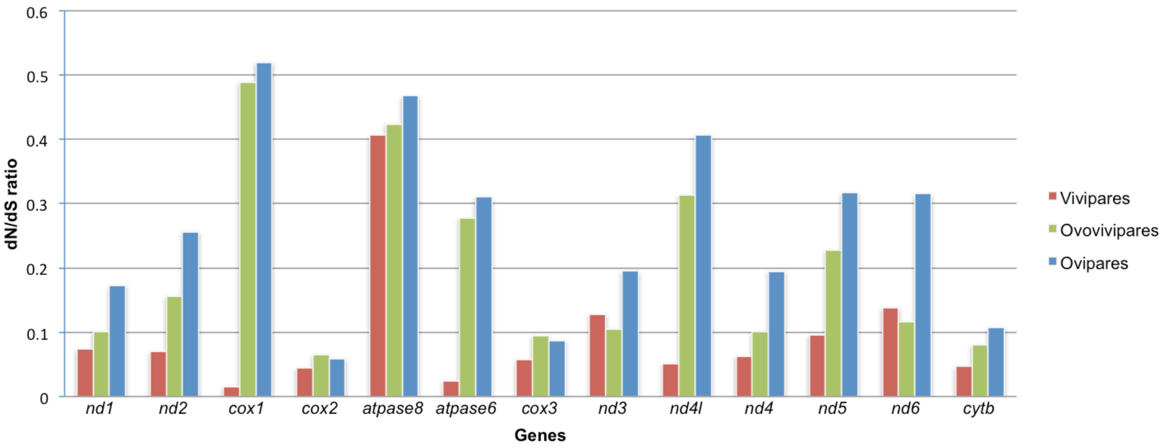


Figure 2. Comparison of the ω ratio (dN/dS) between viviparous, ovoviviparous and oviparous encoding regions**.**

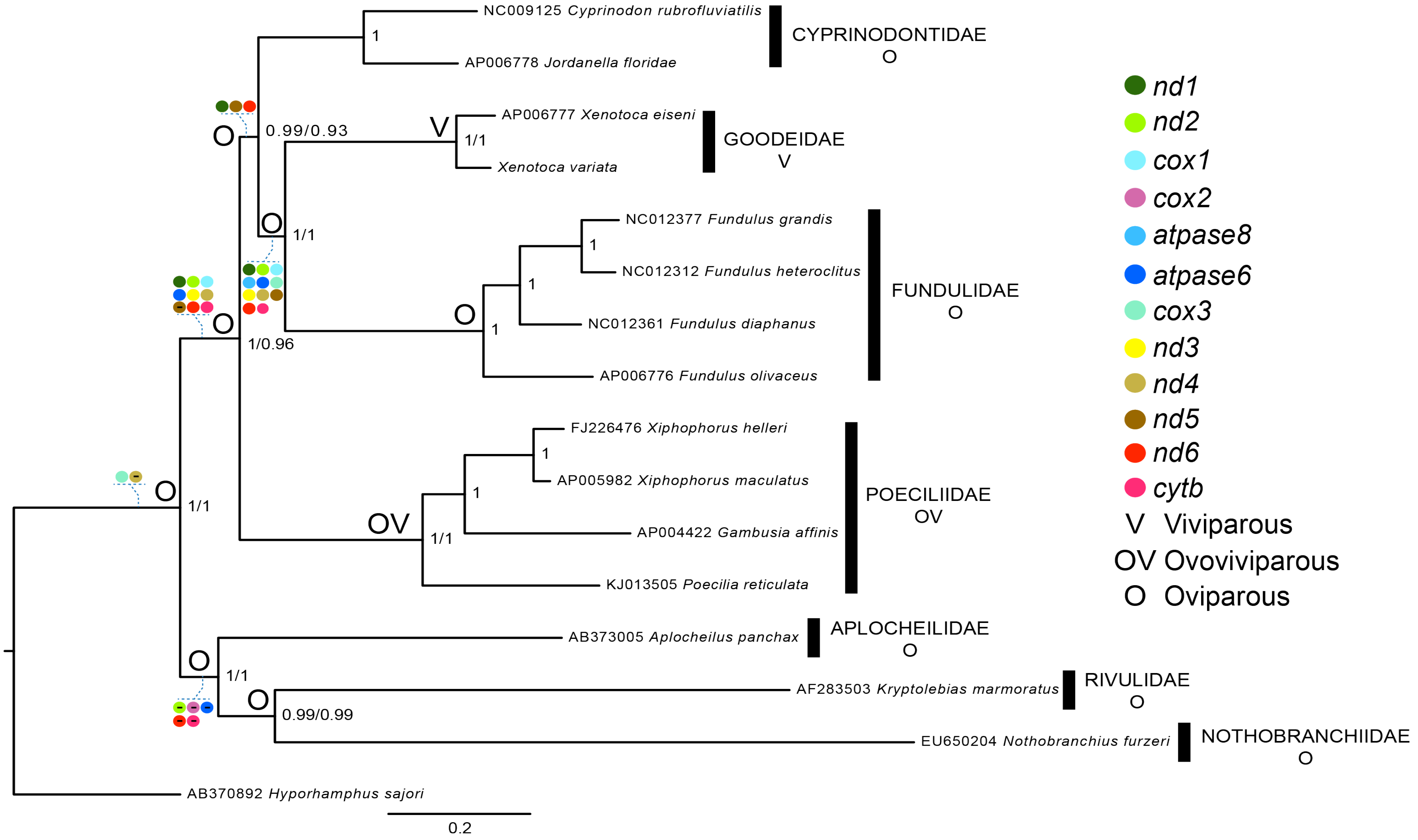
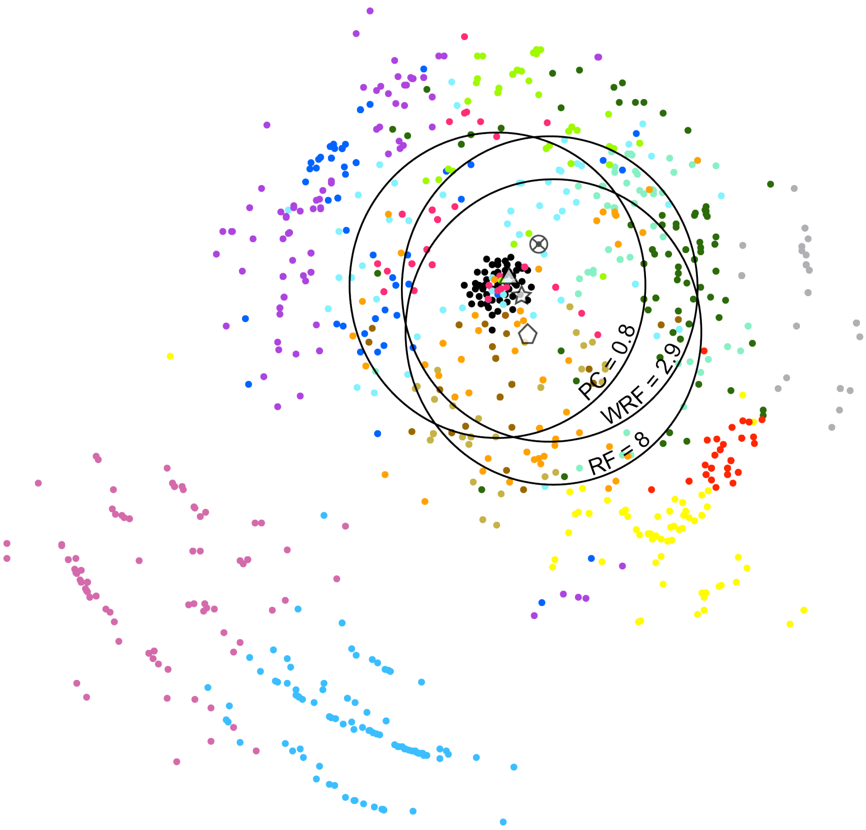


Figure 3. Bayesian phylogenetic inference based on complete mitogenomes, ancestral states estimation (letters in nodes) and neutrality deviation (color circles) for mitochondrial genes estimated with MKT. The values represent the posterior probabilities in phylogenetic inference (left) and ancestral states (right). The circles with a dash represent genes under negative selection.



Trees	PC	RFw	RF
■ Mito			
△ 3g	**1**	**0.9968**	**0**
☆ 5g	**1**	**2.5943**	**0**
⬠ 2g	**0.9768**	**1.3662**	**1**
⊗ 3g RFw	**0.9269**	**3.1302**	**1**
■ *cox1*	**0.9786**	**2.7160**	**1**
■ *nd4l*	**0.9768**	**2.9025**	**1**
■ *cytb*	**0.8933**	**2.9512**	**5**
■ *atpase6*	**0.8767**	**3.4679**	**4**
■ *nd5*	**0.8473**	**2.7703**	**4**
■ *nd4*	0.8285	2.8092	6
■ *nd2*	0.7773	2.8500	7
■ *cox3*	0.7542	2.9306	6
■ *nd1*	0.7394	2.8428	6
■ *cox1* BC	0.6530	3.1076	9
■ *d-loop*	0.6348	3.1858	12
■ *nd3*	0.6085	3.1752	9
■ *nd6*	0.5883	3.1589	12
■ *atpase8*	0.4990	8.8555	13
■ *cox2*	0.2799	5.1793	16

Figure 4. Robinson-Foulds distances and Patristic correlation using MDS of phylogenetic trees based on Bayesian MCMC analyses of different mitochondrial regions. PC values are the patristic correlation between majority-rule consensus trees of each gene and mitogenome. RF and RF_W represent the weighted and unweighted Robison-Foulds distances. Mito = phylogenetic trees based on the complete mitogenome; 3g = trees based on three genes (*cox1*, *nd4l* and *cytb*) with the highest PC values; 5g = trees based on five genes (*cox1*, *nd4l*, *cytb*, *atpase6* and *nd5*) with the highest PC values; 2g = trees based on two genes (*cox1* and *nd4l*) with the lowest RF and PC values; 3g RF_W = trees based on three genes (*cox1*, *nd4l* and *nd5*) according with the lower RF_W values. Circles are the boundaries for each PC, RF_W and RF. Only majority-rule consensus tree was plotted for 3g, 2g, 5g and 3g RF_W inferences. The *cox1* BC represent the barcode region (628bp).

Tables

Table 1. Mitogenomes composition of different reproductive modes and families in Cyprinodontiformes.

Species	**RM**	**Family**	**M size**	**A%**	**T%**	**A+T**	**AT Skew**	**G%**	**C%**	**G+C**	**GC Skew**	**Gene duplication (Position)**
Aplocheilus panchax AB373005	O	Aplocheilidae	16 519	29.2	28.1	57.3	0.0206	15.8	26.9	42.7	-0.2605	
Nothobranchius furzeri EU650204	O	Nothobranchiidae	19 527	31.2	30.2	61.4	0.0162	14.9	23.6	38.5	-0.2254	*d-loop* (17 750-18 932)
Kryptolebias marmoratus AF283503	O	Rivulidae	17 329	27.5	29.2	56.7	-0.0288	15.8	27.5	43.3	-0.2699	*d-loop* (2813-3607)
Xenotoca eiseni AP006777	V	Goodeidae	16 735	29.2	29.3	58.5	-0.0003	14.8	26.7	41.5	-0.2861	tRNA-Met (3952-4020)
Xenotoca variata *XXX*	V		16 511	28.7	29.2	57,8	-0.0086	15.3	26.8	42.1	-0.2741	
Gambusia affinis AP004422	OV	Poeciliidae	16 614	28.0	27.1	55.1	0.0165	15.7	29.2	44.9	-0.3021	
Poecilia reticulata KJ013505	OV		16 570	28.1	30.2	58.3	0.0365	14.6	27.1	41.7	-0.3469	
Xiphophorus hellerii FJ226476	OV		16 635	27.0	25.1	52.1	0.0361	16.3	31.6	47.9	-0.3199	

Xiphophorus maculatus AP005982	OV		16 646	27.6	25.6	53.2	0.0373	15.7	31.2	46.9	-0.3309
Cyprinodon rubrofluviatilis NC009125	O	Cyprinodontidae	16 501	26.0	27.4	53.4	-0.0264	17.1	29.4	46.5	-0.2641
Jordanella floridae AP006778	O		16 177	27.1	27.9	55	-0.0130	16.1	28.9	45	-0.2848
Funfulus diaphanus NC012361	O	Fundulidae	16 531	28.9	32.1	61	-0.0530	15.4	23.6	39	-0.2103
Funfulus grandis NC012377	O		16 524	28.5	31.1	59.6	-0.0448	15.9	24.5	40.4	-0.2122
Funfulus heteroclitus NC012312	O		16 526	28.7	31.6	60.3	-0.0491	15.8	23.9	39.7	-0.2051
Funfulus olivaceus AP006776	O		16 509	27.9	29.9	57.8	-0.0331	16.1	26.1	42.2	-0.2365

RM = Reproductive modes (O = Oviparous; V = Viviparous; OV = Ovoviviparous), M size = Mitogenome size

Table 2. Variation in the encoding genes and *d-loop* region comparing all Cyprinodontiformes mitogenomes included.

Gene or region	**Length bp**	**Variable Sites**	**Informative Sites**	**Saturation in codon position 1/2/3**	**ω**	**Indels events (sites)**	**Pi**
nd1	975	549	455	-/-/*	0.1680	0	0.2663
nd2	1046	680	583	-/-/*	0.2724	2(8)	0.3180
cox1	1551	618	538	-/-/-	0.5141	12(20)	0.1927
cox2	691	300	240	-/-/*	0.0663	0	0.1887
atpase8	168	123	107	*/-/*	0.4762	0	0.3669
atpase6	683	441	377	-/-/*	0.3237	4(6)	0.3189
cox3	785	345	276	-/-/*	0.0920	3(6)	0.1957
nd3	349	206	175	-/-/*	0.2048	0	0.2784
nd4l	297	184	146	-/-/-	0.3890	2(7)	0.2625
nd4	1381	805	673	-/-/*	0.1897	1(5)	0.2777
nd5	1838	1079	881	-/-/-	0.3015	9(27)	0.2738
nd6	492	358	286	-/-/*	0.3042	6(44)	0.3333
cytb	1140	569	471	-/-/*	0.1063	2(3)	0.2302
d-loop	848	256	152	*	--	94(1527[1])	0.2287

ω = dN/dS Ratio of nonsynonymous sites/synonymous nucleotide substitutions, Pi = Nucleotide diversity, [1] Including *d-loop* duplication in *Nothobranchius furzeri*, * Substitution saturation stastistly significant for DNA sequences via Xie *et al.,* 2003.

Supplementary information

Table 1S. Model of nucleotide substitution for encoding genes and partition used in phylogenetic reconstructions under Akaike information criteria AIC.

Gene or region	**Model abbreviation**	**Pinv**	**Gamma**
nd1	TIM2+I+G	0.306	0.451
nd2	GTR+I+G	0.271	0.801
cox1	TPM3uf+I+G	0.538	0.508
cox2	TPM1uf+I+G	0.379	0.379
atpase8	HKY+G	--	0.488
atpase6	TIM2+G	--	0.323
cox3	TPM1uf+I+G	0.473	0.596
nd3	HKY+I+G	0.337	0.712
nd4l	TPM2uf+G	--	0.425
nd4	GTR+I+G	0.314	0.568
nd5	TIM3+I+G	0.338	0.829
nd6	GTR+I+G	0.078	0.490
cytb	TPM3uf+I+G	0.361	0.436
d-loop	GTR+I+G	1.051	0.091
Barcode *cox1*	GTR+I+G	0.506	0.539

Pinv = Proportion of invariable sites

Table 2S. McDonald-Kreitnab neutrality test for mitochondrial genes in groups with different mode of reproduction (V viviparous, OV ovoviviparous and O oviparous).

Groups	**Gene**	**NI**	**α**
V(Goodeidae)O(Fundulidae)	nd1	0.372*	0.627
	nd2	0.301*	0.698
	cox1	0.136*	0.863
	cox2	1.008	0.008
	atpase8	0.125*	0.874
	atpase6	0.090*	0.909
	cox3	0.411*	0.058
	nd3	0.355*	0.644
	nd4l	0.523	0.469
	nd4	0.300*	0.699
	nd5	0.319*	0.680
	nd6	0.414*	0.585
	cytb	0.241*	0.758
O(Cyprinodontidae)(V,O(Goodeidae,Fundulidae))	nd1	0.335*	0.664
	nd2	0.600	0.399
	cox1	0.451	0.548
	cox2	1.000	0.000
	atpase8	0.232	0.767
	atpase6	1.706	-0.706
	cox3	1.578	-0.578
	nd3	1.084	-0.084
	nd4l	0.379	0.620
	nd4	0.910	0.089
	nd5	0.632*	0.367
	nd6	0.342*	0.657
	cytb	0.732	0.267
OV(Poeciliidae)(O,VO(Cyprinodontidae(Goodeidae,Fundulidae))	nd1	0.092*	0.907
	nd2	0.101*	0.898

	cox1	0.331*	0.668
	cox2	0.322	0.677
	atpase8	0.932	0.067
	atpase6	0.331*	0.668
	cox3	0.650	0.349
	nd3	0.229*	0.770
	nd4l	1.4	-0.400
	nd4	0.328*	0.671
	nd5	2.622*	-1.622
	nd6	0.123*	0.876
	cytb	0.296*	0.703
O(OV(Poeciliidae)O,VO(Cyprinodontidae(Goodeidae,Fundulidae)	nd1	0.922	0.077
	nd2	0.270	0.729
	cox1	0.302	0.697
	cox2	0.414	0.585
	atpase8	0.000	1.000
	atpase6	0.299	0.700
	cox3	0.075*	0.924
	nd3	0.405	0.594
	nd4l	0.502	0.497
	nd4	2.970*	-1.970
	nd5	2.262	-1.262
	nd6	0.377	0.622
	cytb	0.518	0.481
O(Aplocheilidae), O(Rivulidae,Nothobranchiidae)	nd1	0.909	0.090
	nd2	1.586*	-0.586
	cox1	1.794	-0.794
	cox2	3.066*	-2.066
	atpase8	0.708	0.291
	atpase6	4.404*	-3.404
	cox3	1.766	-0.766

nd3	1.440	-0.440
nd4l	1.006	-0.006
nd4	1.269	-0.269
nd5	1.134	-0.134
nd6	2.151*	-1.151
cytb	3.758*	-2.758

* Values with significant deviation from neutrality, NI = Neutrality index, α = Proportion of adaptive substitutions corrected by Jukes&Cantor.

Capítulo 5: The complete mitogenome of the Moroccan *Luciobarbus rifensis* Doadrio, Casal-López & Yahyaoui, 2015 (Actinopterygii: Cyprinidae)

DIUSHI KERI CORONA-SANTIAGO, MIRIAM CASAL-LÓPEZ, SILVIA PEREA and IGNACIO DOADRIO

Departamento de Biodiversidad y Biología Evolutiva, Museo Nacional de Ciencias Naturales, CSIC, Madrid, España.

Published in: Mitochondrial DNA Part B: Resources
http://dx.doi.org/10.1080/23802359.2016.1250137

Abstract

The Rifian barbel (*Luciobarbus rifensis*) is a tetraploid cyprinid species from North Africa. The aim of this work is to characterize the mitogenome of *Luciobarbus rifensis* in order to contribute in the future exploration of regions. The circular mitogenome of the Rifian barbel consist of 37 genes in 16 607bp. The cyprinid mitogenome most related with *L. rifensis* available is *L. capito* with pairwise identity of 94.3%. The mitogenome of *L. rifensis* will contribute searching new molecular markers for the study of this tetraploid fish group.

Keywords: Rifian barbel, cyprinid, mitogenome, North Africa.

Main text

The species *Luciobarbus rifensis* Doadrio, Casal-López and Yahyaoui, 2015 is representative of the fish family Cyprinidae and endemic of rivers from the Rif area of Morocco (Casal-López *et al.*, 2015). The molecular marker discovered is relevant due the tetraploid condition in *Luciobarbus.* The aim of this work was to characterize the mitogenome of *L. rifensis* to the future exploration of regions for genetic and evolution studies in this fish group.

For mitogenome sequencing we used a tissue (caudal fin) of *Luciobarbus rifensis* from Laou River (Laou basin, Mediterranean slope, Morocco; Locality: Chefchaouen N35.12154058700472 E -5.289761162745499) stored in the fish collection of the National Museum of Natural Sciences from Madrid (Spain) (Voucher specimen: MNCN-235026). Quality evaluation and assemblage of 6 092 528 reads (from Illumina MiSeq platform PE300) was performed using FastQC (www.bioinformatics.babraham.ac.uk/projects/fastqc/) and SOAPdenovo2 (Luo *et al.,* 2012), respectively. The mitogenome was extracted from the contigs obtained. Gene annotations and identification of tRNA genes was using MitoAnnotator (Iwasaki et al. 2013) and tRNAscanSE v1.21 (Schattner *et al.,* 2005). Phylogenetic reconstruction for sample validation was performed under Neighbor-Joining assumption including 23 species of the Cyprinidae. The full alignment was built in MAFFT v7.222 (Katoh *et al.,* 2002).

The base composition of the circular mitogenome of the *Luciobarbus rifensis* (GenBank accession: KX348041) was as follows: A=30.9%, C=27.9%, G=16.8% and T=24.5% (GC-rich=44.6%), consisting of 37 genes in 16 607bp (13 protein-coding genes, 2 rRNA genes and 22 tRNA genes). The size of the tRNA genes in mitogenome is 67-77bp. In general, the phylogenetic relationships for the species included in this work (Fig. 1) were consistent with those obtained in previous works for Cyprinidae (Wang *et al.,* 2012; Wang *et al.,* 2013). As it has been previously reported (Casal-López *et al.,* 2015), of the mitogenomes available the most related with *L. rifensis* is *L. capito* (GenBank accession: JX987313, 16 603bp), with pairwise identity of 94.3%, followed by *Barbus barbus* (GenBank accession: AB23965, 16 600bp) with 90.8%. No gene rearrangement was identified in the mitogenome of *L. rifensis*. On the light of our results, the mitogenome of *L. rifensis* will contribute searching new molecular markers for the study of this tetraploid fish group.

Acknowledgments

This work was supported by the Ministry of Economy and Competiveness Council of Spain under Grant CGL2013-41375-P. We thanks to the Supercomputing Center of Galicia (CESGA) for the informatics resources facilities.

Declaration of interest

The authors report no conflicts of interest and are responsible for the content and writing of the article.

References

Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. Available from: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Casal-López M, Perea S, Yahyaoui A, Doadrio I. 2015. Taxonomic review of the genus *Luciobarbus* Heckle, 1843 (Actinopterygii, Cyprinidae) from northwestern Morocco with the description of three new species. *Graellsia.* 71(2):e027. Available from: http://dx.doi.org/10.3989/graellsia.2015.v71.135

Iwasaki W, Fukunaga T, Isagozawa R, Yamada K, Maeda Y, Satoh TP, Sado T, Mabuchi K, Takeshima H, Miya M, Nishida M. 2013. MitoFish and MitoAnnotator: A mitochondrial genome database of fish with an accurate and automatic annotation pipeline. *Molecular Biology and Evolution*. 30(11):2531-2540 doi: 10.1093/molbev/mst141.

Katoh K, Misawa K,Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*. 30(14):3059-3066.

Luo R, Liu B, Xie Y, Li Z, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q *et al.,,* 2012. SOAPdenovo2: and empirically improved memory-efficient short-read *de novo* assembler. *GigaScience*. 1:18.

Schattner P, Brooks AN, Lowe TM. 2005. The tRNAscan-SE, snoscan and snoGPS web servers for detection of tRNAS and snoRNAS. *Nucleic Acids Research.* 33(2):W686-W689 doi: 10.1093/nar/gki366.

Wang J, Wu XY, Chen Z, Yue Z, Ma W, Chen SY, Xiao H, Murphy RW, Zhang YP, Zan RG, Luo J. 2013. Molecular phylogeny of European and African *Barbus* and their West Asian relatives in the Cyprininae (Teleostei: Cypriniformes) and orogenesis of the Qinghai-Tibetan Plateau. *Chinese Science Bulletin.* 58(31):3738-3746 doi: : 10.1007/s11434-013-5878-z.

Wang XZ, Gan XN, Li JB, Mayden RL, He SP. 2012. Cyprinid phylogeny base don Bayesian and máximum likelihood analyses of partitioned data: implications for Cyprinidae systematics. *Science China Life Sciences*. 55(9):761-773 doi: 10.1007/s11427-012-4366-z.

Figures

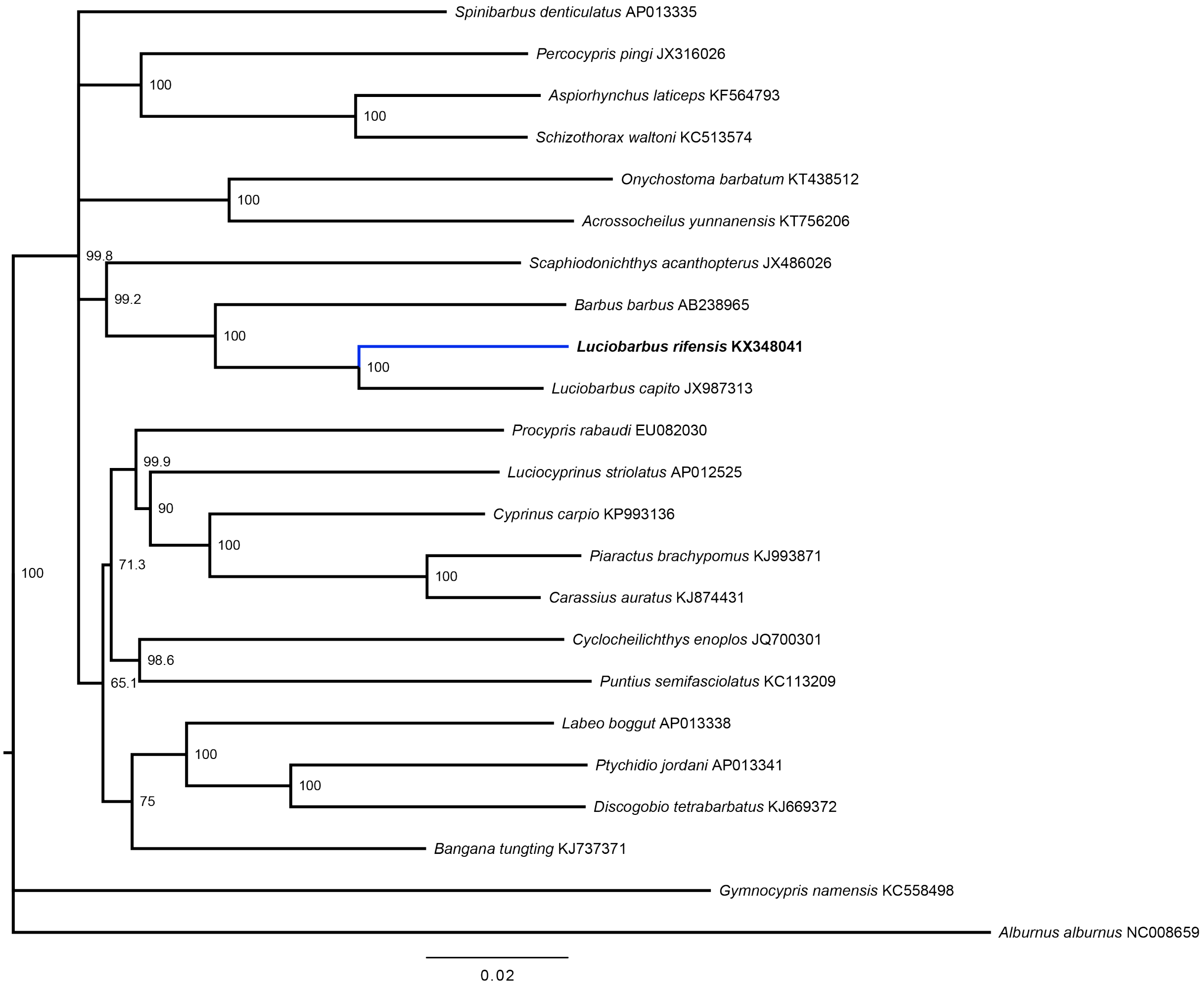


Figure 1. Sample validation based on Neighbor-Joining tree of 23 species of the Cyprinidae family (1000 bootstrap replicates using uncorrected *p*-distances).

Capítulo 6: The complete mitogenome of the live-bearing fish *Xenotoca variata* (Bean, 1887) (Actinopterygii: Goodeidae)

DIUSHI KERI CORONA-SANTIAGO[1], CAROLINA GALVÁN TIRADO[2], FRANCISCO JAVIER GARCÍA-DE LEÓN[2], OMAR DOMÍNGUEZ-DOMÍNGUEZ[3] and IGNACIO DOADRIO[1]

[1]*Departamento de Biodiversidad y Biología Evolutiva, Museo Nacional de Ciencias Naturales, Madrid, Spain*

[2]*Laboratorio de Genética para la Conservación, Centro de Investigaciones Biológicas del Noroeste, S.C. La Paz, Mexico.*

[3]*Laboratorio de Biología Acuática, Facultad de Biología, Universidad Michoacana de San Nicolás de Hidalgo, Morelia, Mexico.*

Published in: Mitochondrial DNA Part B: Resources

http://dx.doi.org/10.1080/23802359.2017.1289350

Abstract

The live-bearing fish *Xenotoca variata* is representative of the viviparous Goodeinae subfamily (Goodeidae) from central Mexico. The mitogenome of the *X. variata* consist of 37 genes in 16 462bp. Comparing with *X. eiseni*, the most related of the mitogenomes included, an identity of 91.1% was found and *trna-met* duplication in *X. eiseni* is absent in *X. variata*. The mitogenome provide important information for future studies in evolution of the live-bearing subfamily.

Keywords: live-bearing; Goodeinae; Central Mexico; mitogenome.

Main text

The species *Xenotoca variata* (Bean 1887), is representative of the live-bearing subfamily Goodeinae (Goodeidae) and is distributed in central Mexico: in the Middle Lerma River, Zacapu basin, Cuitzeo Lake, Chapala Lake, Pánuco and Aguanaval basins (Domínguez-Dominguez, 2008a). Goodeines have been a model for the study in evolution, due their peculiar characteristics of breeding strategies and embryo development. Important genetic divergences (Domínguez-Domínguez *et al.,* 2010, Corona-Santiago & Domínguez-Domínguez, 2013), sexual selection (Ritchie *et al.,* 2005, Moyaho *et al.,* 2004), substantial phenotypic plasticity (Fitzsimons, 1972; Macías-Garcia, 1998) have been observed among *X. variata* populations. Hence, the aim of this work is the characterization of mitochondrial genome of *Xenotoca variata*, which could provide relevant information for future studies.

For mitogenome sequencing we use a sample tissue (pectoral fin) of *Xenotoca variata* from Huingo spring (19º54'44.0"N 100º50'00.3"W), Cuitzeo basin, and were storage in the Colección de Peces de la Universidad Michoacana-UMSNH, Mexico (Voucher specimen: CPUM-7031). Briefly DNA was sheared using a Covaris S2 (Woburn, MA) ultrasonicator, and Illumina (Illumina, San Diego, CA) adapters were ligated on using methods derived from Fisher *et al.,* (2011), but using adapters equivalent to Illumina TruSeq with 10nt indexes (Faircloth & Glenn, 2012). Genomic DNA was subjected to sequencing at the Georgia Genomics Facility (University of Georgia). Reads quality was analyze using FastQC (Andrews, 2010), adapters and poorly quality sequences were trimmed using Trimmomatic v0.36 (Bolger *et al.,*

2014) to assembly using SOAPdenovo2 (Luo *et al.,* 2012). Genome annotation was performed using MitoAnnotator (Iwasaki *et al.,* 2013) but the position of all tRNA genes was confirmed using tRNAscanSE v1.21 (Schattner *et al.,* 2005). Phylogenetic reconstruction was performed under Neighbour-Joining analysis including 14 species of the Cyprinodontiformes order available on GenBank. The analysis was conducted with a full alignment built in MAFFT v7.222 (Katoh *et al.,* 2002).

The circular mitogenome of *Xenotoca variata* (GenBank accession: KY471393) consists of 37 genes in 16 511bp (13 protein-coding genes, 2 rRNA genes and 22 tRNA genes) with 12 intergenic spacer sequences (of 1-37bp). The base composition of the genome was as follow: A = 28.7%, C = 26.8%, G = 15.3% and T = 29.2% (GC-rich = 42.1%). Comparing with *X. eiseni*, the most related of the mitogenomes included in Genbank (Fig. 1), an identity of 91.1% was founded and the duplication of *trna-met* gene in *X. eiseni* is absent in *X. variata*. The number of nucleotide differences between both mitogenomes is 1 206bp corresponding to 7.3% of divergence. The complete mitochondrial genome of *X. variata* provides relevant information to posterior genetic and evolutionary studies of *Xenotoca* genus and the Goodeidae family.

Funding

This work was supported by the SAGARPA-FIRCO 2012 under Grant RGA-BCS-12-000003.

Acknowledgments

We thanks to the Supercomputing Center of Galicia (CESGA) for the informatics resources facilities.

Declaration of interest

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

References

Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. Available from: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 30(15):2114-2120 doi: 10.1093/bioinformatics/btu170.

Corona-Santiago DK, Domínguez-Domínguez O. 2013. Systematik der Hochlandkärplinge . In Kempkes M, Köck M, Stawikouski R, editors: Beiträge zur Biologie und zum Artenschutz der Hochlandkärpflinge. Germany: *Westarp Wissenschaften*; p 13-28.

Domínguez-Domínguez O, Pedraza-Lara C, Gurrola-Sánchez N, Pérez-Rodríguez R, Israde-Alcántara I, Garduño Monroy VH, Doadrio I, Pérez-Ponce De Léon G, Brooks DR. Historical biogeography of the Goodeinae (Cyprinodontiforms). In: Uribe MC, Grier HJ, editors. 2010. Viviparous fishes II. Homestead, Florida: *New Life Publications*; p 13-30.

Domínguez-Domínguez O. 2008a. Filogeografía de *Zoogoneticus quitzeoensis*, *Xenotoca variata* y *Alloophorus robustus* (Cyprinodontiformes: Goodeidae) en el centro de México: Implicaciones taxonómicas y de conservación [Dissertation]. Mexico: Instituto de Biología, Universidad Autónoma de México.

Faircloth BC, Glenn TC. 2012. Not all sequence tags are created equal: Designing and validating sequence identification tags rogust to indels. *PLos ONE*. Available from: http://dx.doi.org/10.1371/journal.pone.0042543.

Fisher S, Barry A, Abreu J, Minie B, Nolan J, Delorey TM, Young G, Fennell TJ, Allen A, Ambrogio L *et al.,* 2011. A scalable, fully automated process for construction of sequence-ready human exome targeted capture libraries. *Genome Biology*. 12(1):R1 doi: 10.1186/gb-2011-12-1-rt.

Fitzsimons JM. 1972. A revision of two genera of Goodeid fishes (Cyprinodontiformes, Osteichthyes) from the Mexican Plateau. *Copeia*. 4:728-756 doi: 10.2307/1442732.

Iwasaki W, Fukunaga T, Isagozawa R, Yamada K, Maeda Y, Satoh TP, Sado T, Mabuchi K, Takeshima H, Miya M, Nishida M. 2013. MitoFish and

MitoAnnotator: A mitochondrial genome database of fish with an accurate and automatic annotation pipeline. *Molecular Biology and Evolution*. 30(11):2531-2540 doi: 10.1093/molbev/mst141.

Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Resources*. 30(14):3059-3066.

Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y. *et al.,* 2012. Soapdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience* 1:18.

Matías-Garcia C. Conducta, conflicto sexual y especiación. In: Martínez-Gómez M, Velázquez-Moctezuma J, editors. 1998. Bases neurobiológicas y ecológicas de la conducta. Mexico: UAT, UAM, UNAM, UV. ISBN: 970-654-229-9.

Moyaho A, Macías-Garcia C, Manjarrez J. 2004. Predation risk is associated with the geographic variation of a sexually selected trait in a viviparous fish, *Xenotoca variata. Journal of Zoology*. 262(3):265-270 doi: 10.1017/S095283690300459X.

Ritchie MG, Hamil RM, Graves JA *et al.,* 2007. Sexual differentiation: population genetic divergence and sexual dimorphism in Mexican goodeid fish. Journal of Evolutionary Biology. 20:2048-2055.

Schattner O, Brooks AN, Lowe TM. 2005. The tRNAscan-SE, snoscan and snoGPS web servers for detection of tRNAS and snoRNAS. *Nucleic Acids Resources*. 33(2):W686-W689 doi: 10.1093/nar/gki366.

Figures

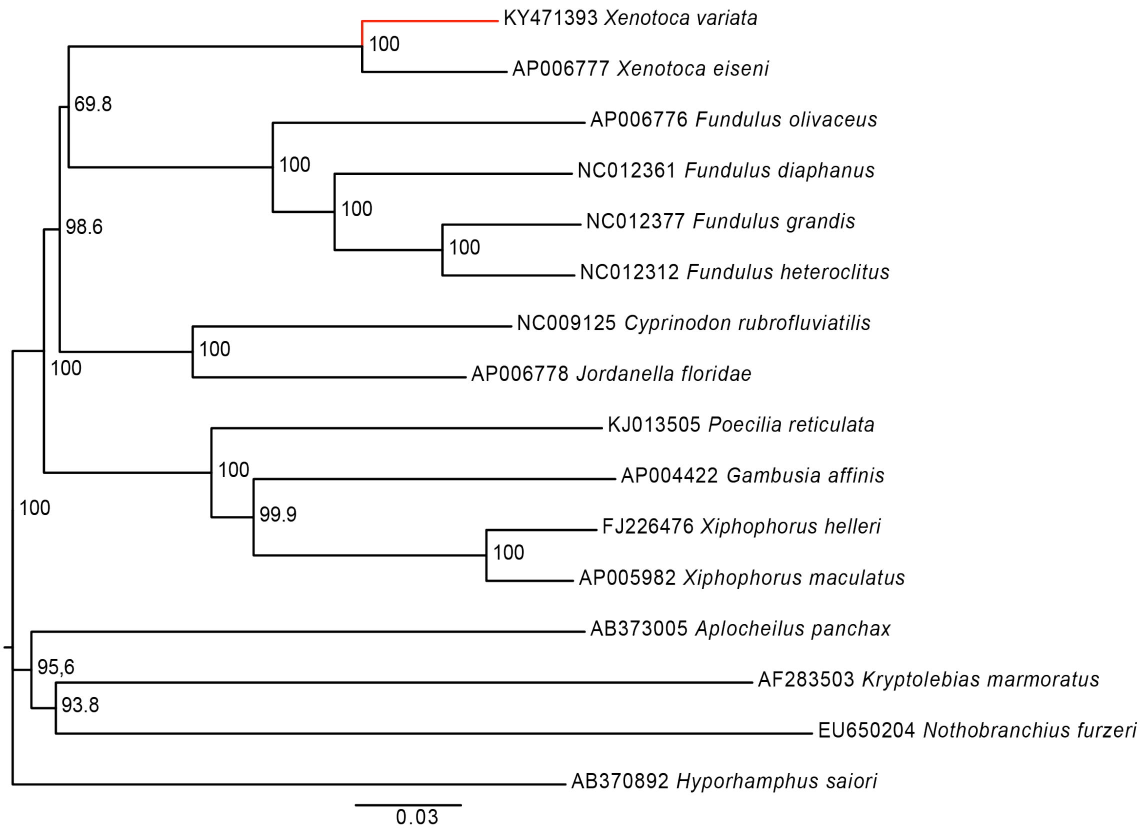


Figure 1. Sample validation based on Neighbor-Joining tree (1000 bootstrap replicated using uncorrected *p*-distances) of 15 species of the Cyprinodontiformes order. In red the position of mitogenome of *Xenotoca variata*.

Discusión general

El desarrollo de la secuenciación Sanger trajo consigo la popularidad de los estudios filogenéticos y filogeográficos aunado a la disponibilidad de oligos para amplificar genes mitocondriales y nucleares de en una extensa variedad de taxa (Hickerson *et al.,* 2010). Para el presente parece obvia la conexión acerca de que los procesos microevolutivos operando dentro de las especies pueden ser extrapolados para explicar diferencias macroevolutivos entre especies y por encima de este nivel taxonómico (Dobzhansky, 1937; Mayr, 1963; Avise *et al.,* 1987; Hickerson *et al.,* 2010). El fundamento para el desarrollo empírico de la filogeografía fue el análisis de marcadores moleculares de origen mitocondrial a un nivel específico (Avise *et al.,* 1987) permitiendo el sometimiento a prueba de hipótesis respecto a las relación causal entre fenómenos geográficos, distribución de especies y mecanismos de especiación. Además, el más sobresaliente y revolucionario aspecto del ADNmt en este campo; además de la ausencia de recombinación, bajo relativa selección neutral, bajo tamaño efectivo poblacional, y consecuentemente un corto tiempo para recuperar monofilias recíprocas entre áreas geográficas, fue el tratado de los segmentos de ANDmt como Unidades Taxonómicas Operacionales OTUS (por sus siglas en inglés) a los individuos de una población o entre poblaciones en un análisis filogenético (Hickerson *et al.,* 2010).

Gracias a la rápida transformación tecnológica, la filogeografía ha sido uno de los campos más integrativos in biología evolutiva, tanto por el desarrollo de diferentes herramientas analíticas, como por la intensificación de la complejidad de las hipótesis evolutivas abordadas. En la presente tesis doctoral, se demostró que el adecuado diseño experimental y la selección de las herramientas moleculares idóneas y su integración, permiten responder a diferentes cuestionamientos en biología evolutiva. En los capítulos 1-4 se integran métodos filogeográficos y evolución molecular junto con modelos de nicho ecológico para explicar procesos históricos y proyecciones a futuro basados en las condiciones y características presentes, como ya ha sido propuesto (Peterson *et al.,* 2002) (Capitulos 1 y 3), la inclusión de análisis de reconstrucción de áreas ancestrales para generar hipótesis biogeográficas (Capítulo 2) como también se ha hecho en otros grupos de la familia Catostomidae (Pérez-Rodríguez *et al.,* 2016), cuantificación de la variación genética intraespecífica para generar información de procesos históricos y actuales con respecto a la dinámica poblacional (Capítulo 1-3) tomando en cuenta la teoría de coalescencia (Hudson, 1983; Kingman, 1982a,b; Tajima, 1983) para generar modelos demográficos históricos (Wakeley, 2008), además de cuantificar la existencia de presiones selectivas (Capitulo 4) para asociarlas a surgimiento de novedades evolutivas en vertebrados como lo es el viviparismo (Consuegra *et al.,* 2015).

Perspectivas generales en la descripción de la historia evolutiva de complejos de especies en México y Europa

De acuerdo a los objetivos propuestos en esta tesis, la descripción de la historia evolutiva de los complejos *Allotoca diazi*, *Catostomus plebeius-nebuliferus* y *Squalius* sp, deja en claro la necesidad de tomar en cuenta cuestiones básicas en biología como es la definición de *especie*, dado que en los tres casos, es difícil ajustar a los linajes identificados genéticamente a cualquiera de las definiciones explicitas o implícitas (Avise y Ball, 1990; De Queiroz, 2007; Templeton, 2001). Los conflictos entre conceptos de especie son observables en los primeros tres capítulos, y aparecen básicamente al definir criterios operativos para identificar estos taxa aún cuando se utilizan conceptos de índole evolutiva.

Bajo estos conceptos, el interés se centra en la identificación de grupos monofiléticos y determinar patrones y procesos de especiación. Dada las dificultades en la aplicación de un concepto de especie, el concepto cohesivo de especie (Templeton 1989) intenta sintetizar los conceptos con un contexto evolutivo, el cual menciona que una especie es la unidad más inclusiva, que presenta cohesión fenotípica y genotípica, mantenida por mecanismos de cohesión como el flujo génico y los factores que determinan el nicho ecológico, es decir, mecanismos que mantienen y promueven el parentesco genético (intercambiabilidad genética) y el uso de un mismo nicho fundamental (intercambiabilidad demográfica o ecológica). Para inferir un suceso de especiación bajo este concepto se deben rechazar estadísticamente tres hipótesis nulas: las poblaciones objeto de estudio forman un único linaje evolutivo, existe intercambiabilidad genética y existe intercambiabilidad ecológica o demográfica (Perfectti, 2002).

Tomando en cuenta el concepto cohesivo de especie se podría decir que el complejo *Allotoca diazi* está conformado por una única especie conservándose la clasificación taxonómica de la primera especie descrita como *Allotoca diazi* (Meek, 1902) y la que posiblemente dio origen de manera natural al menos a *A. meeki*. Sin embargo en otros términos evolutivos, el complejo puede consistir en un sorteo de linajes incompletos (Rosenfel *et al.,* 2012) por la estructura genética encontrada y las divergencias obtenidas entre los miembros del complejo.

A diferencia del complejo *Allotoca diazi*, el complejo *Catostotomus plebeius* presenta una alta estructura y divergencias genéticas entre sus poblaciones basándose en el marcador mitocondrial, cuyas poblaciones fueron aisladas durante el Mioceno Tardío y el Plio-Pleistoceno, suficiente para considerar en la existencia de más de dos linajes evolutivos independientes (Bradley y Baker, 2001) y soportado por otros estudios de la ictiofauna en la región (i.e. Domínguez-Domínguez *et al.,* 2011 y Schönhuth *et al.,* 2014). Sin embargo, las divergencias genéticas encontradas entre estos linajes basadas en el intrón nuclear reflejan un sorteo de linajes incompleto. Aún así, las distancias para el marcador nuclear obtenidas para Catostomidae (Bart *et al.,* 2010) pueden estar sobre estimadas, puesto que es una media estimada para los géneros de la familia, cuya dispersión no está clara entre especies de un mismo género. Lo cual imposibilita el establecimiento del estatus taxonómico de los miembros del complejo y de los posibles linajes evolutivos independientes identificados en este trabajo.

Dado que las aproximaciones filogeográficas permiten identificar eventos de hibridación histórica, zonas de hibridación y ocurrencias de introgresión (Hewitt, 2001; Swenson y Howard, 2005). el reconocimiento y delimitación de *especie* puede complicarse aún más. Un ejemplo de esto, es el caso del complejo de origen híbrido *Squalius* sp abordado en el capítulo 3. La hibridación puede ser una causa de especiación, a través del aislamiento reproductivo del en nuevo linaje de origen híbrido de sus parentales (Baack y Rieseberg, 2007). Sin embargo, el complejo *Squalius* sp no se encuentra aislado de uno de sus parentales y el otro parental ancestral se encuentra actualmente extinto. Lo que sugiere que este complejo se podría encontrar en un proceso intermedio al surgimiento de una nueva especie hasta que una nueva barrera que interrumpa el flujo génico con *S. alburnoides* se establezca. Al poseer características particulares como la hibridogénesis, le permite conservar intacto su genoma ancestral e inclusive obtener individuos no híbridos (nuclearmente) con un genoma mitocondrial perteneciente a *S. alburnoides* como sucede en esta última especie (Cunha *et al.,* 2011). Es por ello que este nuevo complejo es un excelente modelo de estudio para el entendimiento del surgimiento de nuevas especies así como de procesos adaptativos. En este último punto así como la posesión de un genoma de una especie de *Squalius* nunca antes reportado, radica en su importancia de reconocerlo taxonómico y conservarlo. Para la identificación de unidades de conservación es crucial el conocimiento de información genética, biológica y ecológica, así como su reconocimiento como una entidad evolutiva independiente y taxonómicamente reconocida (Doadrio *et al.,* 1996; Crandall *et al.,* 2000; Moritz, 2002; Domínguez-Domínguez *et al.,* 2007), por lo que de acuerdo a que su distribución se encuentra bien definida y que sus poblaciones comparten patrones genéticos únicos (Doadrio *et al.,* 1996), considero que las poblaciones del complejo *Squalius* sp deben ser considerado como unidades operacionales de conservación, con una exigida necesidad de describirlo formalmente tomando en cuenta sus características genéticas únicas (genoma R) y morfológicas reconocidas con anterioridad (Cunha y Doadrio, 2009), y siendo suficientes pruebas para delimitar como especie a este *Squalius* sp como se hizo con *S. alburnoides* (Carmona *et al.,* 2007).

Procesos macroevolutivos y consideraciones para estudios futuros en filogenética y evolución molecular

El estudio de la evolución de la viviparidad en Cyprinodontiformes basado en mitogenomas, permitió determinar y corroborar varias hipótesis como la que explica que la viviparidad tiene como ancestro común el oviparismo, mientras que el ovoviviparismo no es un paso intermedio hacia el viviparismo (Meyer and Lydeard, 1993; Pollux *et al.,* 2009). Este tipo de cuestionamientos evolutivos puede ser abordados a partir de la obtención de una gran cantidad de datos, además de la búsqueda de herramientas analíticas que permitan explicar cuestiones concretas como son las presiones selectivas o cambios estructurales en el genoma mitocondrial asociadas al surgimiento de novedades evolutivas como ocurre también. Además de que permiten aportar nuevas hipótesis evolutivas a grupos bajo controversias sobre sus relaciones filogenéticas (Parenti, 1981; Meyers y Lydeard, 1993; Webb, 1998; Setiamarga *et al.,* 2008; Costa, 2011; Pohl *et al.,* 2015, Hertwig, 2008).

A su vez, el capítulo 4, aportó información acerca de las implicaciones de determinar y seleccionar marcadores moleculares en filogenias a grandes niveles taxonómicos, cuáles son los factores más importantes y qué y cuantos genes son necesario para resolver una filogenia de este tipo. Como se ha descrito, la inclusión de multiples genes para inferencias filogenéticas son esenciales (Comas *et al.,* 2007) pero dependen de su eficiente selección y de tomar en cuenta sus limitaciones (Delsuc, Brinkmann y Philippe, 2005; Feschotte and Pritham, 2007; Philippe *et al.,* 2011; Chan and Ragan, 2013), por lo el capítulo 4 junto con el capítulo 5 y 6 comprenden trabajos que servirán para su aplicación en futuros trabajos filogeográficos y filogenéticos en peces, y ayudará en el diseño de estudios en que la inclusión de gran cantidad de datos o pocos, sea crucial para su desarrollo.

Finalmente, la presente tesis doctoral soporta las bases tradicionales en filogeografía y filogenética a través de la interpretación cuantitativa de cada genealogía y la información referente a la diversidad genética para el estudio de las poblaciones (Hickerson *et al.,* 2010), así como aporta información consistente con trabajos previos sobre temas en macroevolución y evolución molecular (MacCallum y Hill, 2006) sugiriendo que marcadores moleculares pueden ser utilizados a la luz de diversas bases y preguntas evolutivas (Cutter *et al.,* 2013).

Referencias

Avise JC, Arnold J, Ball RM, Bermingham E, Lamb T, Neigel JE et al. 1987. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecology and Systematics*. 18(1), 489-522.

Baack EJ, Rieseberg, LH. 2007. A genomic view of introgression and hybrid speciation. *Current Opinion in Genetics & Development*. 17(6), 513-518.

Ball Jr RM, Neigel JE, Avise JC. 1990. Gene genealogies within the organismal pedigrees of random-mating populations. *Evolution*. 360-370.

Bart HL, Reneau PC, Doosey MH, Bell CD. 2010. Evolutionary divergence of duplicate copies of the growth hormone gene in suckers (Actinopterygii: Catostomidae). *International journal of molecular sciences*. 11(3), 1090-1102.

Bradley RD & Baker RJ. 2001. A test of the genetic species concept: cytochrome-b sequences and mammals. *Journal of Mammalogy*. 82(4): 960–973.

Chan CX, Ragan MA. 2013. Next-generation phylogenomics. *Biology Direct*. 8(1), 3.

Comas I, Moya A, González-Candelas F. 2007. From phylogenetics to phylogenomics: the evolutionary relationships of insect endosymbiotic γ-Proteobacteria as a test case. *Systematic biology*. 56(1), 1-16.

Consuegra S, John E, Verspoor E, De Leaniz CG. 2015. Patterns of natural selection acting on the mitochondrial genome of a locally adapted fish species. *Genetics Selection Evolution*. 47(1), 58.

Crandall KA, Bininda-Emonds OR, Mace GM, Wayne RK. 2000. Considering evolutionary processes in conservation biology. *Trends in Ecology & Evolution*. 15(7), 290-295.

Cunha C, Doadrio I, Abrantes J, Coelho MM. 2011. The evolutionary history of the allopolyploid *Squalius alburnoides* (Cyprinidae) complex in the northern Iberian Peninsula. *Heredity*. 106(1), 100-112.

Cunha C, Doadrio I. 2009. Evidencias de un nuevo complejo híbrido unisexuado del género *Squalius* (Teleostei, Cyprinidae) en el río Estena (Parque Nacional de Cabañeros): una aproximación filogenética. *Proyectos de Investigación en Parques Nacionales*. 357-375.

Cutter AD, Payseur BA. 2013. Genomic signatures of selection at linked sites: unifying the disparity among species. *Nature Reviews Genetics*. 14(4), 262-274.

De Queiroz K. 2007. Species concepts and species delimitation. *Systematic Biology*. 56(6), 879-886.

Delsuc F, Brinkmann H, Philippe H. 2005. Phylogenomics and the reconstruction of the tree of life. *Nature Reviews Genetics*. 6(5), 361-375.

Doadrio I, Perdices A, Machordom A. 1996. Allozymic variation of the endangered killifish *Aphanius iberus* and its application to conservation. *Environmental Biology of Fishes.* 45(3), 259-271.

Dobzhansky T, Dobzhansky, TG. 1937. *Genetics and the Origin of Species* (Vol. 11). Columbia University Press.

Domínguez-Domínguez O, Boto L, Alda F, Pérez-Ponce de León G, Doadrio I. 2007. Human impacts on drainages of the Mesa Central, Mexico, and its genetic effects on an endangered fish, *Zoogoneticus quitzeoensis*. *Conservation Biology*. 21(1), 168-180.

Domínguez-Domínguez O, Vila M, Pérez-Rodríguez R, Remón N, Doadrio I. 2011. Complex evolutionary history of the Mexican stoneroller *Campostoma ornatum* Girard, 1856 (Actinopterygii: Cyprinidae). *BMC evolutionary biology*. 11(1), 153.

Feschotte C, Pritham EJ. 2007. DNA transposons and the evolution of eukaryotic genomes. *Annual Review of Genetics.* 41, 331-368.

Hertwig ST. 2008. Phylogeny of the Cyprinodontiformes (Teleostei, Atherinomorpha): the contribution of cranial soft tissue characters. *Zoologica Scripta*. 37(2), 141-174.

Hewitt GM. 2001. Speciation, hybrid zones and phylogeography—or seeing genes in space and time. *Molecular ecology*. 10(3), 537-549.

Hickerson MJ, Carstens BC, Cavender-Bares J, Crandall KA, Graham CH, Johnson JB et al. 2010. Phylogeography's past, present, and future: 10 years after. *Molecular Phylogenetics and Evolution*. 54(1), 291-301.

Hudson RR. 1983. Properties of a neutral allele model with intragenic recombination. *Theoretical Population Biology*. 23(2), 183-201.

Kingman JFC. 1982a. The coalescent. *Stochastic processes and their applications*. 13(3), 235-248.

Kingman JFC. 1982b. On the genealogy of large populations. *Journal of Applied Probability*. 19(A), 27-43.

MacCallum C, Hill E. 2006. Being positive about selection. *PLoS Biol*. 4(3), e87.

Mayr E. 1963. Animal species and evolution. *Animal species and evolution.*

Meyer A, Lydeard C. 1993. The evolution of copulatory organs, internal fertilization, placentae and viviparity in killifishes (Cyprinodontiformes) inferred from a DNA phylogeny of the tyrosine kinase gene X-src. *Proceedings of the Royal Society of London B: Biological Sciences.* 254(1340), 153-162.

Moritz C. 2002. Strategies to protect biological diversity and the evolutionary processes that sustain it. *Systematic biology*. 51(2), 238-254.

Parenti LR. 1981. A taxonomic revision of the Andean killifish genus *Orestias* (Cyprinodontiformes, Cyprinodontidae). *American Museum of Natural History*.

Pérez-Rodríguez R, Domínguez-Domínguez O, Mar-Silva AF, Doadrio I, Pérez-Ponce de León, G. 2016. The historical biogeography of the southern group of the suker genus *Moxostoma* (Teleostei: Catostomidae) and the colonization of central Mexico. *Zoological Journal de Linnean Society.* DOI: 10.1111/zoj.12383

Perfectti F. 2002. Especiación: Modos y Mecanismos. in Soler M (edr): *Evolución: La base de la biología*. 1–15.

Peterson AT. 2002. Distributional prediction based on ecological niche modeling of primary occurrence data. Á In: Scott, JM et al. (eds), Predicting species occurrences: issues of scale and accuracy.

Philippe H, Brinkmann H, Lavrov DV, Littlewood DTJ, Manuel M, Wörheide G, Baurain D. 2011. Resolving difficult phylogenetic questions: why more sequences are not enough. *PLoS Biol*. 9(3), e1000602.

Pohl M, Milvertz FC, Meyer A, Vences M. 2015. Multigene phylogeny of cyprinodontiform fishes suggests continental radiations and a rogue taxon position of *Pantanodon*. *Vertebrate Zoology*. 65(1), 37-44.

Pollux BJA, Pires MN, Banet AI, Reznick DN. 2009. Evolution of placentas in the fish family Poeciliidae: an empirical study of macroevolution. *Annual Review of Ecology, Evolution and Systematics.* 40. 271-289.

Rosenfeld JA, Payne A & DeSalle R. 2012. Random roots and lineage sorting. *Molecular Phylogenetics and Evolution*. 64: 12–20.

Schönhuth S, Perdices A, Lozano-Vilano L, García-de-León FJ, Espinosa H, Mayden RL. 2014. Phylogenetic relationships of North American western chubs of the genus *Gila* (Cyprinidae, Teleostei), with emphasis on southern species. *Molecular Phylogenetics and Evolution*. 70, 210-230.

Setiamarga DH, Miya M, Yamanoue Y, Mabuchi K, Satoh TP, Inoue JG, Nishida M. 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), 598-605.

Swenson NG, Howard DJ. 2005. Clustering of contact zones, hybrid zones, and phylogeographic breaks in North America. *The American Naturalist*. 166(5), 581-591.

Tajima F. 1983. Evolutionary relationship of DNA sequences in finite populations. *Genetics*. 105(2), 437-460.

Templeton AR. 1989. The meaning of species and speciation: A genetic perspective. *The Units of Evolution*. 159–183.

Templeton AR. 2001. Using phylogeographic analyses of gene trees to test species status and processes. *Molecular Ecology*. 10(3), 779-791.

Wakeley J. 2008. Complex speciation of humans and chimpanzees. *Nature*. 452(7184), E3-E4.

Webb SA. 1998. *A phylogenetic analysis of the Goodeidae (Teleostei: Cyprinodontiformes)* (Doctoral dissertation).

Conclusiones Generales

A raíz de los trabajos realizados en la presente tesis, las siguientes conclusiones se pueden enumerar:

1. Las técnicas moleculares tradicionales y nuevas en estudios integrativos pueden resolver ayudar a resolver preguntar evolutivas tanto generales como concretas.
2. Basado en marcadores moleculares, la historia evolutiva de los tres complejos de especies estudiados está asociada a procesos históricos (geológicos y climáticos) que configuraron su actual distribución.
3. En el caso del complejo *Allotoca diazi*, se determinó a través de microsatélites que parte de su historia evolutiva pudo estar asociada a la actividad humana durante el periodo prehispánico reflejado en sus reducidas divergencias genéticas entre los miembros del complejo y las incongruencias de conectividad entre los cuerpos de agua que habitan actualmente.
4. El complejo *Catostomus plebeius-nebuliferus* presenta una alta estructuración entre sus poblaciones, sugiriendo la presencia de al menos 4 linajes evolutivos independientes que requieren de una profunda revisión taxonómica.
5. El complejo *Squalius* sp representa un nuevo complejo hibridogenético con machos y hembras poliploides, con una distribución muy reducida en el Guadiana Medio de la Península Ibérica, que posee un genoma perteneciente a una especie jamás descrita y extinta que posiblemente se aisló de sus grupos hermanos del sur de Portugal, durante el Plioceno.
6. La viviparidad dentro de los Cyprinodontiformes se origino de manera independiente teniendo como ancestro común a grupos ovíparos.
7. Un número diferencial de genes mitocondriales bajo selección positiva fueron observados durante la evolución de grupos vivíparos (incluyendo ovovivíparos).
8. Los genes que estuvieron bajo con cambios adaptativos significativos pueden estar asociados al origen de la viviparidad en Cyprinodontiformes.

9. Las relaciones filogenéticas basadas en genomas mitocondriales entre las familias del orden Cyprinodontiformes fue bien soportada, y fue posible obtener una inferencia filogenética con valores de soporte altos y con topología idéntica a través de unos pocos genes mitocondriales.
10. La descripción del mitogenoma de *Xenotoca variata* permitió el descubrimiento y prueba de nuevos marcadores moleculares para la aplicación en estudios evolutivos en el futuro, así como para el caso del mitogenoma de *Luciobarbus rifensis* descrito en la presente tesis.