



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
POSGRADO EN CIENCIAS BIOLÓGICAS
INSTITUTO DE BIOLOGÍA, U.N.A.M.
BIOLOGÍA EVOLUTIVA

Reconstrucción filogenómica, estructura genética y flujo génico del complejo
Astyanax mexicanus (De Fillipi 1853) (Actinopterygii, Characidae).

TESIS

QUE PARA OPTAR POR EL GRADO DE:
MAESTRO EN CIENCIAS BIOLÓGICAS

PRESENTA:

GARDUÑO SÁNCHEZ MARCO ANTONIO ALAN

TUTOR PRINCIPAL DE TESIS:

Dra. Claudia Patricia Ornelas García
Instituto de Biología, U.N.A.M

COMITÉ TUTOR:

Dr. Alejandro Zaldívar Riverón
Instituto de Biología, U.N.A.M

Dr. Díaz Jaimes Píndaro

TUTOR INVITADO:

Instituto de Ciencias del Mar y Limnología, U.N.A.M
Ph.D. Joshua Benjamin Gross
University of Cincinnati, EE.UU.

Ciudad Universitaria, CD. MX., Junio, 2021.



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
POSGRADO EN CIENCIAS BIOLÓGICAS
INSTITUTO DE BIOLOGÍA, U.N.A.M.
BIOLOGÍA EVOLUTIVA

Reconstrucción filogenómica, estructura genética y flujo génico del complejo
Astyanax mexicanus (De Fillipi 1853) (Actinopterygii, Characidae).

TESIS

QUE PARA OPTAR POR EL GRADO DE:
MAESTRO EN CIENCIAS BIOLÓGICAS

PRESENTA:

GARDUÑO SÁNCHEZ MARCO ANTONIO ALAN

TUTOR PRINCIPAL DE TESIS:

Dra. Claudia Patricia Ornelas García
Instituto de Biología, U.N.A.M

COMITÉ TUTOR:

Dr. Alejandro Zaldívar Riverón
Instituto de Biología, U.N.A.M

Dr. Díaz Jaimes Píndaro

TUTOR INVITADO:

Instituto de Ciencias del Mar y Limnología, U.N.A.M
Ph.D. Joshua Benjamin Gross
University of Cincinnati, EE.UU.

Ciudad Universitaria, CD. MX., 2021.

COORDINACIÓN DEL POSGRADO EN CIENCIAS BIOLÓGICAS
INSTITUTO DE BIOLOGÍA
OFICIO CPCB/525/2021
ASUNTO: Oficio de Jurado

M. en C. Ivonne Ramírez Wence
Directora General de Administración Escolar, UNAM
P r e s e n t e

Me permito informar a usted que el Comité Académico del Posgrado en Ciencias Biológicas, celebrada el día **30 de noviembre de 2020** se aprobó el siguiente jurado para el examen de grado de **MAESTRO EN CIENCIAS BIOLÓGICAS** en el campo de conocimiento de **BIOLOGÍA EVOLUTIVA** del estudiante **GARDUÑO SÁNCHEZ MARCO ANTONIO ALAN** con número de cuenta **308005424** con la tesis titulada **“Reconstrucción filogenómica, estructura genética y flujo génico del complejo *Astyanax mexicanus* (De Fillipi 1853) (Actinopterygii, Characidae)”**, realizada bajo la dirección de la **DRA. CLAUDIA PATRICIA ORNELAS GARCÍA**, quedando integrado de la siguiente manera:

Presidente: DRA. SUSANA AURORA MAGALLÓN PUEBLA
Vocal: DR. OMAR DOMÍNGUEZ DOMÍNGUEZ
Vocal: DR. MILAN JANDA
Vocal: DRA. ALEJANDRA CITLALLI MORENO LETELIER
Secretario: DR. PÍNDARO DÍAZ JAIMES

Sin otro particular, me es grato enviarle un cordial saludo.

ATENTAMENTE
“POR MI RAZA HABLARÁ EL ESPÍRITU”
Ciudad Universitaria, Cd. Mx., a 03 de junio de 2021

COORDINADOR DEL PROGRAMA



DR. ADOLFO GERARDO NAVARRO SIGÜENZA



Agradecimientos Institucionales

Al Posgrado en Ciencias Biológicas de la Universidad Nacional Autónoma de México por la oportunidad de crecer en mi carrera profesional.

Agradezco el apoyo y financiamiento a esta investigación a través de los proyectos de PAPIIT IN212419, IA203017, y al proyecto CONACYT, ECOS-NORD. De la misma manera agradezco al CONACyT (Consejo Nacional de Ciencia y Tecnología) por haberme otorgado la beca de maestría durante el periodo 2018-2020 y al apoyo brindado por el PAEP (Apoyo a los estudios de Posgrado, U.N.A.M.) para poder llevar a acabo mi estancia académica en la Universidad de Minnesota (University of Minnesota) durante el mes de octubre de 2019. Al Instituto de Supercómputo de Minnesota (Minnesota Supercomputing Institute MSI) de la Universidad de Minnesota por proporcionar los recursos que contribuyeron a los resultados de la investigación informados en esta tesis.

Agradezco a mi tutora principal Dra. Claudia Patricia Ornelas García por su dirección y apoyo durante el desarrollo de este trabajo. A los otros miembros de mi comité tutor, compuestos por el Dr. Alejandro Zaldívar (Instituto de Biología, U.N.A.M.), el Dr. Díaz Jaimes Píndaro (Instituto de Ciencias del Mar y Limnología, U.N.A.M) y al Ph.D. Joshua Benjamin Gross (University of Cincinnati, EE.UU.), por el constante apoyo y su invaluable contribución a este trabajo.

Agradecimientos Personales

A la Universidad Nacional Autónoma de México y al Instituto de Biología, U.N.A.M por brindar los espacios y actividades que han resultado en mi crecimiento profesional, académico y personal.

A los miembros de mi jurado: La Dra. Susana Aurora Magallón Puebla, el Dr. Omar Domínguez Domínguez, el Dr. Píndaro Días Jaimes, el Dr. Milan Janda y la Dra. Alejandra Citlalli Moreno Letellier por sus enriquecedoras contribuciones a este trabajo, y la dedicación que han invertido tanto al desarrollo de sus actividades científicas, como al de sus estudiantes.

A mi tutora la Dra. Claudia Patricia Ornelas García (Paty) por su esfuerzo continuo, incansable y persistente para verme realizado como un profesional de las ciencias biológicas, y por recordarme siempre que los resultados que uno consigue son proporcionales al esfuerzo que uno aplica.

A la Dra. Suzanne McGaugh y la Universidad de Minnesota (University of Minnesota) por haberme dado un lugar en su laboratorio durante mi estancia académica, por el uso de la supercomputadora Mesabi para el análisis de mis datos, por hacerme sentir bienvenido en un país que no era el mío y permitirme aprender de ella y de sus alumnos, Rachel Moran y Jeffrey Miller.

A mis compañeros de laboratorio Jorge Hernandez Garduño, Jazmín Terán Martínez, Carlos Garita Alvarado, Vladimir de Jesus Bonilla, Alberto Hernandez, Victor Hugo Rodríguez Ballesteros, por sus consejos en el manejo y análisis de mis datos, por su trabajo en las salidas de campo y por su amistad incondicional.

A mis padres, Aidé Sánchez Sánchez y Ramón Garduño Lozano por estar conmigo aun cuando no estuviéramos cerca. Por estar ahí en momentos de necesidad y enseñarme que siempre podré contar con ellos.

A mis amigos de Versus, Cesar, Marcos, Gabriel, Jorge, Mario y los demás, por las horas incansables sentados, platicando y jugando Magic the Gathering. Sin esos momentos de diversión probablemente no habría concluido este trabajo.

A mis amigas Sandra Grether, Lucy Vite, Catalina Vázquez, Fer de Alba y Ariana Romero por ser mi red de apoyo durante todo el posgrado.

Por último quiero agradecer a mi persona de máxima confianza, Anayansi Sierralta Gutiérrez. Siempre has sido un apoyo para mi en todo momento, desde lo personal hasta lo profesional. Espero algún día poder ser tan brillante como tu.

Dedicatoria.

Quiero dedicar este trabajo a Александра Асановна Элбакян, por su esfuerzo en eliminar las barreras al acceso de la información científica.



“There should be no obstacles to accessing knowledge... Science should belong to people and not the publishers”.

Александра Асановна Элбакян

Índice

Resumen	1
Abstract	3
Introducción	5
Antecedentes.	11
Objetivos	17
Objetivo General	17
Objetivos Particulares	17
Materiales y Métodos	18
Muestreo y Secuenciación	18
Calidad y Limpieza de las Secuencias	19
Alineamiento, Llamado de Variantes y Genotipado.	21
Inferencia de las Relaciones Filogenéticas	22
Estimación de Flujo Génico y Eventos de Migración en <i>Astyanax</i>	26
Resultados	29
Reconstrucción Filogenética	29
Evaluación de la Estructura Genética en las Poblaciones de <i>Astyanax mexicanus</i> .	34
Resultados del Análisis Discriminante	35
Flujo Génico Reciente y Ancestral	38
Discusión	41
Historia Evolutiva de las Poblaciones de <i>Astyanax mexicanus</i> en la Región de El Abra.	41
Poblaciones de cueva de <i>Astyanax mexicanus</i> clasificadas dentro del grupo del Linaje Viejo	43
Poblaciones de cueva de <i>Astyanax mexicanus</i> clasificadas dentro del grupo del Linaje Nuevo	44
Nuevos Conocimientos Sobre las Relaciones Filogenéticas del Complejo <i>Astyanax mexicanus</i> .	45
Estructura y Flujo Genético de las Poblaciones Troglobias de la Sierra de El Abra	46
Estructura y Flujo Genético de las Poblaciones Troglobias de la Sierra de Guatemala	48
Estructura y Flujo Genético de las Poblaciones Troglobias de la Sierra de la Colmena	50
Incongruencia Filogenética, Evolución Reticular y Flujo Génico Histórico en <i>A. mexicanus</i>	51
Conclusiones	55
Referencias Bibliográficas	56

Resumen

Los cambios fenotípicos, resultado de adaptaciones a la vida en cuevas en algunas especies de peces, ofrecen una gran oportunidad de estudiar la evolución regresiva. Estos animales troglobios nos ofrecen ventajas sobre otros sistemas, ya que de acuerdo a sus grupos hermanos, es posible inferir la dirección del cambio fenotípico (e.g. desde un ancestro de superficie, a las poblaciones de cueva), en respuesta a las presiones selectivas que plantean estos ecosistemas, como son la oscuridad perpetua, la escasez de nutrientes, entre otras. El complejo *Astyanax mexicanus*, distribuido en el noreste de México, se ha convertido en un modelo adecuado para el estudio de la evolución de los caracteres asociados a la vida en cuevas.

Actualmente existen 32 poblaciones troglobias descritas, distribuidas principalmente en la región de La Huasteca en México, en la Sierra de El Abra, la Sierra de Guatemala y la Sierra de la Colmena, en los estados de San Luis Potosí y Tamaulipas. Estas poblaciones exhiben un conjunto de rasgos comunes a otros animales de cuevas, como la reducción o pérdida del sistema visual y la pigmentación. Existe evidencia de múltiples orígenes de las poblaciones de cuevas, sin embargo, el momento y número de las invasiones por parte de las poblaciones de superficie es incierto, así como la conectividad a través de flujo génico, entre poblaciones de cueva y superficie, aún está en debate.

En el presente estudio se llevó a cabo una reconstrucción filogeográfica utilizando SNPs obtenidos a partir de genomas completos, y secuenciación restringida del genoma (Rad-seq), para reconstruir las relaciones entre las poblaciones de cueva y superficie de *A. mexicanus*, así como su estructura genético-poblacional, y estimar los niveles de flujo génico entre estas. Los resultados de este trabajo sugieren que la evolución de las poblaciones de cueva del complejo ha ocurrido en dos linajes distintos y que la evolución a los ambientes de cuevas ha ocurrido en tres eventos distintos en la historia evolutiva del grupo, con una fuerte concordancia con su distribución geográfica. Esto sugiere que la adaptación a la vida en cuevas ha ocurrido de manera repetida en la historia evolutiva del complejo, es decir, de manera polifilética. Los patrones de estructura genética y flujo genético nos

permitieron hacer una reconstrucción del grado de conectividad subterránea entre los sistemas de El Abra, Guatemala y La Colmena. Los resultados muestran la existencia de flujo génico ancestral por parte de poblaciones de superficie hacia las poblaciones de cueva, el cual ha ocurrido en diferentes momentos de la historia evolutiva del grupo, y en diferente intensidad entre poblaciones. Este flujo está determinado en cierta medida de la cercanía geográfica de las poblaciones de superficie con respecto a la entrada de las cuevas (como en el caso de la población de la cueva de Chica).

Abstract

The phenotypic changes observed in cave adapted animals offers a great opportunity to study regressive evolution. These troglobite animals offer advantages over other systems, since we can know the direction of the phenotypic change (e.g. from a surface ancestor to cave populations), in response to strong selective pressures such as perpetual darkness, poor in nutrients, among others. The complex *Astyanax mexicanus*, distributed to the northeast of Mexico, has become a model organism to study the evolution of the characters associated with adaptation to cave environments. Currently there are 34 described troglonian populations, distributed mainly in the La Huasteca region in Mexico, in the Sierra de El Abra, the Sierra de Guatemala and the Sierra de la Colmena, in the states of San Luis Potosí and Tamaulipas. Those populations exhibit a set of traits such as reduction or loss of vision and pigmentation. Despite evidence for the multiple origins of cave populations, the timing of invasions by surface populations is unknown, and gene flow between cave and surface populations is still under debate. In the present study, We carried out a phylogenomic reconstruction using SNPs from whole genomes and restricted genome sequencing (Rad-seq), to reconstruct the relationships between *A. mexicanus* cave and surface populations, as well as its population structure and ancestral and recent gene flow. Our results suggest that the evolution of the cave populations has occurred in two different lineages and that the evolution to cave environments has occurred at three different times in the evolutionary history of *A. mexicanus*, with a strong agreement with its geographical distribution, suggesting that adaptation to cave environments is polyphyletic. The patterns of genetic structure and gene flow allowed us to reconstruct the degree of underground connectivity between the El Abra, Guatemala and La Colmena systems. Also we found the existence of ancestral gene flow from surface populations to cave populations. This gene flow has occurred at different times in the evolutionary history of the group, and is different between different populations. The weight of this gene flow is

determined to some extent by the proximity of the surface populations to the entrance of the caves (as in the case of the Chica cave population).

Introducción

Uno de los principales retos en la biología es entender cómo los organismos pueden adaptarse a los cambios ambientales. La evolución fenotípica, definida como la modificación de los caracteres de un organismo, puede llevar a un proceso adaptativo en respuesta a las variación de las condiciones en el medio (Imam et al., 2020). Esta variación fenotípica puede surgir durante la colonización de nuevos ambientes y en este sentido, la colonización a ambientes extremos representan una gran oportunidad para conocer mejor los procesos de adaptación. Un ejemplo particularmente extendido y bien estudiado de tales adaptaciones puede encontrarse en los animales de cuevas (Krishnan et al., 2017; Bilandžija et al., 2020).

Los organismos que viven en las cuevas se conocen como troglobios. Estos van desde invertebrados como caracoles (Gladstone et al., 2018), gusanos (González et al., 2018), escarabajos (Pacheco et al., 2019), arañas (Mammola et al., 2017), escorpiones (Beron, 2018) y camarones (Chávez-Solís et al., 2018) hasta vertebrados como salamandras (Lunghi et al., 2018) y peces (Keene et al., 20015). La transición a la vida en las cuevas ha llevado a estos organismos a desarrollar un conjunto de rasgos, conocidos como troglomórficos, (Protas et al., 2008; Trontelj et al., 2012; Howarth & Moldovan, 2018; Howarth, 2019), en respuesta a las condiciones ambientales presentes en las cuevas.

Existe una gran variedad de rasgos troglomórficos compartidos entre los organismos de cuevas (Culver et al., 1995; Christiansen 2012; Trontelj et al., 2012; Krishnan et al., 2017; Howarth & Moldovan, 2018), pero los más comunes son la pérdida de ojos, la alteración de los ciclos circadianos y la reducción de la pigmentación corporal. Sin embargo, muchos también han desarrollado una variedad de rasgos, llamados "rasgos constructivos", como órganos sensoriales alargados, papilas gustativas adicionales, aumento en el número de neuromastos, entre otros (Yoshizawa et al. 2010; Krishnan et al., 2017; Howarth & Moldovan, 2018). La adquisición de los rasgos troglomórficos suele atribuirse a los

procesos de evolución en paralelo, o evolución convergente; ya sea por eventos de selección natural, o a través de la acumulación de mutaciones y deriva génica (Jeffery, 2009).

Los animales troglobios ofrecen ciertas ventajas para estudiar los procesos adaptativos. Primero, la falta de luz suele ser una constante durante la historia evolutiva de los organismos de cueva, y conduce a la pérdida de los sistemas visuales y la pigmentación. (Culver et al. 2010; Derkarabetian et al. 2010). Por lo tanto, los rasgos troglomórficos pueden relacionarse directamente con las condiciones ambientales (Jeffery, 2009; Howarth & Moldovan, 2018). En segundo lugar, el marco de tiempo evolutivo en el que se presentan los rasgos troglomórficos puede ser relativamente corto (Bilandžija et al., 2020), lo que permite identificar y estudiar los primeros cambios que conducen a la adaptación a las cuevas (Jeffery, 2009). En tercer lugar, cada cueva es un único experimento natural, por lo que ofrecen la oportunidad de estudiar los mecanismos de evolución convergente y paralela de los rasgos troglobios (Jeffery, 2009).

En los últimos años se han producido grandes avances con respecto a la comprensión de cómo evolucionan las especies dentro de los ambientes de cuevas, gracias a las investigaciones realizadas sobre algunas especies troglobias, como el anfípodo *Gammarus minus*, el crustáceo *Asellus aquaticus* (Protas et al., 2011; Gross et al., 2015), o la sardinita mexicana *Astyanax mexicanus* (Keene et al., 2015).

La sardinita *Astyanax mexicanus* (De Fillipi 1853) es una especie que pertenece a la familia Characidae, dentro del orden de los Characiformes (Wiley y Mitchell, 1971; Jeffery, 2009; Miller et al., 2009). Su distribución abarca desde la cuenca del río Pánuco, hasta la cuenca del río Bravo en la vertiente Atlántica, y las cuencas del río Nazas-Aguanaval y Mezquital en la vertiente Pacífica, (Ornelas-García et al., 2008; Ornelas-García y Pedraza-Lara, 2016), donde se le puede encontrar habitando fondos rocosos y arenosos de estanques y remansos de arroyos y ríos pequeños a grandes y manantiales (Mercado-Silva et al., 2020). Para esta especie, se han reportado dos morfotipos: uno

presente en ambientes de superficie, como en ríos, arroyos y lagos; y otro adaptado a la vida en las cuevas (Miller et al., 2009).

Las poblaciones de cuevas de *Astyanax mexicanus* habitan principalmente en la región de la Huasteca, al noreste de México (Ornelas-García et al., 2016). Ésta se encuentra ubicada en la región de la Sierra Madre Oriental, en el noreste de México, en el límites entre los estados de San Luis Potosí y Tamaulipas. Esta región está compuesta por una serie compleja de facies depositadas en bancos de conchas, lavados por las olas y arrecifes de rudistas, separados por áreas entre arrecifes donde se depositaron lodos calcáreos (Elliot, 2019). De este a oeste, hay una serie de cordilleras: La Sierra de El Abra, La Sierra de Guatemala y La Sierra de la Colmena. Esta última presenta unos pequeños cerros a los que la comunidad científica de *Astyanax* se ha referido como la región de Micos (Espinasa y Espinasa, 2016).

Las primeras poblaciones de cueva, encontradas en la Sierra de El Abra, fueron descritas como tres especies diferentes (Jeffery, 2008). La primera población descubierta fue en la cueva de Chica, y fue asignada a una nueva especie, con un nuevo género: *Anoptichthys jordani* (Hubbs e Innes, 1936). La segunda población encontrada fue en la cueva El Pachón, y fue nombrada como *Anoptichthys antrobius* (Alvarez, 1946), y la tercera fue encontrada en la cueva de Los Sabinos, y nombrada como *Anoptichthys hubbsi* (Alvarez, 1947). Con el paso del tiempo, el género *Anoptichthys* fue sinonimizado con *Astyanax*, quedando las especies de cueva y de superficie como morfotipos distintos (Hubbs e Innes, 1936; Schemmel 1967), bajo un concepto biológico de especie (Gross et al., 2015).

En la década de los 70 's, un gran número de poblaciones de cueva de *A. mexicanus* empezaron a ser descritas y exploradas. Mitchell y colaboradores (1977) describieron 29 poblaciones de cuevas en la Sierra de El Abra, La Sierra de Guatemala y el área de Micos. Posteriormente, fueron descritas las localidades de cuevas Granadas y La Joya, en Guerrero, al Sur de México (Espinasa et al., 2001; Espinasa et al., 2018), para la especie hermana *Astyanax aeneus* (Günther, 1860). En los últimos años, se descubrieron las localidades de Cueva Chiquitita y el Sótano de Toro 2, al sur de la Sierra de El Abra

(Espinasa et al., 2018; Espinasa et al., 2020), y al norte de la Sierra de el Abra, el Sótano de La Calera (Espinasa et al., 2020). Actualmente se conocen 32 poblaciones troglobias descritas para la región de La Huasteca, al noreste de México (fig. 1, Mitchell et al., 1977; Jeffery, 2008; Elliot, 2018; Espinasa et al., 2018; Espinasa et al., 2020).

La mayoría de las cuevas habitadas por estos peces comparten una serie de condiciones ambientales, como la oscuridad constante y la ausencia de depredadores, asociados a la vida en las cuevas (Jeffery, 2009; Borowsky, 2018). Las modificaciones morfológicas y fisiológicas en los organismos troglobios de *Astyanax mexicanus* incluyen: pérdida total o parcial de las estructuras visuales, reducción o pérdida de la pigmentación, (Jeffery y Martasian 1998; Jeffery 2001; Strecker et al., 2003; Jeffery et al., 2009), aumento en el almacenamiento de lípidos (Rose y Mitchell, 1982), disminución del consumo de oxígeno (Hüppop 1986), alteraciones del ciclo circadiano (Erckens y Martin, 1982; Erckens y Martin, 1982b, Borowsky, 2018), y un sentido olfativo extremadamente sensible (Quinn, 1980). Aunque los fenotipos entre diferentes poblaciones de cuevas son similares, estos no son idénticos. Por ejemplo, es posible observar morfologías cráneo-faciales distintas entre diferentes poblaciones de cueva (Jeffery, 2009), así como modificaciones en el eje antero-posterior (Borowsky, 2018); diferencias en el número de vértebras (Dowling et al., 2002), entre otros caracteres (ver Jeffery, 2009).

Una ventaja de este sistema sobre otros organismos de cueva, es que las formas de superficie de *A. mexicanus* siguen siendo nativas de los sistemas fluviales circundantes (Krishnan et al., 2017), lo cual no es tan común en otras especies de cueva (Howart, 2019). Aunque geográficamente aislados, los peces de cueva siguen siendo interfértiles con la forma de superficie, lo que permite el análisis genético de la herencia de los rasgos específicos de la cueva y de la superficie, y entre otras poblaciones de cueva, lo que permite evaluar la complementación genética (Protas et al., 2006; Protas et al., 2008; Kowalko et al., 2013; Krishnan et al., 2017).

En las últimas décadas se han realizado grandes esfuerzos para esclarecer las relaciones de ancestría entre las poblaciones de cuevas y superficies de *Astyanax mexicanus* al noreste de México. Trabajos previos han sugerido que las poblaciones de cuevas fueron colonizadas en al menos dos episodios distintos a partir de peces de superficie (Dowling et al., 2002; Strecker et al., 2004; Ornelas-García et al., 2008; Jeffery, 2009; Hausdorf et al., 2011; Bradic et al., 2012; Strecker et al., 2012; Bradic et al., 2013). El primer episodio de colonización habría dado lugar a poblaciones de cueva de la Sierra de El Abra, designadas como el linaje viejo, mientras que las cuevas de la región occidental de Micos y la Sierra de Guatemala habrían sido colonizadas de manera más reciente, dando lugar al conjunto de poblaciones de cueva conocido como el linaje nuevo (Strecker et al., 2004; Hausdorf et al., 2011; Bradic et al., 2012). Sin embargo, existe cierta controversia con respecto al origen, tiempo de divergencia y relación entre las poblaciones de cueva y de superficie de *Astyanax mexicanus*.

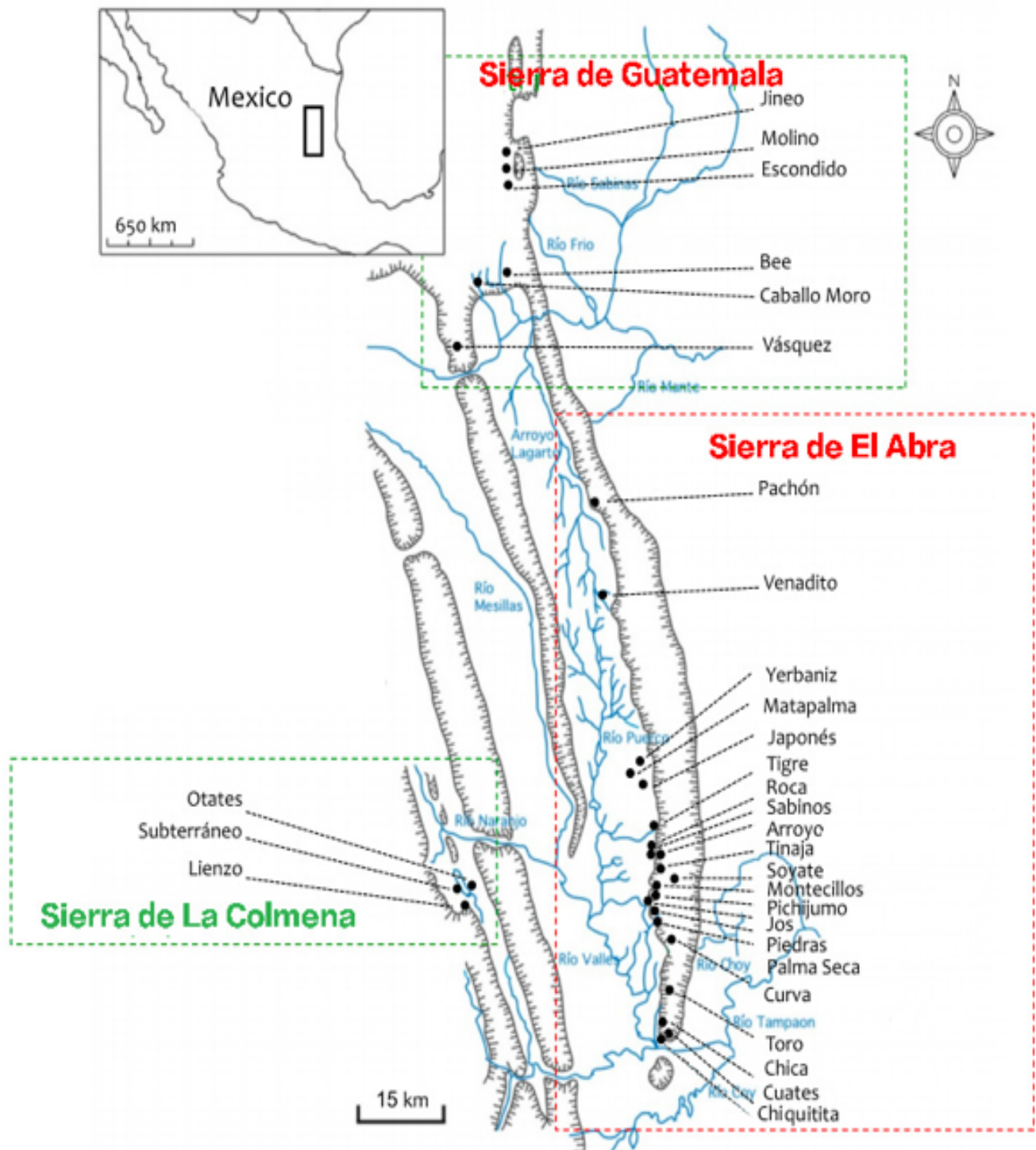


Figura 1. Mapa de las poblaciones de cuevas de *A. mexicanus* descritas en la región de La Huasteca, en la Sierra de El Abra, La Sierra de Guatemala y La Sierra de la Colmena, en los estados de San Luis Potosí y Tamaulipas, al noreste de México. Modificado de Gross (2012).

Antecedentes.

Los primeros estudios sobre la evolución de las poblaciones de cueva de *Astyanax mexicanus* fue realizado por Avise y Selander (1972), quienes estimaron el grado de variabilidad genética entre poblaciones de cueva y superficie para determinar si las poblaciones de cueva eran menos diversas genéticamente que las poblaciones de superficie. En este trabajo se reportó que los niveles de heterocigosidad entre las poblaciones de cueva eran bajos (heterocigosidad promedio = $\sim 7.7\%$) en comparación con las de superficie analizadas (heterocigosidad promedio = $\sim 11.2\%$). Dado el número de individuos que conforman las poblaciones de cueva, la reducción observada en la variabilidad en las poblaciones de las cuevas de *Astyanax* se atribuyó en gran medida a procesos estocásticos.

Adicionalmente, Avise y Selander (1972) consideraron que la aparición de los fenotipos asociados a los ambientes de cueva (i.e. pérdida de pigmentación y de las estructuras visuales) ocurrió previo a la subdivisión actual de estas poblaciones, por lo que puede decirse que apoyaban la teoría de un sólo origen para las poblaciones troglobias (Gross, 2012), el cual que pudo ocurrido, en el Pleistoceno tardío, en un intervalo de 10 mil hasta 2 millones de años (Avise y Selander, 1972).

Por su parte, Wilkens (1971), prueba que las bases genéticas de la pérdida de las estructuras visuales es diferente entre estas poblaciones de cueva, aunque en ambas está involucrado el mismo número de genes. Wilkens sugiere que los patrones de estructura poblacional observados y la disminución de la variabilidad genética en las poblaciones de cuevas son resultado de la distribución discontinua del hábitat acuático (Wilkens, 1971).

Wilkens (1988) explica el grado de regresión evolutiva en los caracteres troglobios a través de la acción conjunta de los procesos de la mutación y de deriva génica en las poblaciones de cueva, y considera a las poblaciones de la Sierra de El Abra (particularmente las poblaciones de Pachón, Sabinos, Piedras, Yerbaniz y Pichijumo), dentro del grupo de cuevas filogenéticamente viejas, debido al alto grado de regresión evolutiva en caracteres troglobios.

Dowling y colaboradores (2002) sugieren la hipótesis de múltiples orígenes para las distintas poblaciones de cueva, en por lo menos dos invasiones independientes (Dowling et al., 2002; Gross, 2012). En este trabajo, la diversidad genética fue evaluada a través de los polimorfismos del gen mitocondrial ND2 en 17 individuos de 12 poblaciones (6 de cueva y 11 de superficie). Dowling y colaboradores (2002) lograron identificar dos linajes distintos, sugiriendo que la evolución del morfotipo troglobio ha ocurrido en más de una ocasión a lo largo de su historia evolutiva.

Strecker y colaboradores (2003) evaluaron la diversidad genética y las relaciones de parentesco entre cuatro poblaciones de cuevas (Pachón, Sabinos, Tinaja, Chica) y cuatro poblaciones de superficie de *Astyanax*. Encontraron que la diversidad genética es altamente variable entre poblaciones, pero significativamente baja en casi todas las poblaciones de las cuevas, en comparación con poblaciones de la superficie (con excepción de la cueva Chica). Este resultado fue congruente con (Avice y Selander 1972), así como secuencias de ADNmt (Dowling et al., 2002), y sugiere un período relativamente largo de separación entre las poblaciones de la cueva y la superficie (Strecker et al., 2003).

Adicionalmente,, encontraron que las poblaciones de cueva de Pachón, Sabinos y Tinaja se agruparon entre sí, en lugar de con los peces de la superficie de la misma área, por lo que se propuso que las poblaciones de las cuevas se originaron en un grupo ancestral diferente de las poblaciones contemporáneas de superficie (Strecker et al., 2003). Los resultados en este trabajo sugieren al menos dos invasiones independientes a las cuevas, a saber, hacia la población de cueva Chica y hacia las poblaciones de cuevas de Sabinos, Tinaja y Pachon (Strecker et al., 2003).

En un trabajo posterior, Strecker y colaboradores (2004) evaluaron las relaciones filogeográficas de 26 poblaciones de superficie, distribuidas en México, Guatemala y Belice y 9 poblaciones de cueva al noreste de México. Para este trabajo, las poblaciones de cueva fueron divididas en dos grupos distintos con base en su morfología. Por un lado, el grupo SEP (“strongly reduced eyes and pigment”), cuyas poblaciones (i.e. Piedras, Sabinos o Tinaja) presentan una fuerte reducción en la pigmentación y en la complejidad de las estructuras visuales, y por otro lado, el grupo VEP (“variable

eye size and pigmentation”), cuyas poblaciones (i.e Chica, Molino, Pachón o Yerbaniz) presentan un tamaño de ojos y pigmentación variables.

Los análisis filogenéticos revelaron que las poblaciones de cueva y superficie se agrupan en siete linajes mitocondriales distintos. Se encontró que las poblaciones de cuevas clasificadas como SEP, forman un linaje mitocondrial independiente, mientras que las poblaciones de cueva clasificadas como VEP, se agrupan junto con poblaciones de superficie geográficamente cercanas. Lo anterior se ha usado como evidencia de múltiples eventos de invasión (al menos dos) a las cuevas por parte de peces de superficie.

Estos autores discuten la posibilidad de que las poblaciones de cueva SEP hayan derivado de la primera invasión, debido al grado de troglomorfia entre estas poblaciones. Por otro lado, se encontró cierto grado de inconsistencia en cuanto a la posición filogenética de algunas poblaciones de cueva VEP por la presencia de fenotipos intermedios. En este sentido, las poblaciones de cueva de Otates, Subterráneo y Lienzo, ubicadas en la Sierra de La Colmena, fueron agrupadas dentro de la categoría VEP, junto con la población de Chica, Molino y Pachón (Wilkens, 1988; Strecker et al., 2004), por la presencia de fenotipos intermedios.

En el trabajo realizado por Ornelas-García et al. (2008), al evaluar las relaciones filogenéticas en 141 poblaciones del género *Astyanax*, usando 3 marcadores mitocondriales y 1 marcador nuclear, se recuperaron a las poblaciones de cueva en dos linajes distintos, de forma similar a los trabajos de Strecker et al. (2003, 2004). El primer linaje (Linaje Ia), incluyó a las poblaciones de superficie de *A. mexicanus* distribuidas en las cuencas del Río Bravo, Conchos, Aguanaval-Mezquital, Pánuco y Tuxpan. Estas se agruparon con las poblaciones de cuevas de las regiones de Micos (cueva de Subterráneo); El Abra: (Cuevas Chica, Pachón, Yerbaniz y Japonés) y Guatemala (cueva de Molino), que previamente habían sido descritas como “linaje nuevo” (agrupadas en VEP, *sensu* Strecker et al. 2004). En el segundo linaje (Lineage Ie), fueron agrupadas las poblaciones de cuevas de “linaje viejo” de Curva, Tinaja, Sabinos y Piedras de la Sierra de El Abra (grupo SEP, *sensu* Strecker et al. 2004),

junto con las poblaciones de superficie de Rascón y Aguanaval - Mezquital. Una de las principales contribuciones de este trabajo fue identificar a las poblaciones de Rascón y Aguanaval - Mezquital como las poblaciones de superficie del linaje viejo, las cuales se creían extintas en estudios previos (Strecker et al. 2004).

Posteriormente, Bradic, et al. (2012), propusieron al menos 5 eventos independientes de colonización a las cuevas, los cuales están a su vez relacionados con los linajes viejo y nuevo. Este trabajo sugiere que los primeros eventos de colonización a las cuevas de la Sierra de El Abra, corresponde con el linaje viejo. Mientras que , la colonización de las cuevas en La Sierra de Guatemala y La Sierra de Micos corresponde con el linaje nuevo (Bradic et al., 2012).

En un estudio posterior, Strecker y col. (2012) evaluaron los patrones de estructura y flujo génico de 7 poblaciones de superficie y 7 de cuevas,. En este trabajo se encontró que la población de Caballo Moro, en la Sierra de Guatemala, forma un grupo genético distinto al resto de las poblaciones de la Sierra de El Abra y genéticamente más cercano a las poblaciones del sur del Cinturón Volcánico Trans-Mexicano (Trans-Mexican Volcanic Belt o TMVB) que a las poblaciones adyacentes del norte de México. Strecker et al., (2012) proponen la colonización de la cueva de Caballo Moro como un tercer evento de invasión, indicando que las poblaciones de superficie de *Astyanax* cruzando el TMVB, al menos en tres ocasiones. En este trabajo encontraron evidencia de flujo génico entre poblaciones de superficie y poblaciones de cueva de la Sierra de El Abra. Sin embargo, describen a este como un fenómeno raro, ya que las cuevas imponen condiciones extremas a la supervivencia de los organismos superficiales.

Coghill et al. (2014) pusieron a prueba el origen monofilético de las poblaciones de cuevas, estimando la historia evolutiva de las poblaciones de cueva y de superficie de *A. mexicanus*, utilizando secuenciación por enzimas de restricción (Rad-seq). restringiendo su reconstrucción bajo el contexto de la deriva génica establecida por Wilkens (1998), contra otra reconstrucción sin restricciones. El restringir la reconstrucción filogenética basada en el grado de troglomorfa produjo un árbol con bajos

soportes, en comparación con el árbol sin restricción. Tradicionalmente el grado de troglomorfia había sido utilizado para asumir cierto tiempo desde la colonización. Sin embargo, Coghill y colaboradores (2014) encontraron que la adaptación fenotípica a las cuevas ocurre relativamente rápido en tiempos evolutivos y que el grado de regresión no es un indicador robusto para clasificar los eventos de colonización en las poblaciones troglobias (Coghill et al., 2014).

Fumey y colaboradores (2018) analizaron secuencias de transcriptomas de las poblaciones de cueva Pachón, Tinaja y Molino, así como marcadores mitocondriales y nucleares previamente publicados, buscando verificar la existencia de los dos linajes de *Astyanax mexicanus*. En este trabajo, se encontró un alto grado de incongruencia en las historias evolutivas recuperadas entre los distintos marcadores empleados. Adicionalmente, utilizando simulaciones de coalescencia, encontraron que la probabilidad máxima del tiempo de divergencia se encontraba dentro del rango de 1,000 a 10,000 años. Estos resultados sugieren un origen reciente para las poblaciones de cueva, independientemente de su clasificación como viejos, nuevos, aislados o híbridos (Fumey et al., 2018).

Finalmente, Herman y colaboradores (2018) secuenciaron 47 genomas completos, provenientes de 3 poblaciones de cuevas (Molino, Tinaja y Pachón) y 2 de superficie (Rascón y Río Choy) de *Astyanax mexicanus* e infirieron las relaciones filogenéticas poblacionales, utilizando la especie hermana *A. aeneus* de la población de la cueva de Granadas, en el estado de Guerrero, México como un grupo externo. Encontraron que las poblaciones de cuevas son polifiléticas, distribuyéndose en 2 linajes independientes: por un lado, el clado monofilético conformado por la población de superficie de Rascón y las poblaciones de cueva de Tinaja y Pachón, reconocido en trabajos previos como el “linaje viejo”, y por el otro el clado monofilético conformado por las poblaciones de superficie de Río Choy y Sótano de Molino, reconocido en la literatura como el “linaje nuevo”. Adicionalmente, estimaron el posible tiempo de divergencia entre estos linajes utilizando un valor de divergencia basada en dXY y modelos demográficos y sugieren que la división del linaje "viejo" y "nuevo" es más reciente de lo que se había propuesto antes (entre 3 y 12 Mda), previo al cierre del Eje Neovolcánico Transversal. Es

posible que probablemente esta barrera geográfica haya sido traspasada por *Astyanax* en varias ocasiones, provocado invasiones a las poblaciones de cuevas en cada evento de migración (Hausdorf et al., 2011; Gross, 2012; Strecker et al., 2012).

A pesar de la repetida evidencia acerca de la invasión a las cuevas en dos linajes distintos, aún existe controversia con respecto al origen y relaciones filogenéticas entre las poblaciones de cueva y de superficie. Por lo tanto, en el presente estudio se propone realizar una reconstrucción filogenómica de las diferentes poblaciones de cuevas y superficie de *Astyanax mexicanus* utilizando el muestreo más exhaustivo hasta la fecha, con el fin de proponer una hipótesis filogenética robusta sobre la historia evolutiva de las poblaciones de cuevas de *Astyanax mexicanus* al norte de México.

Objetivos

Objetivo General

Construir una hipótesis filogeográfica de las poblaciones de cueva y superficie de *Astyanax mexicanus* (De Fillipi 1853), basada en SNPs de genomas completos y secuenciación por sitios de restricción (Rad-seq), así como evaluar la estructura y flujo entre sus poblaciones.

Objetivos Particulares

- Construir una hipótesis filogenómica de las poblaciones de cueva y superficie de *Astyanax mexicanus* basada en SNPs de genomas completos y secuenciación por sitios de restricción (Rad-seq).
- Caracterizar la estructura genética y la entremezcla entre poblaciones de *Astyanax mexicanus*.
- Evaluar el papel del flujo génico ancestral y reciente, para entender los patrones de diversificación de las poblaciones de *Astyanax mexicanus* en el norte de México.

Materiales y Métodos

Muestreo y Secuenciación

En este estudio se colectaron 122 individuos de *Astyanax mexicanus* provenientes de 20 poblaciones de cueva de la Sierra de El Abra, Sierra de Guatemala y Sierra de Micos, 9 poblaciones de superficie de las regiones cercanas y 5 localidades de superficie del norte de México. Las localidades del norte de México fueron incluidas para una mejor representatividad geográfica, en la reconstrucción filogenómica de las cuevas y superficie. Dado que existe cierta controversia sobre cuál es el grupo hermano de las poblaciones de *Astyanax mexicanus* (Ornelas-García et al., 2008; Coghill et al., 2014; Herman et al., 2018), se utilizaron 2 linajes como grupos externos: 2 muestras del linaje ID (Costa de Chiapas) de Ornelas-García et al., (2008) y 2 muestras del linaje de *Astyanax aeneus* de la cuenca del río Balsas (Herman et al., 2018). Para la colecta de los ejemplares de cueva se contó con el permiso por parte de las autoridades competentes (SEMARNAT SGPA/DGVS/2438/16). Los especímenes fueron catalogados y depositados en la Colección Nacional de Peces, del Instituto de Biología, U.N.A.M.

Se utilizaron dos estrategias de secuenciación: secuenciación asociada a sitios de restricción (RADseq) y secuenciación de genomas completos (Whole Genome Sequencing o WGS). Para la secuenciación de los genomas completos, 78 muestras fueron ultracongeladas con nitrógeno líquido, siguiendo el protocolo de “flash frozen” (Passow et al., 2018), y almacenadas a -72°C. Posteriormente, se prepararon para la secuenciación mediante la adición de códigos de barras genéticos individuales por cada muestra y la creación de las librerías utilizando el kit TruSeq Nano de Illumina. Las muestras fueron secuenciadas utilizando la plataforma de Illumina NovaSeq 6000, con un tamaño de lectura promedio de 150 pb y un rendimiento promedio de 4 millones de lectura por muestra. Para la secuenciación por el método de RADseq, se extrajo el ADN genómico de 37 muestras. La extracción de ADN se realizó mediante el método de extracción de cloruro de sodio (NaCl) 4,5 M (Sonnenberg et al. 2007) y se cuantificó utilizando el espectrofotómetro NANODROP 2000 (Thermo). La selección de la enzima de digestión se basó en simulaciones de digestiones *in silico* del genoma de *Astyanax*

mexicanus V. 2.0 (McGaugh et al., 2014), utilizando la paquetería en R Sim-RAD (Lepais et al., 2014), con diferentes enzimas de restricción. Las muestras de DNA extraídas con el método de cloruro de sodio fueron digeridas usando la enzima de restricción MspI y preparadas para la secuenciación de Illumina mediante la codificación de barras individuales, y procesadas con el kit de preparación de muestras de ADN TruSeq Nano de Illumina, utilizando los reactivos v3. Las muestras fueron multiplexadas y secuenciadas en la plataforma Illumina NextSeq, secuenciando un solo extremo, con 150 pb de longitud promedio por lectura, obteniendo un rendimiento promedio de 4 M lecturas por muestra. la preparación de las bibliotecas, así como la secuenciación fue llevada a cabo por el Servicio de Secuenciación del Centro Genómico de la Universidad de Minnesota (<https://genomics.umn.edu>). Adicionalmente, se incluyeron en este estudio 30 genomas completos de *Astyanax mexicanus* previamente secuenciados (Herman et al., 2008), para un total de 152 muestras totales (Tabla 1, Fig. 2 B).

Calidad y Limpieza de las Secuencias

Para la limpieza de los datos se utilizó el programa FastQC (Andrews, 2010), tanto para las secuencias de los genomas completos, cómo para las de RADseq. Dependiendo del estado de la calidad de las muestras, las secuencias se depuraron y limpiaron con el programa trimmomatic V.0.39 (Bolger et al., 2014). Se utilizó un enfoque de recorte de ventana deslizante, que corta las bases dentro de la ventana de tamaño una vez que la calidad promedio cae por debajo de un umbral establecido. El tamaño de la ventana seleccionado fue de 4 bases y el límite inferior de calidad fue de 15 en la escala Phred 33. También se buscaron y eliminaron los adaptadores utilizados para el proceso de secuenciación, y se cortaron las primeras 14 bases de todas las lecturas independientemente de su calidad. Al final, solo se conservaron las lecturas que tuvieran una longitud mínima de 75 bases.

	Localidad	Área geográfica	Clado	Sistema Hidrológico	Ambiente	N
1	Sótano de Jineo	Sierra de Guatemala	1D	Gomes Farías	Cueva	5
2	Sótano de Molino	Sierra de Guatemala	1D	Gomes Farías	Cueva	5
3	Sótano Escondido	Sierra de Guatemala	1D	Gomes Farías	Cueva	5
4	Sótano del Caballo Moro	Sierra de Guatemala	1E	Chamal - Ocampo	Cueva	3
5	Sótano de Vásquez	Sierra de Guatemala	1E	Chamal - Ocampo	Cueva	5
6	Cueva de El Pachón	Sierra de El Abra	2A	Norte de El Abra	Cueva	5
7	Sótano de Yerbaniz	Sierra de El Abra	2B	Sistema Yerbaniz	Cueva	7
8	Sótano de Japonés	Sierra de El Abra	2B	Sistema Yerbaniz	Cueva	4
9	Sótano del Tigre	Sierra de El Abra	2C	Sistema Los Sabinos	Cueva	5
10	Cueva de Los Sabinos	Sierra de El Abra	2C	Sistema Los Sabinos	Cueva	2
11	Sótano del Arroyo	Sierra de El Abra	2C	Sistema Los Sabinos	Cueva	5
12	Sótano de la Tinaja	Sierra de El Abra	2C	Sistema Los Sabinos	Cueva	5
13	Sótano de Montecillos	Sierra de El Abra	2C	Sistema Los Sabinos	Cueva	4
14	Sótano de Jos	Sierra de El Abra	2C	Sistema Los Sabinos	Cueva	5
15	Sótano de Las Piedras	Sierra de El Abra	2C	Sistema Los Sabinos	Cueva	1
16	Sótano de la Palma Seca	Sierra de El Abra	2C	Sistema Los Sabinos	Cueva	5
17	Sótano del Toro	Sierra de El Abra	2D	Región Sur del Abra	Cueva	3
18	Cueva de Chica	Sierra de El Abra	2D	Región Sur del Abra	Cueva	5
19	Cueva de Chiquitita	Sierra de El Abra	2D	Región Sur del Abra	Cueva	2
20	Cueva del Río Subterráneo	Sierra de La Colmena	1C	Región de Micos	Cueva	5
21	Río Coy	Río Coy	1B	Cuenca del Río Pánuco	Superficie	6
22	Río Choy	Río Choy	1B	Cuenca del Río Pánuco	Superficie	6
23	Río Mante	Río Mante	1B	Cuenca del Río Pánuco	Superficie	10
24	Río Sabinas	Río Sabinas	1B	Cuenca del Río Pánuco	Superficie	4

25	Bravo	Bravo	1A	Cuenca Bravo - Conchos	Superficie	1
26	Conchos	Conchos	1A	Cuenca Bravo - Conchos	Superficie	5
27	Texas	Texas	1A	Cuenca Bravo - Conchos	Superficie	1
28	San Fernando	San Fernando	1A	Cuenca Bravo - Conchos	Superficie	2
29	Hualahuises	Hualahuises	1A	Cuenca Bravo - Conchos	Superficie	3
30	Santa Anita	Santa Anita	1F	Cuenca del Río Pánuco	Superficie	2
31	Río Rascón	Río Rascón	2E	Cuenca del Río Pánuco	Superficie	6
32	Río Gallinas	Río Gallinas	2E	Cuenca del Río Pánuco	Superficie	3
33	Tamasopo	Tamasopo	2E	Cuenca del Río Pánuco	Superficie	6
34	Río Peroles	Río Peroles	2E	Cuenca del Río Pánuco	Superficie	5
35	<i>Astyanax aeneus</i>	Cuenca del Río Balsas	Grupo Externo	Cuenca del Río Balsas	Superficie	2
36	<i>Astyanax sp.</i>	Costa de Chiapas	Grupo Externo	Costa de Chiapas	Superficie	2

Tabla 1 Número de ejemplares de las poblaciones de cueva y superficie de la especie *A. mexicanus* en la región de La Huasteca, en la Sierra de Guatemala, la Sierra de El Abra y la Sierra de la Colmena, en la región de Micos utilizados en este trabajo.

Alineamiento, Llamado de Variantes y Genotipado.

Una vez depuradas, las muestras fueron alineadas de manera individual utilizando la herramienta de alineamiento Burrows-Wheeler con el algoritmo *mem*, implementado en el programa *bwa-0.7.1* (Li y Durbin, 2009, 2010). Se utilizó el genoma completo de *Astyanax mexicanus* V 2.0 (McGaugh et al., 2014), como referencia para llevar a cabo los alineamientos. Posteriormente, se

utilizaron los programas samtools (Li et al., 2009) y Picard v1.83

(<http://broadinstitute.github.io/picard/>), para la manipulación y edición de los alineamientos.

El llamado de variantes se llevó a cabo utilizando el módulo *HaplotypeCaller* del software *Genome Analysis Toolkit* (GATK, ver. 4.1.2.0. McKenna et al. 2010). Para el procesamiento de las muestras, primero se identificaron, filtraron y fusionaron los RADloci de las muestras secuenciadas por RADseq en un solo archivo en formato gVCF (el cual conserva los sitios variables e invariables), para realizar el llamado de variantes y el genotipado de las muestras de RADseq y de genomas completos, utilizando los programas GATK y BCFTOOL (Danecek et al., 2015). Dado que el módulo *HaplotypeCaller* determina en qué regiones del genoma se debe hacer el llamado de variantes (usando la opción -L), en función de la presencia de evidencia significativa de variación, se omitieron las regiones ausentes del gVCF de los RADloci, mejorando la velocidad del análisis para las muestras de genomas completos, homogeneizando así ambos tipos de datos. Finalizando el llamado de variantes se obtuvieron archivos en formato gVCF por individuo para las muestras RADseq y WGS. Para emitir un modelo detallado de la confianza del llamado de variantes, se utilizó el parámetro BP_RESOLUTION, en la opción -ERC de *HaplotypeCaller*.

Después del llamado de variantes, las muestras se genotiparon individualmente utilizando el módulo *GenotypeGVCFs* en GATK. Esta herramienta está diseñada para realizar genotipado conjunto en una sola entrada que posee las probabilidades de los genotipos producidas por *HaplotypeCaller* con las opciones: '-ERC GVCF' o '-ERC BP_RESOLUTION'. El resultado de este paso fue un archivo gVCF compuesto de sitios variables e invariables para cada muestra secuenciada.

Inferencia de las Relaciones Filogenéticas

Los archivos gVCF ya genotipados fueron fusionados en un sólo set de datos utilizando la función *merge* del programa BCFtools (Danecek, 2015). En este set de datos, las regiones con una profundidad inferior a 10 lecturas se descartaron. De manera similar, se eliminaron las inserciones y

deleciones (INDELs) así como los sitios invariantes y las secuencias con una calidad de mapeo inferior a 40. Para disminuir la redundancia de los datos, se eliminaron los loci en desequilibrio de ligamiento con $r^2 > 0.6$, cada 600 pb.

El archivo resultante se transformó a formato fasta utilizando el script en python `vcf2phylip` ver. 2.0 (Ortiz, 2019). Este script toma como entrada un archivo VCF y usa los genotipos en el archivo para crear una matriz adecuada para el análisis filogenético en los formatos PHYLIP (versión relajada), FASTA, NEXUS o NEXUS binario. Para los SNP heterocigotos, se utilizan los códigos de ambigüedad de nucleótidos de la IUPAC. También permite detectar automáticamente cualquier nivel de ploidía.

La reconstrucción se realizó mediante el software RAxML Ver. 8.2.12 (Stamatakis, 2014), implementado en el servidor web CIPRES (url: <http://www.phylo.org>; Miller et al., 2012). Se especificaron 100 réplicas de bootstrap rápido como medida de soporte de árbol. El modelo de sustitución utilizado fue GTRCAT. La aproximación GTRCAT es una solución computacional para el modelo GTR de sustitución nucleotídica, ampliamente utilizado bajo el modelo Gamma de heterogeneidad de tasas. La idea principal detrás de la aproximación GTRCAT es permitir la integración de la heterogeneidad de tasas en los análisis filogenéticos en conjuntos de datos con más de 50 taxa a un costo computacional y consumo de memoria significativamente menor (resulta en un análisis aproximadamente 4 veces más rápido con un consumo de memoria 4 veces menor) (Stamatakis, 2016). Para realizar una reconstrucción filogenética adecuada, utilizamos la corrección por sesgo de verificación de Lewis. Las nuevas versiones de RAxML tienen este modelo de corrección resulta útil al analizar alineamientos de SNPs (Stamatakis, 2016). La topología resultante y los soportes de bootstrap se visualizaron utilizando el programa FigTree Ver. 1.4.4 (Rambaut, 2018).

Adicionalmente, se realizó un análisis de redes filogenética utilizando el algoritmo ‘Neighbor-Net’, implementado en el programa SplitsTree4 (Huson et al., 2006; Huson et al., 2014), para visualizar las relaciones en casos donde la historia evolutiva subyacente puede no ser puramente

jerárquica. Este método está basado en la unión aglomerativa de vecinos más cercanos ('Neighbor-Joining' o NJ) de Saitou y Nei (1987), utilizando matrices de distancia (Bryant et al., 2004), y es particularmente adecuado para grupos donde se ha encontrado evidencia de escenarios evolutivos complejos, como eventos de hibridación, duplicación y recombinación de genes (Huson et al., 2006). Para la reconstrucción de esta red, se creó un subset con 73 muestras de los datos totales, utilizando la función 'query' de BCFtools (Danecek, 2015), seleccionando hasta 2 individuos por localidad. Este subset fue importado al lenguaje de programación R (R core team, 2018) con la función 'snpGDSVCF2GDS' de la paquetería SNPRelate versión 1.24.0 (Zheng et al., 2012). Posteriormente se calcularon de manera pareada las proporciones de identidad por estado ('Identity-By-State' o IBS), utilizando la función 'snpGDSIBS' de la paquetería SNPRelate (Zheng et al., 2012). Estas proporciones fueron exportadas en formato NEXUS, utilizando la función 'write.nexus.dist' de la paquetería phangorn (Klaus, 2011), para la construcción de la red utilizando el algoritmo 'Neighbor-Net' en el programa SplitsTree4 versión 4.16.2 (Huson et al., 2006; Huson et al., 2014).

Estructura Poblacional

Se estimó la estructura genética y las proporciones de entremezcla entre los ejemplares, usando el software ADMIXTURE (Alexander et al., 2009; Alexander & Lange, 2011). El algoritmo ADMIXTURE estima las proporciones de ancestría y las frecuencias alélicas ancestrales utilizando un grupo de datos compuesto por una matriz con un número J de SNPs y columnas con N individuos no relacionados (Alexander et al., 2009; Alexander et al. 2015).

Realizamos una serie de análisis no supervisados (esto es, la estimación de las frecuencias alélicas ancestrales y las proporciones de entremezcla, sin información previa sobre la asignación de cada individuo), utilizando 148 muestras de *Astyanax mexicanus*, sin considerar las muestras de los grupos externos. La matriz de datos se transformó al formato requerido por ADMIXTURE utilizando el software PLINK (Purcell et al., 2007), y posteriormente se llevaron a cabo 10 corridas independientes

en ADMIXTURE, de manera no supervisada. En cada uno de estos análisis se evaluaron diferentes valores de K (el número de poblaciones ancestrales), de $K = 1$ a $K = 20$. Para conocer el mejor valor de K , usamos el procedimiento de validación cruzada de ADMIXTURE y presentamos los resultados para el cual encontró el valor de validación cruzada más bajo.

Se probó la discriminación de las poblaciones ancestrales estimadas por ADMIXTURE a través de un Análisis Discriminante de Componentes Principales (*Discriminant analysis of principal components* o DAPC Jombart et al., 2010). Este análisis discriminante está basado en la obtención de los componentes principales de los datos genéticos, para el desarrollo de variables sintéticas denominadas funciones discriminantes. Estas funciones discriminantes son la descomposición de la varianza total de los datos, en la varianza entre grupos y la varianza dentro de grupos, minimizando la primera y maximizando la segunda (Jombart et al., 2012). La estimación de los componentes principales como paso previo permite que las funciones discriminantes no estén correlacionadas.

La matriz de datos utilizada para el análisis de ADMIXTURE se importó al lenguaje de programación R (R core team, 2018) y se transformó en un objeto *genlight*, utilizando las funciones *read.vcf* y *vcfR2genlight* respectivamente, de la paquetería *vcfR* (Knaus y Grünwald, 2017). Posteriormente se realizó el DAPC utilizando la función *DAPC.genlight* del paquete *adegenet* (Jombart y Ahmed, 2011). Este análisis discriminante requiere que los componentes principales utilizados sean suficientes para asegurar una correcta discriminación entre grupos, sin que sean demasiados para generar un sobreajuste en el modelo (Jombart y Collins, 2015). Para la selección objetiva del número de componentes principales, se llevó a cabo un análisis de validación cruzada. En esta prueba, los datos se dividen en dos conjuntos: uno de entrenamiento (90% de los datos totales) y otro de validación (con el 10% de los datos restantes). El conjunto de validación se selecciona con un muestreo estratificado aleatorio, lo que garantiza que al menos un miembro de cada conglomerado o cluster esté representado en ambos conjuntos (Jombart y Collins, 2015). Posteriormente, el DAPC se lleva a cabo en 30 repeticiones en el conjunto de entrenamiento con números variables de componentes principales

retenidos, y se evalúa qué tan acertado es el análisis de predecir la pertenencia entre los individuos del conjunto de validación (a través de la raíz del error cuadrático medio) para identificar el número óptimo de componentes principales a retener (Jombart y Collins, 2015). Los clusters o conglomerados resultantes del DAPC se visualizaron en un diagrama de dispersión, utilizando las dos primeras funciones discriminantes.

Estimación de Flujo Génico y Eventos de Migración en *Astyanax*

Los eventos de flujo génico ancestral y reciente fueron estimados utilizando el algoritmo de TreeMix (Pickrell & Pritchard, 2012). Este algoritmo modela los patrones de divergencia y entremezcla a lo largo de la historia evolutiva de múltiples poblaciones, utilizando una aproximación gaussiana de la deriva génica. TreeMix estima el gráfico \mathbf{G} , que representa la topología de la historia evolutiva de las poblaciones. Este es un gráfico acíclico, similar a un árbol filogenético enraizado, con longitudes de rama específicas y eventos de flujo génico.

Para simplificar la búsqueda, el programa asume que la historia evolutiva de las poblaciones muestreadas es aproximadamente similar a la de un árbol filogenético, por lo que comienza buscando el árbol de máxima verosimilitud, adoptando un algoritmo similar al de Felsenstein (Felsenstein, 1981). Para buscar en el espacio de posibles gráficos, TreeMix adopta un enfoque de “Hill-climbing”. Luego, selecciona al azar tres poblaciones, optimiza las longitudes de las ramas para los tres árboles posibles y elige el mejor árbol en términos de verosimilitud compuesta (la verosimilitud compuesta deriva de la matriz de covarianza de las frecuencias alélicas entre las poblaciones comparadas). Después, comienza a agregar las poblaciones restantes una por una en un orden aleatorio, realizando rondas de reordenamientos locales, manteniendo el árbol resultante sólo si aumenta la verosimilitud (Pickrell & Pritchard, 2012).

Después de construir el árbol, TreeMix corrige la posición de la raíz utilizando información previa sobre el grupo externo. Luego calcula la matriz de covarianza residual R y agrega bordes de migración en pares de poblaciones según los residuos máximos de la matriz R . Cuando TreeMix añade un borde de migración que aumenta la verosimilitud del gráfico, realiza modificaciones en la topología de la red, realizando intercambios con el vecino más cercano y el destino del evento de migración. Una vez que se ha alcanzado el máximo local de la verosimilitud por este método, se realizan modificaciones en la topología de la red haciendo intercambios de vecinos más cercanos en todos los nodos internos. El programa itera sobre este procedimiento para un número predeterminado de eventos de migración. Para cada gráfico que se prueba, el programa optimiza las longitudes de las ramas y los pesos de los bordes de migración (Pickrell y Pritchard, 2012).

En este trabajo se visualizaron los eventos de migración con el programa TreeMix entre los clusters observados en el programa ADMIXTURE. Siguiendo una pauta similar a la de Herman y colaboradores (2018), para este análisis no se consideraron los individuos con evidencia de entremezcla genética. Si bien la existencia de híbridos es consistente con el intercambio genético a largo plazo, tales híbridos no proporcionan evidencia de introgresión histórica. Este subset de datos se importó en el lenguaje de programación R (R core team, 2018) y fue transformada en un objeto *genlight* usando la librería *vcfR* (Knaus & Grünwald, 2017). Después de eso, usamos la función *gl2treemix* del paquete *DartR* (Gruber y Georges, 2019) para convertir la matriz en el formato correcto para TreeMix.

Usando TreeMix, se evaluaron diferentes escenarios de flujo génico, con hasta diez eventos de migración (m) diferentes (de $m = 1$ a $m = 10$) por escenario. Para tener en cuenta que los SNPs cercanos podrían no ser independientes, el análisis se llevó a cabo utilizando diferentes particiones de la matriz de datos original, en ventanas de tamaño K , utilizando el parámetro $-k$ en TreeMix. Se utilizaron 10 valores diferentes de k , comenzando con $k = 600$ y aumentando 50 bases por iteración ($k = 600, 650, 700 \dots 1050$). Este procedimiento generó un total de 100 modelos de migración diferentes, con 10 modelos por cada evento de migración estimado. Para evaluar la confianza en la topología de la

red estimada por TreeMix, se construyó un árbol consenso utilizando las topologías de los 100 modelos estimados mediante el programa Mesquite Ver. 3.51 (Maddison y Maddison, 2019).

La selección del mejor modelo de migración puede ser difícil de determinar. Agregar eventos de migración adicionales siempre mejorará la verosimilitud del modelo, pero con rendimientos menores a medida que se van agregando eventos adicionales, pues éstos explican poca variación en los datos. Para la selección del número de eventos de migración, utilizamos el Paquete *OptM* versión 0.1.3 (Fitak, 2019). *OptM* selecciona el número de eventos de migración utilizando estadísticos *ad hoc* que automatizan el proceso de selección, en función de la tasa de cambio de segundo orden en la verosimilitud (Fitak, 2019). En este trabajo utilizamos como estadístico *ad hoc* el método de Evanno (Δm) para la selección del número de eventos de migraciones.

Resultados

Reconstrucción Filogenética

En general, el rendimiento de la secuenciación de las muestras de genoma completo fue de 95.41 M lecturas promedio por individuo (± 12.96 M de lecturas; min: 63.43 M, Max: 128.97 M) y para las muestras de RADseq fue de 4.69 M lecturas promedio por individuo (± 1.72 M de lecturas; min: 2.08 M, Max: 9.96 M). La profundidad promedio de mapeo sobre el genoma completo de *Astyanax* v. 2.0 (McGaugh et al., 2004) por individuo fue de 14.17 lecturas (S.D. 7.25 lecturas promedio por individuo) y la profundidad de cada sitio en el alineamiento tuvo un promedio de 13.85 lecturas (S.D. 2 lecturas por sitio). Después de los filtros aplicados, se retuvo un total de 360,651 SNP's bialélicos, con un porcentaje de datos faltantes de 3.7%. En total la matriz que se utilizó para la reconstrucción filogenómica incluyó 148 individuos del complejo *Astyanax mexicanus*, incluyendo tanto poblaciones de cuevas como de superficie, más 4 individuos de *A. aeneus* como grupo externo (tabla 1). Como se observa en la figura 2 A, las poblaciones del complejo *A. mexicanus* se dividen en dos linajes (linaje 1 y linaje 2), los cuales incluyen tanto poblaciones de cuevas como de superficie entremezclados, con elevados valores de soporte (>95).

El linaje 1 está compuesto de seis clados principales. El primer clado, el clado 1F, incluye a la población de superficie de la localidad de Santa Anita en San Luís Potosí. Después se recuperan en un clado a las poblaciones de la Sierra de Guatemala, las cuales se dividen en dos subclados distintos, los cuales resultan congruentes con la cercanía geográfica de las poblaciones. Las poblaciones de cueva de Jineo, Molino y Escondido se agrupan en el subclado 1D, y corresponde con la región de Gomes Farías (Elliot, 2018), mientras que las poblaciones de cueva de Caballo Moro y Vásquez, se agrupan en el subclado 1E y corresponden con la región de Chamal-Ocampo (Elliot, 2018).

El siguiente clado incluyó a dos subclados: el 1C el cual incluyó a la población de cueva Subterráneo, de la Sierra de La Colmena, en la región de Micos. El clado 1B incluyó a las poblaciones

de superficie de los ríos de la cuenca del río Pánuco Choy, Coy, Tampoan y Mante. Adicionalmente, 3 individuos de la población de la cueva Arroyo, de la Sierra de El Abra, fueron agrupados dentro de este clado. Finalmente el clado 1A incluye a las poblaciones de superficie de la cuenca del Bravo-Conchos, al Norte de México: río Bravo en Texas, Río Hualahuises, Río San Fernando y Río Conchos.

Dentro del Linaje 2, las poblaciones de cueva de la Sierra de El Abra y las poblaciones de superficie formaron 5 clados diferentes. Las poblaciones de superficie de este linaje, compuestas por las poblaciones de río Rascón, río Gallinas y río Tamasopo, formaron un clado monofilético (2E). Las poblaciones de cueva, formaron 4 clados diferentes, con una elevada congruencia geográfica a lo largo de la Sierra de El Abra.

El clado 2D incluyó a las poblaciones de cueva de Chica, Chiquitita y Toro, ubicadas en la región sur de la Sierra del Abra. El clado 2C, recuperó a la mayoría de las poblaciones de cueva de la región media de la Sierra de El Abra (Cuevas de Tigre, Sabinos, Arroyo, Tinaja, Montecillos, Jos, Piedras y Palma Seca), con excepción de las poblaciones de Japonés y Yerbaniz, las cuales formaron un clado diferente, El clado 2B. Finalmente el clado 2A incluyó a los individuos de la población de cueva de El Pachón, la cual se encuentra en la región norte de la Sierra de El Abra.

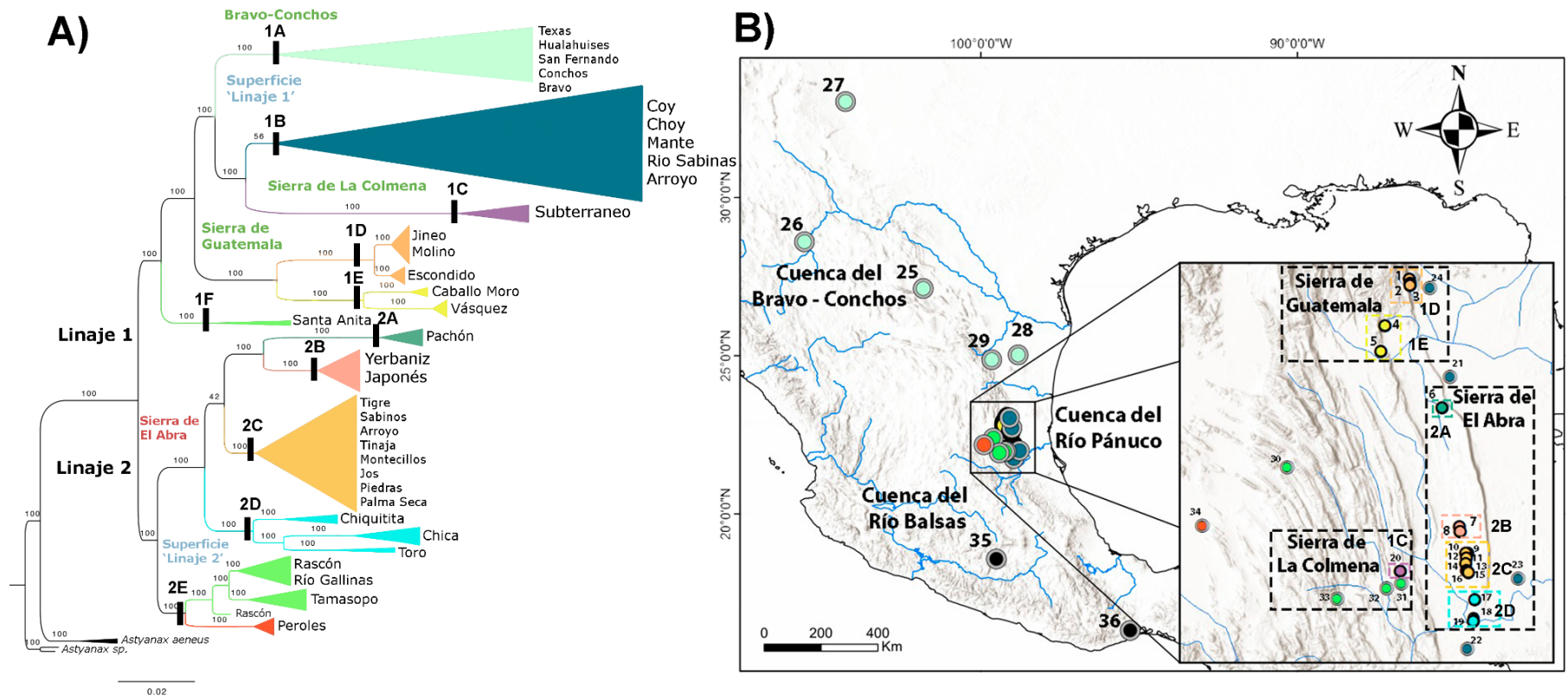


Figura 2. A) Reconstrucción filogenética de Máxima Verosimilitud en RAxML v8.2.8 (Stamatakis, 2014) de los SNPs obtenidos de las poblaciones de *A. mexicanus* de cueva y de superficie al noreste de México. B) Distribución de los sitios de muestreo de las poblaciones de cueva (circulo externo negro) y superficie (circulo externo blanco) de la especie *A. mexicanus* en la región de La Huasteca, en la Sierra de Guatemala, la Sierra de El Abra y la Sierra de la Colmena, en la región de Micos. Los números corresponden a las localidades descritas en la tabla 1, y los colores y códigos utilizados corresponden con los linajes encontrados en la reconstrucción filogenética.

La Fig. 3 ilustra la red filogenética para el complejo *Astyanax mexicanus*, obtenida usando las proporciones pareadas de IBS y estimada usando el algoritmo 'Neighbour-Net'. El índice de divisibilidad para la red, el cual es un estimado de bondad de ajuste de cuán precisa es la representación del conjunto de datos (Dress et al., 1996), fue de 98.059%. A pesar de que la red obtenida no muestra un patrón totalmente jerarquizado (en forma de árbol), es congruente con la inferencia por máxima verosimilitud (fig. 2 A), y también apoya la subdivisión de las diferentes poblaciones de cueva en dos linajes distintos, con la correspondencia geográfica de las poblaciones de cueva. La presencia de cajas en la red sugieren posibles eventos de introgresión entre taxones, que no son observados por el árbol filogenético. Sin embargo, incluso en presencia de estos patrones de reticulación, la señal filogenética subyacente es evidente para las poblaciones de cueva y superficie del complejo *Astyanax mexicanus*.

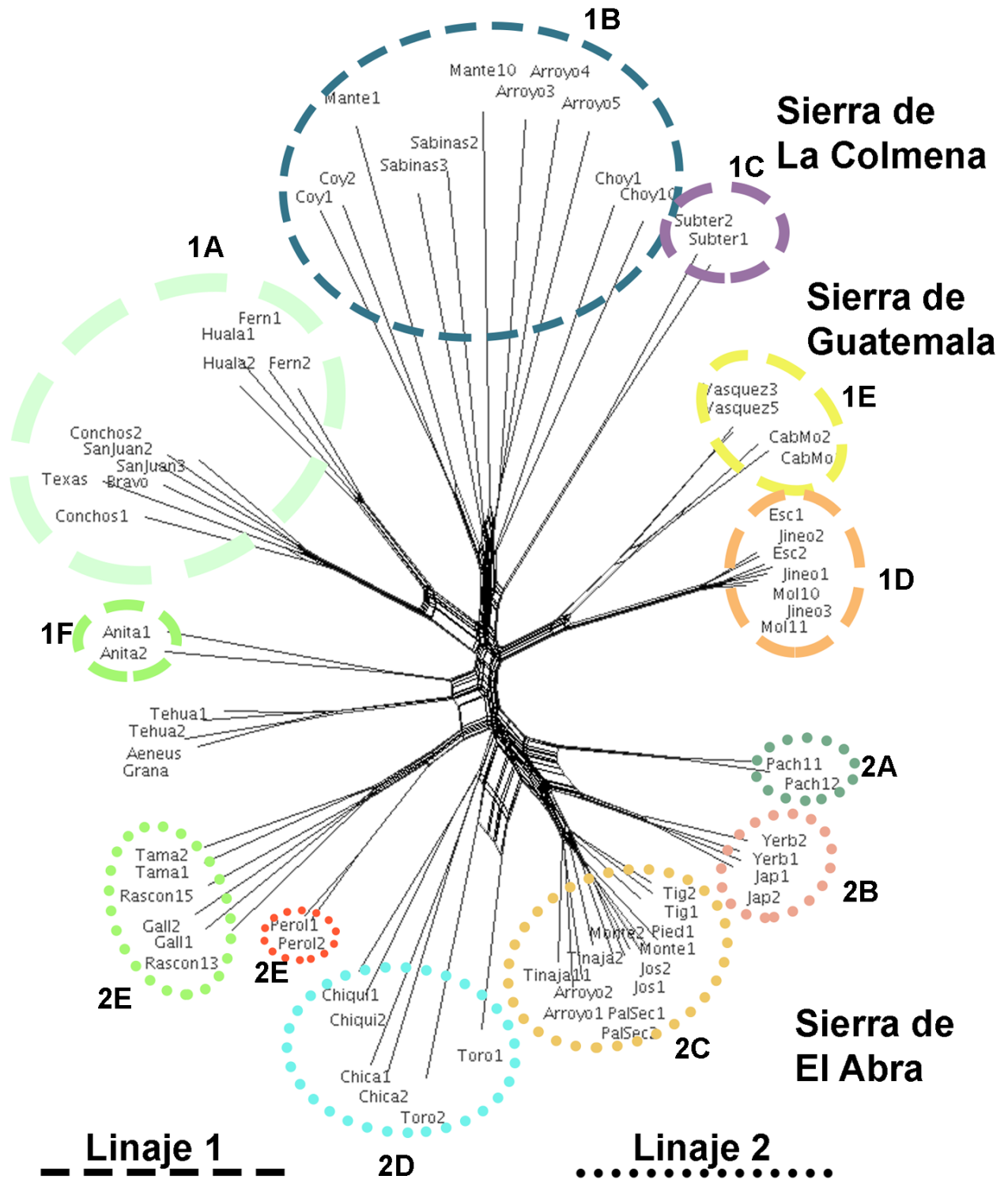


Figura 3. Diagrama de la red filogenética para el complejo *Astyanax mexicanus*, construido mediante el procedimiento 'Neighbour-Net' (Huson et al., 2006) utilizando las proporciones pareadas de IBS. El índice de divisibilidad de la red fue de 98.059%

Evaluación de la Estructura Genética en las Poblaciones de *Astyanax mexicanus*.

Los resultados del análisis de ADMIXTURE apoyaron la asignación de los individuos de *Astyanax mexicanus* en K=11 grupos genéticos (figura 4 B), basados en los valores de la prueba de validación cruzada, para 4 de las 10 iteraciones realizadas.

El análisis de asignación de ADMIXTURE recuperó en gran medida los clados principales obtenidos en la reconstrucción de máxima verosimilitud de RaxML (fig. 2 A). Para la asignación poblacional de K=11 grupos genéticos, las poblaciones de cueva y de superficie del Linaje 1 fueron asignadas a 5 grupos y las poblaciones de cueva y de superficie del Linaje 2 fueron asignadas a 6 grupos. En general, los individuos de las distintas poblaciones presentaron altos valores de probabilidad de asignación, con un número relativamente bajo de individuos con evidencia de entremezcla entre poblaciones.

Las poblaciones de superficie de Linaje 2 (clado 2E) se dividieron en dos clusters genéticos, sin evidencia de entremezcla entre ellos: el primer grupo separó a la población del río Peroles del resto de las poblaciones de superficie, mientras que el segundo grupo incluyó a las poblaciones de superficie de Río Gallinas, Río Rascón, Tamasopo y Santa Anita. Sin embargo, los individuos de la población de Santa Anita presentan evidencia de entremezcla con otros grupos genéticos, incluso poblaciones de cueva y superficie de linajes distintos.

Las poblaciones de cueva del Linaje 2 fueron asignadas a 4 grupos genéticos diferentes. Las poblaciones de las cuevas Chica, Chiquitita y Toro (clado 2D), al sur de la Sierra de El Abra, formaron un grupo genético (cluster 3) en el cual se puede observar que algunos individuos de las poblaciones de Chiquitita y Toro presentan evidencia de introgresión con las poblaciones de superficie del clado 1B, del linaje 1, así como con otras poblaciones de cueva del clado 2C de la Sierra de El Abra (cluster 4). Las poblaciones de cueva de la región media de la Sierra de El Abra (clado 2C) formaron un grupo genético (cluster 4). Se encontró evidencia de introgresión en las poblaciones de cueva Tigre, Arroyo y

Tinaja con las poblaciones de cueva de la región media alta de la Sierra de El Abra (clado 2B, cluster 5), así como con las poblaciones de superficie de linaje nuevo (clado 1B, cluster 7). La región media alta, conformada por las poblaciones de cueva de Japonés y Yerbaniz (clado 2B), formaron un grupo genético (cluster 5) sin evidencia de introgresión. De igual manera, la población de cueva de Pachón, en la región más al norte de la Sierra de El Abra (clado 2A), formó un grupo genético independiente (cluster 6), sin evidencia de introgresión con otras poblaciones.

Las poblaciones de superficie de Linaje 2 (clado 1B) formaron un grupo genético (cluster 7) con evidencia de entremezcla con la población de cueva de Subterráneo en la Sierra de la Colmena, en la región de Micos (clado 1C, cluster 8). Algunos individuos de las población de Cueva Arroyo (los individuos Arroyo3, Arroyo4 y Arroyo5) en la Sierra de El Abra, fueron asignados a este grupo genético. Este resultado se observó de manera similar en la reconstrucción por máxima verosimilitud.

Similar al caso de las poblaciones de cueva de la Sierra de El Abra, las poblaciones de cueva de la Sierra de Guatemala también muestran una correspondencia entre la asignación encontrada por ADMIXTURE y su distribución geográfica. Las poblaciones de cueva de la Sierra de Guatemala formaron dos grupos genéticos diferentes. El primero (cluster 9), agrupó a las poblaciones de Vazquez y Caballo Moro, de clado 1D y el segundo (cluster 10) incluyó a las poblaciones de Escondido Jineo y Molino del clado 1E. Por último, las poblaciones del Norte de México (clado 1A) formaron un grupo genético (cluster 11) sin evidencia de entremezcla con algún otro grupo genético.

Resultados del Análisis Discriminante

La prueba de validación cruzada para el análisis discriminante de componentes principales (DAPC) resultó en 20 componentes principales (PCs) y 10 funciones discriminantes (DAs) retenidas para la evaluación de los $K=11$ clusters estimados por ADMIXTURE. La varianza conservada fue de 63% de los datos totales. Se construyó una gráfica de dispersión utilizando las primeras 2 funciones

discriminantes, representando a los individuos como puntos y los grupos genéticos fueron encerrados por elipses de inercia, dibujados con un coeficiente positivo de 1.5. (figura 3 A). Los clusters fueron conectados por su cercanía en el espacio discriminante a través de un árbol de expansión mínima (*Minimum Spanning tree* o MST).

La distribución de los clusters en el espacio discriminante, sugiere una buena separación entre los grupos, esto es, que no se sobrelapan entre sí (figura 3 A). El análisis muestra una fuerte diferenciación entre las poblaciones de cueva de la Sierra de Guatemala (clusters 9 y 10), y con el resto de los clusters. Tanto el MST, como la reconstrucción por máxima verosimilitud, muestran que las poblaciones de Vázquez y Caballo Moro (cluster 9) y las poblaciones de Escondido, Jineo y Molino (cluster 10) están más relacionadas entre sí, que con el resto de las poblaciones troglobias del sistema de El Abra y Subterráneo. Resulta interesante observar que las poblaciones de superficie del linaje 1 y del linaje 2 no se sobrelapan en el espacio discriminante, ni se relacionan directamente a través del árbol de distancias.

La población de Peroles (cluster 1) y el resto de las poblaciones de superficie de linaje viejo (cluster 2) presentan una distancia genética más corta con respecto a las poblaciones de superficie de linaje nuevo (cluster 7). El árbol de distancias también muestra que las poblaciones de cueva de la región media baja de la Sierra de El Abra (cluster 4) presentan una menor distancia genética con las poblaciones de río Gallinas, Rascón y Tamasopo (cluster 2). Por otro lado, las poblaciones de Chica, Chiquitita y Toro (Cluster 3) son más cercanas a las poblaciones de superficie del linaje 1 (cluster 7). Los clusters conformados por poblaciones de cueva en las que se ha reportado la presencia de híbridos (poblaciones como El Pachón, Subterráneo, Chica, Chiquitita y Toro), se encuentran más próximos a los clusters por poblaciones de superficie, mientras que las poblaciones que han estado sujetas a fuertes presiones de deriva (como Sótano de Jineo, Sótano de Japonés,, etc.) se encuentran en los extremos del gráfico.

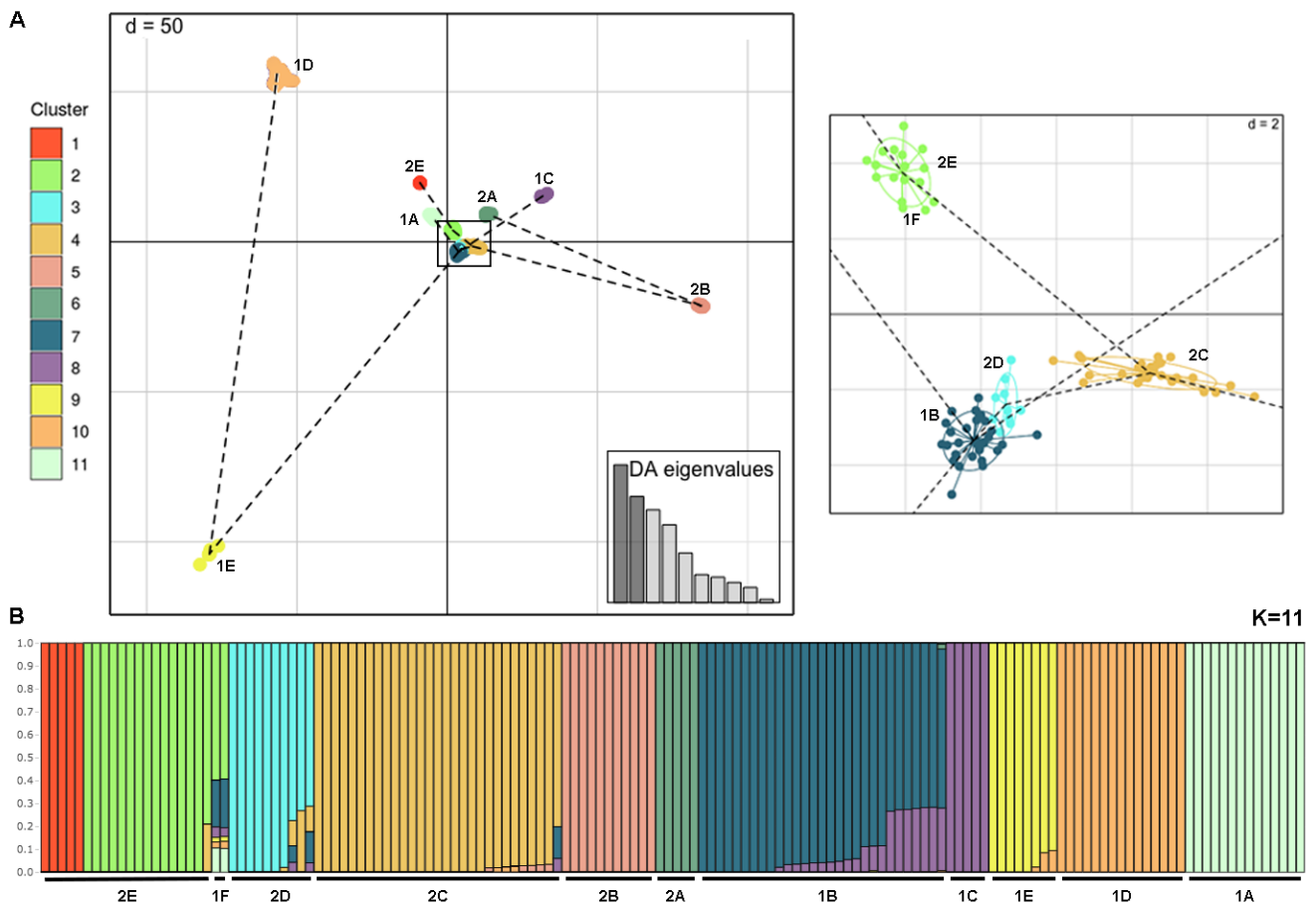


Figura 4. A) Análisis Discriminante de Componentes Principales (DAPC) de las 11 categorías genéticas estimadas por el análisis de ADMIXTURE. Para la evaluación de los clusters, se utilizaron 20 componentes principales (63% de la variación acumulada de los datos). El análisis discriminante descompuso los componentes principales en 10 funciones discriminantes (DA). Las dos primeras funciones discriminantes fueron utilizadas para realizar el gráfico. La línea punteada representa un árbol de expansión mínima (Minimum Spanning tree o MST), construido con base en las distancias de los clusters. B) Análisis no supervisado de ADMIXTURE para los individuos utilizados en la reconstrucción filogenética, sin considerar los grupos externos. Este análisis muestra entremezcla entre las poblaciones de linaje nuevo y viejo, y entre poblaciones de cueva y de superficie.

Flujo Génico Reciente y Ancestral

Utilizando el programa TreeMix (Pickrell & Pritchard, 2012), se calcularon los eventos de migración, usando la información obtenida del análisis de ADMIXTURE. La prueba *ad hoc* de Δm basada en el método de Evanno, para la selección del número de los eventos de migración (Fitak, 2018), identificó cuatro eventos principales de flujo génico entre los clusters inferidos por ADMIXTURE (fig. 4 B), con una varianza explicada de 99.48% del total de los datos. Las relaciones de ancestría recuperadas entre los grupos genéticos en la topología resultante (fig. 5 A), fueron similares a las observadas en la reconstrucción por máxima verosimilitud.

Los resultados obtenidos por TreeMix ilustran el patrón complejo de flujo de génico ocurrido en repetidas ocasiones a partir de poblaciones de superficie hacia las poblaciones de cueva en la región de El Abra, al tiempo que proporcionan evidencia de flujo génico de las poblaciones al Sur del Eje Neovolcánico Transversal (i.e. *Astyanax aeneus*) hacia las poblaciones de superficie de linaje Nuevo. Tres de los cuatro eventos de migración sugieren que la región de la Sierra de El Abra ha experimentado episodios de invasión, por parte de peces de superficie, hacia los distintos sistemas de cuevas, durante diferentes periodos de tiempo. El evento más reciente de migración se estimó hacia la región sur, por parte de las poblaciones actuales de superficie del Linaje 1. Para el norte de la Sierra de El Abra (cluster 6), el episodio de invasión de peces de la superficie tuvo lugar a partir de las poblaciones ancestrales que dieron origen a las poblaciones del linaje nuevo. Para las regiones media y sur de la Sierra de El Abra, es posible identificar un evento de migración más antiguo, en comparación a los dos previamente descritos, a partir de organismos de superficie ancestrales previos a la divergencia de ambos linajes.

En este análisis también podemos observar *la presencia de ramas terminales largas para* algunos grupos (como el cluster 3 y el cluster 6). Es importante mencionar que la topología recuperada por TreeMix coloca a las poblaciones de superficie asignadas como parte del linaje 2 por fuera de

ambos linajes, y no dentro del linaje 2. Este resultado es diferente del observado en nuestra reconstrucción por máxima verosimilitud y de la red filogenética.

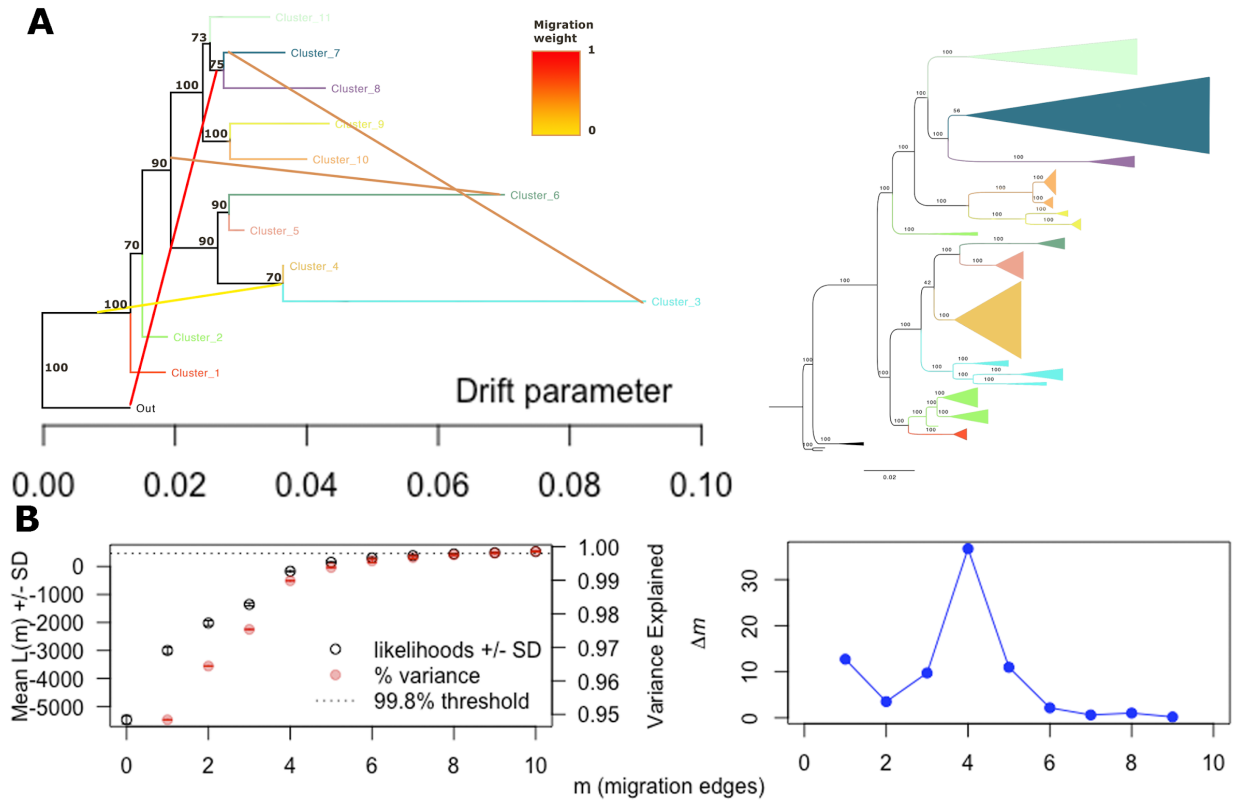


Figura 5 A) Análisis de flujo génico con TreeMix (Pickrell and Pritchard 2012) utilizando los clusters estimados como poblaciones. Las poblaciones que mostraron evidencia de entremezcla en el Análisis de ADMIXTURE fueron descartadas. B) Resultados del estadístico ad hoc ΔM implementando en el paquete OptM (Fitak, 2018), basado en el método de Evanno, el cual lleva la selección del modelo con base en la tasa cambio en el logaritmo de la verosimilitud, para la selección de los eventos de migración en TreeMix. La gráfica poblacional de TreeMix con 4 eventos de migración. En esta gráfica se pueden observar los patrones de entremezcla entre *A. aeneus* con las poblaciones de superficie de linaje nuevo, así como peces de superficie entre poblaciones de cueva de la Sierra de El Abra en diferentes momentos de la historia evolutiva.

Discusión

En este trabajo se evaluaron las relaciones de ancestría, la estructura genética y los eventos de flujo génico, entre las poblaciones de cueva y superficie, con el fin de aportar más información que pueda ayudar a resolver la controversia con respecto al origen y relación entre las poblaciones de *A. mexicanus* de cueva y de superficie. En este trabajo se encontró que las poblaciones de cueva y superficie de *A. mexicanus* se agrupan en dos linajes distintos. La estructura genética entre las poblaciones se corresponde con la distribución geográfica de las poblaciones y estas han experimentado diversos eventos de flujo génico a lo largo de su historia evolutiva.

Historia Evolutiva de las Poblaciones de *Astyanax mexicanus* en la Región de El Abra.

En estudios previos, el consenso en cuanto a la evolución de *Astyanax mexicanus* a la vida en las cuevas, basada en la variación del mtDNA y de nucDNA, es que las poblaciones de cuevas en la Sierra de Guatemala, Sierra de la Colmena y Sierra de El Abra, fueron colonizadas por al menos dos linajes distintos de superficie (Dowling et al.2002; Strecker et al.2003, 2004, 2012; Ornelas-García et al. 2008; Hausdorf et al., 2011; Bradic et al., 2012; Herman et al., 2018). Estos trabajos han propuesto que las cuevas de la Sierra de El Abra habrían sido colonizadas en un primer episodio a partir de peces de superficie ancestrales, clasificadas como linaje viejo (Strecker et al., 2004), mientras que las cuevas de la Sierra de la Colmena y la Sierra de Guatemala (clasificadas como linaje nuevo), fueron invadidas por un evento más reciente de colonización, por parte de peces que habitan actualmente en los ríos de esta región (Hausdorf et al., 2011; Ornelas-García, et al. 2008; Bradic et al., 2012). En este trabajo se encontró que la invasión a las cuevas se ha dado de manera polifilética, en dos linajes diferentes, con una elevada correspondencia geográfica. Tanto en la reconstrucción por máxima verosimilitud (fig. 2 A), como en la red filogenética (fig. 3), las poblaciones de nueva de la Sierra de El Abra formaron un clado monofilético dentro del linaje 2, mientras que las poblaciones de cueva de la Sierra de Guatemala

y la Sierra de La Colmena, formaron 2 clados independientes, dentro del linaje 1, recuperando así la hipótesis de los dos linajes. Dentro de estos dos linajes se identificaron tres eventos independientes de colonización a las cuevas, demostrando que las características fenotípicas asociadas a la vida en cuevas han surgido en repetidas ocasiones a lo largo de la historia evolutiva del complejo.

Adicionalmente, se encontró que las relaciones de ancestría de las poblaciones de cueva, están asociadas a su distribución geográfica, lo cual corresponde con un claro patrón biogeográfico. Este patrón ha sido observado en otros organismos troglobios de la región. Por ejemplo, el camarón de cueva *Spelaeomysis quinterensis* (Villalobos, 1951), presentó dos linajes, uno distribuido en la sierra de El Abra, mientras que otras poblaciones troglobias de la región de La Huasteca mostraron un linaje distinto (Espinasa et al., 2020). Esto sugiere que las barreras geográficas entre regiones han moldeado los patrones biogeográficos no sólo de los peces de cueva, sino de otras comunidades acuáticas (Espinasa et al., 2020).

Finalmente, nuestros resultados fueron inciertos en cuanto la asignación de la población de superficie de Santa Anita (clado 1F). En la reconstrucción filogenética por máxima verosimilitud, la población de superficie de Santa Anita forma un clado monofilético (clado 1F), dentro del linaje 1, mientras que en el análisis de ADMIXTURE presenta una probabilidad de asignación mayor al cluster 2, que incluye las población de superficie del linaje 2 (clado 2 E), con evidencia de entremezcla con otras poblaciones, y en algunos casos, geográficamente muy distantes (i.e. con las poblaciones del clado 1A, restringidas a la cuenca del Bravo-Conchos, al norte de México (fig. 2 B)). Las proporciones de asignación de la población Santa Anita pudieron ser modeladas como una mezcla de las poblaciones muestreadas, en lugar de tener su propia población ancestral (efecto de “*Ghost admixture*”; Lawson et al., 2018).

Poblaciones de cueva de *Astyanax mexicanus* clasificadas dentro del grupo del Linaje Viejo

Las poblaciones de cueva de la Sierra de El Abra (fig 2, linaje 2) han sido clasificadas dentro del linaje viejo, y diversos trabajos las han relacionado con el primer episodio de colonización a las cuevas (Strecker et al., 2004; Bradic et al., 2012; Bradic et al., 2013; Herman et al., 2018). Sin embargo, en diversos trabajos, ha existido cierta controversia respecto a las relaciones filogenéticas de las poblaciones de cueva de Pachón y Chica. Utilizando marcadores mitocondriales, las poblaciones de la cueva de Pachón y Chica se han agrupado con los peces de la superficie geográficamente cercanos (Dowling et al., 2002; Strecker et al., 2003, Ornelas et al., 2008). En contraste, con datos de microsatélites, estas poblaciones se agrupan con poblaciones de cueva de la Sierra de El Abra (Strecker et al., 2003). En este trabajo, tanto el árbol como la red filogenética agrupó a estas poblaciones como parte del linaje 2 (linaje viejo), compuesto por las poblaciones de cueva de la Sierra de El Abra, resultado similar a la evidencia encontrada con microsatélites (Strecker et al., 2003).

Por otra parte, el análisis de ADMIXTURE mostró presencia de entremezcla genética entre las poblaciones de cueva del clado 2D (compuesto por las poblaciones de Chica, Chiquitita y Toro) y las poblaciones de superficie del clado 1B. Estas poblaciones de cueva se encuentran relativamente cerca de cuerpos de agua, como el río Tampaón, localizado a 200 m de la cueva de Chiquitita (Espinasa et al., 2018), por el cual, en temporadas de lluvias, los peces de superficie pueden ingresar hacia las cuevas y luego dispersarse a través de los conductos subterráneos, hasta llegar a las pozas internas de la cueva de Chica (Mitchell et al. 1977). Se ha sugerido que la posibilidad de intercambio genético entre poblaciones de cueva y de superficie permite la introgresión de haplotipos mitocondriales (Strecker et al., 2003; Hausdorf et al., 2011; Bradic et al., 2012), por lo que las conclusiones basadas solamente en marcadores mitocondriales podrían ser engañosas; las inferencias utilizando este tipo de marcadores podrían reflejar eventos recientes de hibridación, y enmascarar las relaciones filogenéticas naturales.

Poblaciones de cueva de *Astyanax mexicanus* clasificadas dentro del grupo del Linaje Nuevo

Estudios previos de estructuración genética, han propuesto que las cuevas de la Sierra de Guatemala y de la Sierra de la Colmena, en la región de Micos, se establecieron posteriormente a las poblaciones de la Sierra de El Abra (Dowling et al. 2002; Strecker et al. 2004; Ornelas-García et al. 2008; Strecker et al. 2011; Bradic et al. 2012). En este trabajo se encontró que la colonización de las poblaciones de cueva de la Sierra de Guatemala (clado 1D y 1E) y de la Sierra de la Colmena (clado 1C), corresponden a dos eventos distintos de colonización, siendo el evento más reciente el que daría lugar a la población de cueva de Subterráneo, en la Sierra de La Colmena (fig. 2 A).

La cueva de Subterráneo, en la región de Micos (Sierra de la Colmena), es la población más aislada geográficamente con respecto al resto de poblaciones de cuevas de *Astyanax mexicanus* (Wilkens y Burns, 1972; Peters et al., 1975). Esta localidad incluye una entrada por el río Subterráneo que se encuentra a 35 km al Oeste de la Sierra de El Abra, separadas por la Sierra de Nicolás Pérez y la Sierra de Colmena (Mitchell et al., 1977). Considerando las poblaciones de cueva de linaje nuevo, la distancia entre la región de Micos y la cueva más cercana de la Sierra de Guatemala, es de aproximadamente 90 km, y existen diversas barreras geográficas entre ellas. Hasta la fecha, no se han documentando pasajes subterráneos entre estas dos regiones, por lo que es muy probable que las poblaciones de la región de Micos y la Sierra de Guatemala representen eventos separados de invasión (Bradic et al., 2012).

El análisis de ADMIXTURE (fig. 4, B), muestra la evidencia de entremezcla de las poblaciones de superficie del linaje 1 (clado 1B) con las poblaciones de cueva de Subterráneo (clado 1C), y con las poblaciones de cueva de la Sierra de El Abra. Durante la temporada de lluvias, la cueva Subterráneo recibe enormes afluencias de agua corriente que transportan peces de superficie que habitan los arroyos circundantes. Una vez dentro de las cuevas, los peces de superficie coexisten con los peces de cuevas, llegando en algunos casos a reproducirse, formando una amplia gama de fenotipos híbridos intermedios (Bibliowicz et al., 2013).

Nuevos Conocimientos Sobre las Relaciones Filogenéticas del Complejo *Astyanax mexicanus*.

Algo que debe tenerse en cuenta durante la inferencia filogenética, es la relación entre el muestreo de taxones y la escala filogenética. El uso de un subconjunto de taxones para la inferencia filogenética puede conducir a uno de dos casos: el primero, que los taxones formen un subárbol pequeño, muestreado de un árbol más grande, y el segundo, que formen un patrón ampliamente disperso en un árbol más grande (Zwickl et al., 2002). Uno esperaría que el primer caso no presentara un problema filogenético difícil (ni un problema comparable al que plantean todos los taxones), mientras que el segundo presentaría un problema mucho más difícil, pero más relevante. (Zwickl et al., 2002).

En muchos casos, la mejor opción disponible para aumentar la precisión del análisis filogenético, es aumentar el número de taxones utilizados. Se ha argumentado que un muestreo extensivo tiene un efecto positivo e importante en la precisión de la estimación filogenética (Zwickl et al., 2002). Sin embargo, aumentar el número de taxones simplemente no es posible, debido a la falta de taxones existentes para muestrear, la dificultad que implica acceder a los sitios de colecta, etc. Este último parece ser el caso del complejo *Astyanax mexicanus*.

Debido al difícil acceso a las poblaciones, el número de cuevas y de superficie utilizadas en la mayoría de trabajos anteriores ha sido limitado (Awise & Selander, 1972; Dowling et al., 2002; Strecker et al., 2003, 2004; Strecker et al., 2012; Coghill et al., 2014; Fumey et al., 2018; Herman et al., 2018), lo que dio lugar en algunos casos a interpretaciones sesgadas sobre la evolución de este complejo.

Un buen ejemplo de lo anterior es la idea de la extinción de las poblaciones de superficie del linaje viejo (Bradic et al., 2012), y que las poblaciones de cueva de la Sierra de El Abra eran un relictos de esas poblaciones de superficie. Sin embargo, en este trabajo encontramos que las poblaciones de superficie de río Gallinas, río Rascón, río Tamasopo y río Peroles formaron un clado monofilético (fig. 2, clado 2E), sin evidencia de mezcla con el resto de poblaciones de superficie (figura 4, B). Este clado

resultó ser el grupo hermano de las Poblaciones Cueva de la Sierra de El Abra, y sugiere que estas poblaciones son remanentes del proceso de invasión a las poblaciones de cueva de la Sierra de El Abra.

Previamente, se planteó la posibilidad de que estas poblaciones de superficie hubieran quedado aisladas, debido a los procesos erosivos que hicieron descender los valles circundantes. Estos procesos erosivos habrían dejando a estas poblaciones de superficie aisladas a través de la Cascada de Tamúl (Espinasa & Espinasa, 2016). Actualmente, esta caída vertical aísla la corriente río arriba de los drenajes inferiores, los cuales están habitados por las poblaciones de superficie del linaje 1 (i.e. linaje nuevo). En este trabajo, a través de un amplio muestreo de las poblaciones de superficie de la región de La Huasteca, el aislamiento geográfico de las poblaciones de superficie del linaje 2 (i.e. linaje viejo), y su relación con las poblaciones de cueva de la Sierra de El Abra, fue comprobada a través de la reconstrucción por máxima verosimilitud, soportado con altos valores de soporte, y congruente con los análisis de estructura poblacional, tanto de ADMIXTURE (fig. 4 B) como de DAPC (figura 4 A).

Estructura y Flujo Genético de las Poblaciones Troglodias de la Sierra de El Abra

Mitchell, Russell y Elliot (1977) propusieron la existencia de un drenaje subterráneo que dividía a la Sierra de El Abra en diferentes sistemas, como resultado de cambios en la elevación ancestral y reciente. En años posteriores, a través de trabajos de espeleología y exploración de las cuevas de esta región, se confirmó el alto grado de conectividad a través de canales de agua subterránea, presentes entre las poblaciones de cueva (Elliot, 2018), lo que permitió clasificar en diferentes sistemas hidrológicos la región de El Abra (Elliot 2018).

Esta propuesta de conectividad entre poblaciones de cueva a través de canales subterráneos, ha sido apoyada parcialmente por algunos estudios moleculares (Strecker et al. 2012, Bradic et al., 2012; Herman et al., 2018). Sin embargo, este estudio es el primero en incluir un muestreo exhaustivo de las poblaciones de cueva de la región de El Abra (14 poblaciones de cueva), permitiendo poner a prueba la extensión de dicha conectividad entre cuevas. En este trabajo se encontró que la asignación genética

recuperada con el programa ADMIXTURE, para las poblaciones del complejo *Astyanax mexicanus*, es congruente con los sistemas hidrológicos propuestos por Elliot (2018).

Para la Sierra de El Abra, existen cuatro sistemas diferentes que pueden clasificarse de acuerdo a sus conexiones hidrológicas:

1) **La región más al norte de la Sierra de El Abra.** Esta región está ubicada en El Valle de Antiguo Morelos. Cerca de Praxedis Guerrero, en el estado de Tamaulipas, México. Esta región capta agua a través de dos afluentes: Arroyo El Lagarto, que drena al norte del río Capote y del Río Ocampo hacia el sur, a Río Valles. Las conexiones hidrológicas subterráneas conectan a las poblaciones de cueva El Pachón, y Venadito (Elliot, 2018). En este trabajo, este sistema hidrológico es recuperado dentro del clado 2A en la reconstrucción filogenética (fig. 2 A).

2) **El sistema de Yerbaniz.** Este sistema está compuesto por tres cuevas hidrológicamente relacionadas: Sótano de Yerbaniz, Sótano de Matapalma y Sótano de Japonés. De ellas, Yerbaniz es la más joven en tiempos geológicos, y probablemente la más activa hidrológicamente de este sistema. Estas poblaciones se ubican a 19.5 – 22.0 km de Ciudad Valles en el estado de San Luis Potosí, México (Elliot, 2018). En este trabajo, este sistema hidrológico es recuperado dentro del clado 2B en la reconstrucción filogenética (fig. 2 A).

3) **El sistema de Los Sabinos.** Este es un sistema de tres cuevas que comprende las poblaciones Cueva de Los Sabinos, Sótano del Arroyo y Sótano de la Tinaja, las cuales han sido confirmadas como sistemas hidrológicamente conectados. Sin embargo se ha sugerido que el sistema podría incluir todas las cuevas del área, hasta el Sótano de la Palma Seca, (incluyendo dentro de este sistema al Sótano del Tigre, Sótano de la Roca, El Sotanito de Montecillos, El sótano de Pichijumo, El Sótano de Jos, El Sótano de las Piedras y el Sótano de la Palma Seca). Estas poblaciones se ubican a 2.8 km de Ciudad Valles en San Luis Potosí (Elliot, 2018). En este trabajo, este sistema hidrológico es recuperado dentro del clado 2C en la reconstrucción filogenética (fig. 2).

4) **La región Sur de la Sierra de El Abra.** Este sistema se encuentra en el extremo sur de la Sierra de El Abra, aproximadamente a 200 m del Río Tampaón en el estado de San Luis Potosí, México. Este sistema está conectado hidrológicamente entre las poblaciones de Cueva Chica, Cueva los Cuates, Cueva Chiquitita y cueva de El Toro (Elliot, 2018; Espinasa,2020). En este trabajo, este sistema hidrológico es recuperado dentro del clado 2D en la reconstrucción filogenética (fig. 2).

Estructura y Flujo Genético de las Poblaciones Troglodias de la Sierra de Guatemala

Las conexiones hidrológicas subterráneas propuestas por Elliot (2018), Para la Sierra de Guatemala,dividen a las 5 poblaciones de cuevas de esta región, en 2 sistemas diferentes.

1) **El sistema de Gómez Farías.** Esta región está localizada geográficamente en el valle Este de Gómez Farías, Tamaulipas, entre La Sierra de los Mangos, y la Base de la Sierra de Guatemala. De norte a sur, las poblaciones de cueva de *Astyanax mexicanus* en este sistema hidrológico son el Sótano de Jineo, El Sótano del Molino y el Sótano Escondido, estando integradas a una red conjunta de drenaje subterráneo que podría estar siendo alimentada por el Río Nacimiento, Río Florido y Río Frío. Esta conexión subterránea propuesta por Elliot (2018), fue recuperada por nuestros análisis de ADMIXTURE (figura 4, B) y DAPC (figura 4 A), y debido a que es la primera vez que se incorporan las poblaciones de cueva de Jineo y Escondido en análisis moleculares, pudimos detectar una diferenciación de la población del Sótano Escondido, de las otras dos poblaciones de esta región, lo que podría estar relacionado con un proceso de deriva génica que deberá ser estudiado más adelante.

2) **El sistema de Chamal-Ocampo.** Este sistema se encuentra próximo al Nacimiento del Río Frío, a un costado del poblado de Chamal en el estado de Tamaulipas. Las cuevas que conforman este sistema, de norte a sur, son: Cueva de Bee, Sótano de Caballo Moro y Sótano de Vásquez. Las primeras dos poblaciones están en el valle formado por la Sierra Cucharas, y la Sierra de Guatemala, mientras que el Sótano de Vásquez se encuentra en la parte norte de la Sierra de Tamalave. Las poblaciones están conectadas hidrológicamente y los cuerpos de agua de esta región desembocan en el Nacimiento de río

Frío, excepto Vázquez, que desemboca en el río Ocampo. Este estudio es el primero en caracterizar molecularmente a la población del Sótano de Vázquez, lo cual sin duda ha permitido corroborar la afinidad genética de esta población con la población de Caballo Moro, geográficamente cercana.

Cabe resaltar que algunos de los sistemas hidrológicos, tanto de la Sierra de Guatemala como de la Sierra de El Abra, presentan evidencia de entremezcla entre sistemas hidrológicos geográficamente cercanos (fig. 4 B), por ejemplo, entre los sistemas de Gómez Farías y Chamal-Ocampo en la Sierra de Guatemala y entre los sistemas de Yerbaniz y Sabinos en la Sierra de El Abra. Adicionalmente, la red filogenética (figura 3) muestra cierto grado de reticulación entre las poblaciones del sistema norte de la Sierra de El Abra, (clado 2A) y el sistema de Yerbaniz (clado 2B), así como entre las poblaciones de los sistema de Gómez-Farías (clado 1E) y Chamal-Ocampo (clado 1D). Esto sugiere la presencia de intercambio genético entre sistemas.

Previamente se ha encontrado evidencia de la migración entre estos sistemas hidrológicos en otras especies troglobias. Por ejemplo, el nicotélido troglobio *Anelpistina quinterensis*, ha sido reportada por Espinasa (2007) en la Cueva de Pachón (sistema norte de la Sierra de El Abra) y Sótano de Yerbaniz (sistema Yerbaniz). Se piensa que este nicotélido ha migrado a través de las conexiones subterráneas entre estos sistemas, para alcanzar y establecer las poblaciones actuales en los últimos 12,000 años (Espinasa 2007; Espinasa et al., 2014). La evidencia de entremezcla entre sistemas hidrológicos geográficamente cercanas, dentro de la Sierra de El Abra, ha sido evaluada previamente (Bradic et al., 2012; Strecker et al.,2012; Herman et al., 2012), pero hasta ahora, no se había encontrado una señal tan clara entre los patrones de estructuración genética poblacional y las conexiones hidrológicas propuestas (Espinaza & Espinaza 2015; Elliot, 2018; Espinasa et al., 2020). Debido a la amplia conectividad entre las cuevas y la amplitud de los cambios geológicos a lo largo de la historia hidrogeológica (Elliot, 2018), es bastante probable que las especies adaptadas a la vida en las cuevas, incluidas las poblaciones de cueva del complejo *Astyanax mexicanus*, migren a través de gran parte de la región de El Abra (Herman et al., 2018).

Estructura y Flujo Genético de las Poblaciones Troglodias de la Sierra de la Colmena

Para la Sierra de la Colmena, se ha propuesto que las poblaciones de Cueva del Río Subterráneo (este trabajo), la Cueva de Otates y la Cueva Lienzo (Elliot, 2018) están conectadas hidrológicamente **en el sistema de Micos**. Este sistema se localiza a 1.5 km de Ciudad Valles en San Luis Potosí, en la Sierra de San Dieguito, al este de la Sierra de la Colmena. Las cuevas de esta región están en un valle cercano a 10 km al suroeste de Micos, en el río Salto, el cual se transforma en el Río Valles después de fluir hacia la Sierra de Guatemala. El sistema hidrológico recibe el flujo del Arroyo La Pagua.

Según nuestros resultados, la población de cueva de Subterráneo presenta evidencia de entremezcla con las poblaciones de superficie del linaje 1 (i.e. linaje nuevo), y con algunas poblaciones de cueva de la Sierra de El Abra (fig. 4 B). Hasta la fecha, no se ha reportado ninguna conexión subterránea entre la Sierra de El Abra y la Sierra de La Colmena. Adicionalmente, la evidencia de flujo génico contemporáneo entre poblaciones de cuevas alopatricas es escasa (Trontelj, 2018). La mayoría de los casos conocidos de migración reciente entre poblaciones no vecinas conllevan la posibilidad de dispersión a través de la superficie, incluso si la especie es un habitante de cuevas obligado (Trontelj, 2018). Dada la gran separación geográfica entre La Sierra de El Abra y la Sierra de La Colmena, la dispersión entre individuos de cuevas de ambas regiones es muy poco probable.

La evidencia de entremezcla de la población de cueva Subterráneo, con las poblaciones de superficie de linaje 1 y las poblaciones de cueva de la Sierra de El Abra, puede explicarse a través de los procesos de deriva y flujo génico. Durante su existencia, las población de cueva tienden a perder parte de su diversidad genética debido a los procesos de deriva; esto está controlado por factores endógenos como el tamaño efectivo, el sistema de apareamiento y la demografía (Trontelj, 2018). Por otra parte, la configuración genética de las poblaciones de la cueva de *Astyanax* resulta en parte del intercambio genético en curso con las poblaciones de la superficie. La población de Subterráneo recibe un gran número de migrantes por parte de las poblaciones de superficie y se ha observado la presencia

de fenotipos intermedios, resultado de la introgresión con las formas epígeas (Dowling et al., 2002; Bradic et al., 2012; Bibliowicz et al., 2013). La disminución del tamaño efectivo en las poblaciones de cueva, aumenta la probabilidad de fijar los alelos en pocas generaciones (Kimura et al., 1969), por lo que, el pool genético de la población de cueva de Subterráneo podría ser un subconjunto de los alelos de las poblaciones de superficie debido al flujo génico, como ha sido propuesto previamente (Bradic et al., 2012), explicando además la evidencia de introgresión de estos alelos en las poblaciones de cueva de la Sierra de El Abra.

En trabajos futuros, la colecta de individuos de las poblaciones de cueva de Otates y Lienzo permitirá evaluar la conectividad entre estas poblaciones con la cueva de Subterráneo, así como la presencia de migrantes por parte de las poblaciones de superficie.

Incongruencia Filogenética, Evolución Reticular y Flujo Génico Histórico en *A. mexicanus*

Una de las premisas centrales de la filogenética es que existe un patrón jerárquico en la relación de los organismos, que se puede inferir al observar y analizar caracteres homólogos, moldeados por la historia evolutiva (Dávalos et al., 2012). Aunque la mayoría de los caracteres de un organismo comparten una historia evolutiva común, las tasas de cambio diferenciales y los mecanismos evolutivos que impulsan esas tasas, producen filogenias conflictivas -filogenias inferidas que se diferencian entre sí en uno o más caracteres- (Wendell & Doyle, 1998; Dávalos et al., 2012). Con la proliferación de nuevas herramientas moleculares, estos conflictos filogenéticos se han vuelto más evidentes, y con frecuencia, son la norma más que la excepción (Wendell & Doyle, 1998; Dávalos et al., 2012).

En *Astyanax mexicanus*, existen claras diferencias entre los árboles obtenidos a partir de caracteres morfológicos (Schmitter-Soto, 2016; Schmitter-Soto, 2017), mitocondriales (Strecker et al., 2003; Strecker et al., 2004; Ornelas-García et al., 2008; Strecker et al., 2012), nucleares, (Ornelas-García et al., 2008; Bradic et al., 2012) y genómicos (Coghill et al., 2014; Herman et al.,

2018; este trabajo). Esta incongruencia filogenética puede reflejar uno o más procesos biológicos subyacentes, como eventos de hibridación, introgresión, sorteo incompleto de linajes, etc. (Wendell & Doyle, 1998; Mallet et al., 2016). Adicionalmente, en especies con divergencia reciente, a menudo se presentan barreras reproductivas incompletas, por lo que éstas pueden llegar a hibridar en simpatria (Fontaine et al., 2015). La introgresión entre las poblaciones de cueva y superficie en *Astyanax* se ha reportado ampliamente en la literatura (Strecker et al., 2004; Bradic et al., 2012; Strecker et al., 2012) y junto con los tiempos rápidos de divergencia estimados (Fumey et al., 2018; Herman et al., 2018), es probable que estas sean las causas más comunes de la incongruencia filogenética en este grupo.

Dado la compleja historia evolutiva de *A. mexicanus*, se evaluaron en este trabajo las relaciones de ancestría del complejo, construyendo una red filogenética utilizando el método ‘Neighbor-Net’ (Bryant et al., 2006) implementado en SplitsTree4 (Huson, 2006). En esta red se encontró un patrón claramente no jerarquizado (esto es, en forma de árbol filogenético), que señala posibles eventos de introgresión entre taxones. Este patrón es similar al encontrado en otros taxones, y se asocia a la divergencia temprana de los grupos (Kelly et al., 2008; Lecaudey et al., 2018). En este sentido, es posible que el patrón reticular de la red filogenética sea evidencia indirecta de un proceso de divergencia reciente entre los linajes, como se ha propuesto en trabajos previos (Fumey et al., 2018, Herman et al., 2018).

También existe evidencia de patrones evolutivos reticulados mucho más profundos, al considerar los eventos históricos de introgresión. En algunos taxones, se ha inferido que estos eventos afectan gran parte del genoma, generando conflicto filogenético, porque tienden a reducir la divergencia entre los taxa (Fontaine et al., 2015; Mallet et al., 2016). A través del análisis de Treemix, hemos podido recuperar al menos cuatro eventos de flujo genético ancestral en *Astyanax mexicanus* (fig. 5 A). Al eliminar los individuos con evidencia de entremezcla en el análisis de Admixture (fig 4 B), pudimos visualizar la naturaleza ancestral de los eventos de flujo génico ancestral, los cuales están indicados por la posición de la flecha de migración a lo largo de las ramas (fig. 5 A), en lugar de las

puntas, en la reconstrucción. Estos eventos de flujo génico ancestral muestran que los eventos de colonización por parte de los peces de superficie hacia los ambientes de cueva han ocurrido en múltiples momentos de la historia evolutiva, pero en este trabajo fue posible poner un marco temporal de estos eventos de flujo génico, con respecto a la divergencia de las diferentes poblaciones. Adicionalmente, el peso de los eventos de migración (fig 5, A) entre las diferentes poblaciones no es similar, lo que significa que algunas poblaciones de cueva han sido más propensas que otras a la entrada de peces de superficie.

Para el sistema de la región norte de la Sierra de El Abra (Cueva de Pachón, cluster 6), se ha reportado con anterioridad la presencia de eventos de hibridación que involucran poblaciones de superficie de linaje nuevo, mediante observaciones de campo (Avice & Selander, 1972; Mitchell et al., 1977) y datos moleculares (Dowling et al., 2002; Strecker et al., 2003; Ornelas-García et al., 2008; Ornelas-García & Pedraza-Lara, 2015; Strecker et al., 2012; Herman et al., 2018).

Para la región sur de la Sierra de El Abra, tanto la prueba de entremezcla con ADMIXTURE, como la evaluación del flujo génico ancestral en Treemix, sugiere la evidencia de entremezcla entre las poblaciones al sur de la Sierra de El Abra (poblaciones de Chica, Chiquitita y Toro, cluster 3) con otras poblaciones de superficie del linaje 1. La población de cueva Chiquitita corresponde con la población de cueva más cercana al río Tampaón, y ésta a su vez se encuentra a una distancia de 1.25 km de la población de cueva Chica (Espinasa y Espinasa., 2020). Esto hace muy plausible que estas poblaciones puedan estar conectadas con la población de superficie de forma intermitente en la temporada de lluvias (Elliot 2018; Espinasa y Espinasa, 2020).

También se encontró evidencia de que las poblaciones de *Astyanax* al sur del TMVB han cruzado esa barrera geográfica. En otros trabajos también se ha encontrado un patrón de introgresión similar entre las poblaciones de *A. mexicanus* y *A. aeneus*, en las cuencas del Pánuco y del Papaloapan

(Strecker et al., 2011; Pérez-Rodríguez et al., 2021). Estos resultados, junto con los de este trabajo, dan evidencia de la mezcla genética entre poblaciones, a pesar del obstáculo físico que presenta el TMVB.

En este trabajo no encontramos evidencia de flujo génico ancestral entre las diferentes regiones de las poblaciones de cuevas. Este resultado contrasta con el de Herman y colaboradores (2018), quienes habían reportado el intercambio genético entre las poblaciones de Cueva del Molino (Sierra de Guatemala, linaje 1) y cueva de La Tinaja (Sierra de El Abra, linaje 2), y sugieren que éste habría surgido como producto de conexiones subterráneas entre las estas cuevas, a pesar de las considerables distancias geográficas (170 km de distancia aprox.).

A pesar de la evidencia de los eventos de flujo génico entre las poblaciones de cueva y de superficie, el que la identidad fenotípica de los peces de cueva se ha mantenido en éstas a pesar del contacto secundario y la hibridación es notable. Se ha propuesto en diferentes trabajos que la evolución regresiva en *Astyanax mexicanus* puede explicarse ya sea a través de procesos de selección natural, eventos de mutación recurrente, con deriva genética y pleiotropía (Protas et al., 2007; Borowsky, 2015; Cartwright et al., 2017). Estas fuerzas directrices del fenotipo de cuevas no son mutuamente excluyentes. Se ha probado que la pérdida de la visión se ha producido a través de procesos de selección (Yamamoto et al., 2004; Borowsky, 2015; Wilkens 2020), mientras que la disminución en el número de melanóforos se debió principalmente la combinación de mutaciones recurrentes con deriva genética, o indirectamente a través de pleiotropía (Protas et al., 2007, Borowsky, 2015). Sin embargo, contar con una hipótesis filogenética robusta, podrá dar marco a los procesos y patrones que han dado lugar a la evolución de las formas troglobias en este género.

Conclusiones

- La hipótesis filogenómica generada en este trabajo sugiere que la evolución de las poblaciones de cuevas de *A. mexicanus* ha ocurrido en tres ocasiones distintas, de manera parafilética (en dos linajes diferentes).
- La estructura genética de las poblaciones de cueva de *A. mexicanus* está fuertemente influenciada por los sistemas hidrológicos (i.e. poblaciones de cueva geográficamente cercanas, conectadas a través de pasajes subterráneos), descritos previamente para la Sierra de Guatemala, la Sierra de El Abra y la Sierra de la Colmena (Elliot, 2018), lo cual corresponde con un claro patrón biogeográfico. En algunos casos, es posible observar la evidencia de entremezcla entre dichos sistemas, la cual pueden explicarse a través de conexiones intermitentes asociadas a pulsos hidrológicos en los sistemas.
- El flujo génico entre las poblaciones de cueva y superficie ha ocurrido en múltiples ocasiones a lo largo de la historia evolutiva del complejo. Éste es distinto entre diferentes poblaciones y el peso de este flujo está determinado en cierta medida de la cercanía de las poblaciones de superficie a la entrada de las cuevas (como en el caso de la población de la cueva de Chica).

Referencias Bibliográficas

- Alexander, D. H., Shringarpure, S. S., Novembre, J., & Lange, K. (2015). **Admixture 1.3 software manual**. Los Angeles: UCLA Human Genetics Software Distribution.
- Alvarez, J. (1946). **Revisión del género Anoptichthys con descripción de una especie nueva (Pisc., Characidae)**. An Esc Nac Cienc Biol Mex, 4, 263-282.
- Alvarez, J. (1947). **Descripción de Anoptichthys hubbsi caracínido ciego de La Cueva de Los Sabinos**. SLP Rev. Soc. Mex. Hist. Nat, 8(1-4), 215-219.
- Andrews, S. (2010). **FastQC: a quality control tool for high throughput sequence data**.
- Avise, J. C., & Selander, R. K. (1972). **Evolutionary genetics of cave-dwelling fishes of the genus Astyanax**. Evolution, 26(1), 1-19.
- Barrett, R. D., & Schluter, D. (2008). **Adaptation from standing genetic variation**. *Trends in ecology & evolution*, 23(1), 38-44.
- Beron, P. (2018). **Cave Arachnida**. In *Zoogeography of Arachnida* (pp. 833-851). Springer, Cham.
- Bilandžija, H., Hollifield, B., Steck, M., Meng, G., Ng, M., Koch, A. D., Gračan, R., Četković, H., Porter, M. L., Renner, K. J., & Jeffery, W. (2020). **Phenotypic plasticity as a mechanism of cave colonization and adaptation**. ELife, 9. <https://doi.org/10.7554/eLife.51830>
- Bibliowicz, J., Alié, A., Espinasa, L., Yoshizawa, M., Blin, M., Hinaux, H., ... & Rétaux, S. (2013). **Differences in chemosensory response between eyed and eyeless Astyanax mexicanus of the Rio Subterráneo cave**. EvoDevo, 4(1), 1-6.
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). **Trimmomatic: a flexible trimmer for Illumina sequence data**. Bioinformatics, 30(15), 2114-2120.
- Borowsky, R. (2008). **Astyanax mexicanus, the blind Mexican cave fish: A model for studies in development and morphology**. *Cold Spring Harbor Protocols*, 2008(11), pdb-emo107.

- Borowsky, R., Keene, A., Yoshizawa, M., & McGaugh, S. E. (2015). **Regressive evolution: Testing hypotheses of selection and drift**. En *Keene A, Yoshizawa, M. McGaugh, SE, editor. Biology and Evolution of the Mexican Cavefish San Diego: Elsevier*, 93.
- Borowsky, R. (2018). **Cavefishes**. *Current Biology*, 28(2), R60–R64.
<https://doi.org/10.1016/j.cub.2017.12.011>
- Boudriot, F., & Reutter, K. (2001). **Ultrastructure of the taste buds in the blind cave fish *Astyanax jordani* (“Anoptichthys”) and the sighted river fish *Astyanax mexicanus* (Teleostei, Characidae)**. *Journal of Comparative Neurology*, 434(4), 428-444.
- Bradic, M., Beerli, P., García-De León, F. J., Esquivel-Bobadilla, S., & Borowsky, R. L. (2012). **Gene flow and population structure in the Mexican blind cavefish complex (*Astyanax mexicanus*)**. *BMC Evolutionary Biology*, 12(1), 9. <https://doi.org/10.1186/1471-2148-12-9>
- Bradic, M., Teotónio, H., & Borowsky, R. L. (2013). **The population genomics of repeated evolution in the blind cavefish *Astyanax mexicanus***. *Molecular Biology and Evolution*, 30(11), 2383–2400. <https://doi.org/10.1093/molbev/mst136>
- Cartwright, R. A., Schwartz, R. S., Merry, A. L., & Howell, M. M. (2017). **The importance of selection in the evolution of blindness in cavefish**. *BMC Evolutionary Biology*, 17(1), 1–14.
<https://doi.org/10.1186/s12862-017-0876-4>
- Chávez-Solís, E. M., Mejía-Ortiz, L. M., & Simões, N. (2018). **Predatory behavior of the cave shrimp *Creaseria morleyi* (Creaser, 1936)(Caridea: Palaemonidae), the blind hunter of the Yucatán cenotes, Mexico**. *Journal of Crustacean Biology*, 38(1), 1-7.
- Coghill, L. M., Darrin Hulsey, C., Chaves-Campos, J., García de León, F. J., & Johnson, S. G. (2014). **Next generation phylogeography of cave and surface *Astyanax mexicanus***. *Molecular Phylogenetics and Evolution*, 79(1), 368–374. <https://doi.org/10.1016/j.ympev.2014.06.029>
- Christiansen, K. (2012). **Morphological adaptations**. En *Encyclopedia of caves* (pp. 517-528). Academic Press.

- Culver, D. C., Kane, T., & Fong, D. W. (1995). **Adaptations and natural selection in caves and wells.** *J. Caves Karst Stud*, 62, 11-17.
- Culver, D. C., Holsinger, J. R., Christman, M. C., & Pipan, T. (2010). **Morphological differences among eyeless amphipods in the genus *Stygobromus* dwelling in different subterranean habitats.** *Journal of Crustacean Biology*, 30(1), 68-74.
- Dávalos, L. M., Cirranello, A. L., Geisler, J. H., & Simmons, N. B. (2012). **Understanding phylogenetic incongruence: lessons from phyllostomid bats.** *Biological Reviews*, 87(4), 991-1024.
- Danecek, P., Schiffels, S., & Durbin, R. (2014). **Multiallelic calling model in bcftools (-m).**
- Derkarabetian, S., Steinmann, D. B., & Hedin, M. (2010). **Repeated and time-correlated morphological convergence in cave-dwelling harvestmen (Opiliones, Laniatores) from montane western North America.** *PLoS One*, 5(5), e10388.
- Dowling, T. E., Martasian, D. P., & Jeffery, W. R. (2002). **Evidence for multiple genetic forms with similar eyeless phenotypes in the blind cavefish, *Astyanax mexicanus*.** *Molecular Biology and Evolution*, 19(4), 446–455. <https://doi.org/10.1093/oxfordjournals.molbev.a004100>
- Elliott, W. R. (2018). **The *Astyanax* Caves of Mexico: Cavefishes of Tamaulipas, San Luis Potosí, and Guerrero.** Association for Mexican Cave Studies.
- Elipot, Y., Hinaux, H., Callebert, J., Launay, J. M., Blin, M., & Rétaux, S. (2014). **A mutation in the enzyme monoamine oxidase explains part of the *Astyanax* cavefish behavioural syndrome.** *Nature communications*, 5(1), 1-11.
- Erckens, W., & Martin, W. (1982). **Exogenous and endogenous control of swimming activity in *Astyanax mexicanus* (Characidae, Pisces) by direct light response and by a circadian oscillator I. Analyses of the time-control systems of an epigeal river population.** *Zeitschrift für Naturforschung C*, 37(11-12), 1253-1265.

- Erckens W, Martin W. (1982b). **Exogenous and endogenous control of swimming activity in *Astyanax mexicanus* (Characidae, Pisces) by direct light response and by a circadian oscillator. II. Features of time-controlled behavior of a cave population and their comparison to an epigeal ancestral form.** *Z Naturforsch.* 37:1266–1273.
- Espinasa, L., Rivas-Manzano, P., & Pérez, H. E. (2001). **A new blind cave fish population of genus *Astyanax*: Geography, morphology and behavior.** *Environmental Biology of Fishes*, 62(1–3), 339–344. <https://doi.org/10.1023/A:1011852603162>
- Espinasa, L., Flick, C., & Giribet, G. (2007). **Phylogeny of the American silverfish Cubacubaninae (Hexapoda: Zygentoma: Nicoletiidae): a combined approach using morphology and five molecular loci.** *Cladistics*, 23(1), 22-40.
- Espinasa, L., Bartolo, N. D., & Newkirk, C. E. (2014). **DNA sequences of troglobitic nicoletiid insects support Sierra de El Abra and the Sierra de Guatemala as a single biogeographical area: Implications for *Astyanax*.** *Subterranean Biology*, 13, 35.
- Espinasa, L., & Espinasa, M. (2016). **Hydrogeology of Caves in the Sierra de El Abra Region.** En *Biology and Evolution of the Mexican Cavefish* (Issue 1977). Elsevier Inc. <https://doi.org/10.1016/B978-0-12-802148-4.00002-5>
- Espinasa, L., Legendre, L., Fumey, J., Blin, M., Rétaux, S., & Espinasa, M. (2018). **A new cave locality for *Astyanax* cavefish in Sierra de El Abra, Mexico.** *Subterranean Biology*, 26, 39–53. <https://doi.org/10.3897/subtbiol.26.26643>
- Espinasa, L., Ornelas-García, C. P., Legendre, L., Rétaux, S., Best, A., Gamboa-Miranda, R., Espinosa-Pérez, H., & Sprouse, P. (2020). **Discovery of Two New *Astyanax* Cavefish Localities Leads to Further Understanding of the Species Biogeography.** *Diversity*, 12(10), 368. <https://doi.org/10.3390/d12100368>
- Felsenstein, J. (1981). **Evolutionary trees from gene frequencies and quantitative characters: finding maximum likelihood estimates.** *Evolution*, 1229-1242.

- Fitak, R. (2019). **optM: an R package to optimize the number of migration edges using threshold models**. *Journal of Heredity*.
- Fontaine, M. C., Pease, J. B., Steele, A., Waterhouse, R. M., Neafsey, D. E., Sharakhov, I. V., ... & Besansky, N. J. (2015). **Extensive introgression in a malaria vector species complex revealed by phylogenomics**. *Science*, 347(6217).
- Fumey, J., Hinaux, H., Noirot, C., Thermes, C., Rétaux, S., & Casane, D. (2018). **Evidence for late Pleistocene origin of *Astyanax mexicanus* cavefish**. *BMC Evolutionary Biology*, 18(1), 1–19. <https://doi.org/10.1186/s12862-018-1156-7>
- Gallo, N. D., & Jeffery, W. R. (2012). **Evolution of space dependent growth in the teleost *Astyanax mexicanus***. *PLoS One*, 7(8), e41443.
- Gladstone, N. S., Carter, E. T., McKinney, M. L., & Niemiller, M. L. (2018). **Status and Distribution of the Cave-Obligate Land Snails in the Appalachians and Interior Low Plateau of the Eastern United States**. *American Malacological Bulletin*, 36(1), 62-78.
- Gonzalez, B. C., Worsaae, K., Fontaneto, D., & Martínez, A. (2018). **Anophthalmia and elongation of body appendages in cave scale worms (Annelida: Aphroditiformia)**. *Zoologica Scripta*, 47(1), 106-121.
- Gross, J. B. (2012). **The complex origin of *Astyanax* cavefish**. *BMC Evolutionary Biology*, 12(1). <https://doi.org/10.1186/1471-2148-12-105>
- Gross, J. B., & Wilkens, H. (2013). **Albinism in phylogenetically and geographically distinct populations of *Astyanax* cavefish arises through the same loss-of-function *Oca2* allele**. *Heredity*, 111(2), 122–130. <https://doi.org/10.1038/hdy.2013.26>
- Gross, J. B., Meyer, B., & Perkins, M. (2015). **The rise of *Astyanax* cavefish**. *Developmental Dynamics*, 244(9), 1031–1038. <https://doi.org/10.1002/dvdy.24253>
- Gruber, B., & Georges, A. (2019). **dartR: Importing and analysing SNP and silicodart data generated by genome-wide restriction fragment analysis**. R package version 1.1. 11.

- Hausdorf, B., Wilkens, H., & Strecker, U. (2011). **Population genetic patterns revealed by microsatellite data challenge the mitochondrial DNA based taxonomy of *Astyanax* in Mexico (Characidae, Teleostei)**. *Molecular Phylogenetics and Evolution*, 60(1), 89–97.
<https://doi.org/10.1016/j.ympev.2011.03.009>
- Herman, A., Brandvain, Y., Weagley, J., Jeffery, W. R., Keene, A. C., Kono, T. J. Y., Bilandžija, H., Borowsky, R., Espinasa, L., O'Quin, K., Ornelas-García, C. P., Yoshizawa, M., Carlson, B., Maldonado, E., Gross, J. B., Cartwright, R. A., Rohner, N., Warren, W. C., & McGaugh, S. E. (2018). **The role of gene flow in rapid and repeated evolution of cave-related traits in Mexican tetra, *Astyanax mexicanus***. *Molecular Ecology*, 27(22), 4397–4416.
<https://doi.org/10.1111/mec.14877>
- Hinaux, H., Poulain, J., Da Silva, C., Noirot, C., Jeffery, W. R., Casane, D., & Rétaux, S. (2013). **De novo sequencing of *Astyanax mexicanus* surface fish and Pachón cavefish transcriptomes reveals enrichment of mutations in cavefish putative eye genes**. *PloS one*, 8(1), e53553.
- Hinaux, H., Rétaux, S., & Elipot, Y. (2016). **Social Behavior and Aggressiveness in *Astyanax***. In **Biology and Evolution of the Mexican Cavefish**. Elsevier Inc.
<https://doi.org/10.1016/B978-0-12-802148-4.00017-7>
- Howarth, F. G., & Moldovan, O. T. (2018). **The ecological classification of cave animals and their adaptations**. In **Cave ecology** (pp. 41-67). Springer, Cham.
- Howarth, F. G. (2019). **Adaptive shifts**. In *Encyclopedia of caves* (pp. 47-55). Academic Press.
- Hubbs, C. L., & Innes, W. T. (1936). **The first known blind fish of the family Characidae: a new genus from Mexico**.
- Hüppop, K. (1986). **Oxygen consumption of *Astyanax fasciatus* (Characidae, Pisces): a comparison of epigean and hypogean populations**. *Environmental Biology of Fishes*, 17(4), 299-308.
- Imam, T. S., Abubakar, K. I., Rabi, S., & Rabi, A. T. (2020) **Phenotypic Plasticity in Response to Ecological Forces: A Review**.

- Jeffery, W. R., & Martasian, D. P. (1998). **Evolution of eye regression in the cavefish *Astyanax*: apoptosis and the Pax-6 gene.** *American Zoologist*, 38(4), 685-696.
- Jeffery, W. R. (2001). **Cavefish as a model system in evolutionary developmental biology.** *Developmental Biology*, 231(1), 1–12. <https://doi.org/10.1006/dbio.2000.0121>
- Jeffery, W. R. (2008). **Emerging model systems in evo-devo: cavefish and microevolution of development.** *Evolution & Development*, 10(3), 265–272.
<https://doi.org/10.1111/j.1525-142X.2008.00235.x>
- Jeffery, W. R. (2009). **Regressive evolution in *Astyanax cavefish*.** *Annual Review of Genetics*, 43, 25–47. <https://doi.org/10.1146/annurev-genet-102108-134216>
- Jombart, T., Devillard, S., & Balloux, F. (2010). **Discriminant analysis of principal components: a new method for the analysis of genetically structured populations.** *BMC genetics*, 11(1), 94.
- Jombart, T., & Ahmed, I. (2011). **adegenet 1.3-1: new tools for the analysis of genome-wide SNP data.** *Bioinformatics*, 27(21), 3070-3071.
- Jombart, T., & Collins, C. (2015). **A tutorial for discriminant analysis of principal components (DAPC) using adegenet 2.0.0.** *London: Imperial College London, MRC Centre for Outbreak Analysis and Modelling.*
- Kelly, C. M., Barker, N. P., Villet, M. H., Broadley, D. G., & Branch, W. R. (2008). **The snake family Psammophiidae (Reptilia: Serpentes): phylogenetics and species delimitation in the African sand snakes (*Psammophis* Boie, 1825) and allied genera.** *Molecular Phylogenetics and Evolution*, 47(3), 1045-1060.
- Keene, A., Yoshizawa, M., & McGaugh, S. E. (2015). *Biology and evolution of the Mexican cavefish.* Academic Press.
- Kimura, M., & Ohta, T. (1969). **The average number of generations until extinction of an individual mutant gene in a finite population.** *Genetics*, 63(3), 701.

- Knaus, B. J., & Grünwald, N. J. (2017). **vcfr: a package to manipulate and visualize variant call format data in R**. *Molecular ecology resources*, 17(1), 44-53.
- Kowalko, J. E. (2013). **The genetic basis of behavior in the blind Mexican cavefish, *Astyanax mexicanus***. 239. <https://dash.harvard.edu/handle/1/11181222>
- Krishnan, J., & Rohner, N. (2017). **Cavefish and the basis for eye loss**. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1713), 20150487.
- Lawson, D. J., Van Dorp, L., & Falush, D. (2018). **A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots**. *Nature Communications*, 9(1), 1-11.
- Lepais, O., & Weir, J. T. (2014). **Sim RAD: an R package for simulation-based prediction of the number of loci expected in RAD seq and similar genotyping by sequencing approaches**. *Molecular ecology resources*, 14(6), 1314-1321.
- Li, H., & Durbin, R. (2009). **Fast and accurate short read alignment with Burrows–Wheeler transform**. *bioinformatics*, 25(14), 1754-1760.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... & Durbin, R. (2009). **The sequence alignment/map format and SAMtools**. *Bioinformatics*, 25(16), 2078-2079.
- Li, H., & Durbin, R. (2010). **Fast and accurate long-read alignment with Burrows–Wheeler transform**. *Bioinformatics*, 26(5), 589-595.
- Lunghi, E., & Bruni, G. (2018). **Long-term reliability of Visual Implant Elastomers in the Italian cave salamander (*Hydromantes italicus*)**. *Salamandra*, 54(4), 283-286.
- Ma, L., Strickler, A. G., Parkhurst, A., Yoshizawa, M., Shi, J., & Jeffery, W. R. (2018). **Maternal genetic effects in *Astyanax* cavefish development**. *Developmental Biology*, 441(2), 209–220. <https://doi.org/10.1016/j.ydbio.2018.07.014>

- Mallet, J., Besansky, N., & Hahn, M. W. (2016). **How reticulated are species?**. *BioEssays*, 38(2), 140-149.
- Mammola, S., & Isaia, M. (2017). **Spiders in caves**. *Proceedings of the Royal Society B: Biological Sciences*, 284(1853), 20170193.
- Maddison, W. P., & Maddison, D. R. (2019). **Mesquite: a modular system for evolutionary analysis**. Version 3.51. 2018.
- McGaugh, S. E., Gross, J. B., Aken, B., Blin, M., Borowsky, R., Chalopin, D., ... & Minx, P. (2014). **The cavefish genome reveals candidate genes for eye loss**. *Nature communications*, 5(1), 1-10.
- McGaugh, S. E., Kowalko, J. E., Duboué, E., Lewis, P., Franz-Odenaal, T. A., Rohner, N., Gross, J. B., & Keene, A. C. (2020). **Dark world rises: The emergence of cavefish as a model for the study of evolution, development, behavior, and disease**. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, December 2019, 1–8.
<https://doi.org/10.1002/jez.b.22978>
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., ... & DePristo, M. A. (2010). **The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data**. *Genome research*, 20(9), 1297-1303.
- Mercado-Silva, Norman; Ornelas-García, Claudia P.; Schmitter-soto, Juan Jacobo; Gidmark, Nicolas J.; Simons, Andrew M. (2002) **Characidae: Characins** en *Freshwater Fishes of North America: Volume 2: Characidae to Poeciliidae* (pp. 2-22) Johns Hopkins University Press.
- Miller, R. R., Minckley, W. L., Soto, S., & Jacobotr, J. (2009). **Peces dulceacuícolas de México** (No. EE/597.092972 M5).
- Miller, R. R., Minckley, W. L., & Norris, S. M. (2005). **Freshwater fishes of Mexico** (No. QL 629. M54 2005).

- Miller, M. A., Pfeiffer, W., & Schwartz, T. (2012, July). **The CIPRES science gateway: enabling high-impact science for phylogenetics researchers with limited resources**. In *Proceedings of the 1st Conference of the Extreme Science and Engineering Discovery Environment: Bridging from the extreme to the campus and beyond* (pp. 1-8).
- Mitchell, R. W., Russell, W. H., & Elliott, W. R. (1977). **Mexican eyeless characin fishes, genus *Astyanax*: environment, distribution, and evolution**. Texas Tech Press.
- Moran, D., Softley, R., & Warrant, E. J. (2014). **Eyeless Mexican cavefish save energy by eliminating the circadian rhythm in metabolism**. *PloS one*, 9(9), e107877.
- Moran, D., Softley, R., & Warrant, E. J. (2015). **The energetic cost of vision and the evolution of eyeless Mexican cavefish**. *Science advances*, 1(8), e1500363.
- Ornelas-García, C. P., & Pedraza-Lara, C. (2016). **Phylogeny and Evolutionary History of *Astyanax mexicanus***. In *Biology and Evolution of the Mexican Cavefish* (Issue November). Elsevier Inc. <https://doi.org/10.1016/B978-0-12-802148-4.00004-9>
- Ornelas-García, C. P., Domínguez-Domínguez, O., & Doadrio, I. (2008). **Evolutionary history of the fish genus *Astyanax* baird & Girard (1854) (Actinopterygii, Characidae) in mesoamerica reveals multiple morphological homoplasies**. *BMC Evolutionary Biology*, 8(1), 1–17. <https://doi.org/10.1186/1471-2148-8-340>
- Orr, H. A., & Betancourt, A. J. (2001). **Haldane's sieve and adaptation from the standing genetic variation**. *Genetics*, 157(2), 875-884.
- Ortiz, E. M. (2019). **vcf2phylip v2.0: convert a VCF matrix into several matrix formats for phylogenetic analysis**. URL [https://doi.org/10.5281/zenodo, 2540861](https://doi.org/10.5281/zenodo.2540861).
- Pacheco, T. L., & Vaz-de-Mello, F. Z. (2019). **New dung beetle genus and species from a cave in the Espinhaço mountain range, Brazil (Coleoptera: Scarabaeidae: Scarabaeinae)**. *Journal of Natural History*, 53(19-20), 1247-1253.

- Passow, C. N., Kono, T. J., Stahl, B. A., Jaggard, J. B., Keene, A. C., & McGaugh, S. E. (2018). **RNA later and flash freezing storage methods nonrandomly influence observed gene expression in RNAseq experiments.** *bioRxiv*, 379834.
- Pérez-Rodríguez, R., Esquivel-Bobadilla, S., Orozco-Ruíz, A. M., Olivas-Hernández, J. L., & García-De León, F. J. (2021). **Genetic structure and historical and contemporary gene flow of *Astyanax mexicanus* in the Gulf of Mexico slope: a microsatellite-based analysis.** *PeerJ*, 9, e10784.
- Peter, B. M., Huerta-Sanchez, E., & Nielsen, R. (2012). **Distinguishing between selective sweeps from standing variation and from a de novo mutation.** *PLoS Genet*, 8(10), e1003011.
- Pickrell, J., & Pritchard, J. (2012). **Inference of population splits and mixtures from genome-wide allele frequency data.** *Nature Precedings*, 1-1.
- Poulson, T. L., & White, W. B. (1969). **The cave environment.** *Science*, 165(3897), 971-981.
- Price, T. D., Qvarnström, A., & Irwin, D. E. (2003). **The role of phenotypic plasticity in driving genetic evolution.** *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 270(1523), 1433-1440.
- Protas, M. E., Hersey, C., Kochanek, D., Zhou, Y., Wilkens, H., Jeffery, W. R., Zon, L. I., Borowsky, R., & Tabin, C. J. (2006). **Genetic analysis of cavefish reveals molecular convergence in the evolution of albinism.** *Nature Genetics*, 38(1), 107–111. <https://doi.org/10.1038/ng1700>
- Protas, M., Conrad, M., Gross, J. B., Tabin, C., & Borowsky, R. (2007). **Regressive Evolution in the Mexican Cave Tetra, *Astyanax mexicanus*.** *Current Biology*, 17(5), 452–454. <https://doi.org/10.1016/j.cub.2007.01.051>
- Protas, M., Tabansky, I., Conrad, M., Gross, J. B., Vidal, O., Tabin, C. J., & Borowsky, R. (2008). **Multi-trait evolution in a cave fish, *Astyanax mexicanus*.** *Evolution & development*, 10(2), 196-209.

- Protas, M. E., Trontelj, P., & Patel, N. H. (2011). **Genetic basis of eye and pigment loss in the cave crustacean, *Asellus aquaticus***. *Proceedings of the National Academy of Sciences*, *108*(14), 5702-5707.
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., ... & Sham, P. C. (2007). **PLINK: a tool set for whole-genome association and population-based linkage analyses**. *The American journal of human genetics*, *81*(3), 559-575.
- Quinn, T. P. (1980). **Locomotor responses of juvenile blind cave fish, *Astyanax jordani*, to the odors of conspecifics**. *Behavioral and neural biology*, *29*(1), 123-127.
- Romero, A. (2011). The Evolution of Cave Life: **New concepts are challenging conventional ideas about life underground**. *American Scientist*, *99*(2), 144-151.
- Rose, F. L., & Mitchell, R. W. (1982). **Comparative lipid values of epigeal and cave-adapted *Astyanax***. *The Southwestern Naturalist*, *27*(3), 357-358.
- Schemmel, C. (1967). **Vergleichende Untersuchungen an den Hautsinnesorganen ober-und unterirdisch lebender *Astyanax*-Formen**. *Zeitschrift für Morphologie der Tiere*, *61*(2), 255-316.
- Schmitter-Soto, J. J. (2016). **A phylogeny of *Astyanax* (Characiformes: Characidae) in Central and North America**. *Zootaxa*, *4109*(2), 101–130. <https://doi.org/10.11646/zootaxa.4109.2.1>
- Schmitter-Soto, J. J. (2017). **A revision of *Astyanax* (Characiformes: Characidae) in Central and North America, with the description of nine new species**. *Journal of Natural History*, *51*(23–24), 1331–1424. <https://doi.org/10.1080/00222933.2017.1324050>
- Sonnenberg, R., Nolte, A. W., & Tautz, D. (2007). **An evaluation of LSU rDNA D1-D2 sequences for their use in species identification**. *Frontiers in zoology*, *4*(1), 6.
- Stamatakis, A. (2014). **RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies**. *Bioinformatics*, *30*(9), 1312-1313.

- Stamatakis, A. (2016). **The RAxML v8. 2. X Manual**. Heidelberg Institute for Theoretical Studies. Available at: <https://cme.h-its.org/exelixis/resource/download/NewManual.pdf>.
- Strecker, U., Bernatchez, L., & Wilkens, H. (2003). **Genetic divergence between cave and surface populations of *Astyanax* in Mexico (Characidae, Teleostei)**. *Molecular Ecology*, 12(3), 699–710. <https://doi.org/10.1046/j.1365-294X.2003.01753.x>
- Strecker, Ulrike, Faúndez, V. H., & Wilkens, H. (2004). **Phylogeography of surface and cave *Astyanax* (Teleostei) from Central and North America based on cytochrome b sequence data**. *Molecular Phylogenetics and Evolution*, 33(2), 469–481. <https://doi.org/10.1016/j.ympev.2004.07.001>
- Strecker, Ulrike, Hausdorf, B., & Wilkens, H. (2012). **Parallel speciation in *Astyanax* cave fish (Teleostei) in Northern Mexico**. *Molecular Phylogenetics and Evolution*, 62(1), 62–70. <https://doi.org/10.1016/j.ympev.2011.09.005>
- Terán, G. E., Benitez, M. F., & Mirande, J. M. (2020). **Opening the Trojan horse: phylogeny of *Astyanax*, two new genera and resurrection of *Psalidodon* (Teleostei: Characidae)**. *Zoological Journal of the Linnean Society*, 1–18. <https://doi.org/10.1093/zoolinlean/zlaa019>
- Team, R. C. (2018). **R: A language and environment for statistical computing**.
- Trajano, E. (2012). **Ecological classification of subterranean organisms**. *Encyclopedia of caves, 1*, 275-277.
- Trontelj, P., Blejec, A., & Fišer, C. (2012). **Ecomorphological convergence of cave communities**. *Evolution: International Journal of Organic*
- Trontelj, P. (2018). **Structure and genetics of cave populations**. en *Cave ecology* (pp. 269-295). Springer, Cham.
- Van Trump, W. J., & McHenry, M. J. (2013). **The lateral line system is not necessary for rheotaxis in the Mexican blind cavefish (*Astyanax fasciatus*)**.

- Via, S. (2017). 3. **The Evolution of Phenotypic Plasticity: What Do We Really Know?**. In *Ecological genetics* (pp. 35-57). Princeton University Press.
- Waddington, C. H. (1953). **Genetic assimilation of an acquired character**. *Evolution*, 118-126.
- Wendel J.F., Doyle J.J. (1998) **Phylogenetic Incongruence: Window into Genome History and Molecular Evolution**. In: Soltis D.E., Soltis P.S., Doyle J.J. (eds) *Molecular Systematics of Plants II*. Springer, Boston, MA.
- Wiley, S., & Mitchell, R. W. (1971). **A bibliography of the Mexican eyeless characin fishes of the genus *Astyanax***. In *Studies of the Cavernicole Fauna of Mexico*, AMCS Bulletin 4 (Eds. Reddell JR, Mitchell RW), Austin, TX: The Speleo Press. p 231–239.
- Wilkens, H. (1971). **Genetic interpretation of regressive evolutionary processes: studies on hybrid eyes of two *Astyanax* cave populations (Characidae, Pisces)**. *Evolution*, 530-544.
- Wilkens, H. (1988). **Evolution and Genetics of Epigean and Cave *Astyanax fasciatus* (Characidae, Pisces)**. In *Evolutionary Biology* (pp. 271–367). Springer US.
https://doi.org/10.1007/978-1-4613-1043-3_8
- Wilkens, H. (2020). **The role of selection in the evolution of blindness in cave fish**. *Biological Journal of the Linnean Society*, 130(3), 421-432.
- Yamamoto, Y., Stock, D. W., & Jeffery, W. R. (2004). **Hedgehog signalling controls eye degeneration in blind cavefish**. *Nature*, 431(7010), 844-847.
- Yoshizawa, M., Gorički, Š., Soares, D., & Jeffery, W. R. (2010). **Evolution of a behavioral shift mediated by superficial neuromasts helps cavefish find food in darkness**. *Current Biology*, 20(18), 1631-1636.
- Yoshizawa, M., Yamamoto, Y., O'Quin, K. E., & Jeffery, W. R. (2012). **Evolution of an adaptive behavior and its sensory receptors promotes eye regression in blind cavefish**. *BMC biology*, 10(1), 108.

Zwickl, D. J., & Hillis, D. M. (2002). **Increased taxon sampling greatly reduces phylogenetic error.**
Systematic biology, 51(4), 588-598.