

LA FILOGENÓMICA COMO HERRAMIENTA FUNDAMENTAL EN EL ESTUDIO DE LA EVOLUCIÓN BIOLÓGICA

PHYLOGENOMICS AS FUNDAMENTAL TOOL IN THE STUDY OF BIOLOGICAL EVOLUTION

TONATIUH RAMÍREZ-REYES

Posgrado en Ciencias Biológicas, Universidad Nacional Autónoma de México. Circuito de posgrados, Ciudad Universitaria, Coyoacán, 04510 Ciudad de México, México

Museo de Zoología, Facultad de Ciencias, Universidad Nacional Autónoma de México

Correspondence: trrecolgia@gmail.com

Abstract.— Sequencing technologies have allowed the accessibility of genomic data, making approaches from multilocus to genomic scales common practice in biodiversity studies. Phylogenomics has become a fundamental tool across a diversity of biological disciplines, since it allows a greater resolution in obtaining phylogenies at different scales (deep or shallow), shedding light on processes involved in the evolution of any taxon. Specifically in biogeography, phylogenomics has allowed in detail investigations into the geographic evolution of species with broad distributions. On the other hand, phylogenomics also generates evolutionary models that recognize the probable mechanisms of cryptic diversification in taxa with little morphological disparity. In this review, I show the profound impact of phylogenomics across different biological disciplines, as well as explain some particularities of this discipline, its practical applications, and the potential that phylogenomic models have to answer complex questions in evolutionary biology.

Keywords.— Biodiversity, cryptic species, genomics, RADseq, sequence capture.

Resumen.— Las tecnologías de secuenciación han permitido la obtención de datos genómicos, por lo que las aproximaciones multilocus a escalas genómicas en los estudios de la biodiversidad son ya una práctica común. La filogenómica se ha convertido en una herramienta fundamental en estudios de diversas disciplinas biológicas, ya que permite una resolución mayor en la obtención de filogenias a distintas escalas (profundas o recientes), evidenciando procesos involucrados en la evolución de cualquier taxón. Específicamente en biogeografía, ha permitido conocer detalles en la evolución geográfica de especies con amplia distribución, la filogenómica, también genera modelos evolutivos sobre los cuales se puede reconocer la diversidad críptica y los probables mecanismos que la generan en múltiples taxones que presentan escasa disparidad morfológica. En esta revisión, trato de mostrar el profundo impacto de la filogenómica en distintas disciplinas biológicas, así como algunas particularidades de los fundamentos de esta disciplina, las aplicaciones prácticas que dependen enteramente de ella y el potencial de los estudios que incorporan modelos filogenómicos para responder preguntas complejas de la evolución biológica.

Palabras clave.— Biodiversidad, especies crípticas, genómica, Radseq, captura de secuencias.

DE LA FILOGENÉTICA A LA FILOGENÓMICA

A pesar de que el término filogenómica fue acuñado a finales de los años 90 y presentado como un método novedoso en la anotación de genes, rápidamente fue incorporado en las disciplinas biológicas involucradas en reconstruir el árbol de la vida (Phillipe et al., 2005; Bleidorn, 2017). Precisamente porque la filogenética

es la rama de la biología encargada del estudio de relaciones evolutivas entre taxones (con una metodología específica) y porque esta disciplina se encuentra vinculada estrechamente con la sistemática. A mediados de la década de los 2000, hubo un incremento dramático en los métodos de obtención de datos a escala genómica, debido a las mejoras sustanciales en las tecnologías y en las estrategias de secuenciación, lo que

permitió generar información genómica incluso en organismos no-modelo (Delsuc et al., 2005; Lemmon & Lemmon, 2013; Da Fonseca et al., 2016; Posada, 2016; Bleidorn, 2017). Este desarrollo en obtención de datos genómicos, transformó el campo de la filogenética molecular (basada en pocos loci, principalmente obtenidos mediante secuenciación Sanger) en la filogenómica (con miles de loci, incluso genomas completos), ya que de ella se obtuvieron de manera inmediata anotaciones funcionales en los árboles de genes y también se generaron árboles de especies con gran precisión (Bleidorn, 2017). Actualmente podemos definir a la filogenómica como el estudio de relaciones evolutivas (de genes, genomas, especies, linajes), basado en análisis comparativos filogenéticos de datos a escala genómica (Delsuc et al., 2005; Chan & Ragan, 2013). Los principales objetivos que se habían planteado inicialmente esta disciplina fueron dos muy importantes: 1) utilizar datos genómicos para inferir relaciones filogenéticas entre especies y obtener información sobre los mecanismos de evolución molecular y 2) comparar genomas de múltiples especies para inferir funciones putativas de genes y proteínas (Phillipe et al., 2005, Phillipe & Blanchete, 2007). Sin embargo, con el paso del tiempo no solo no se ha limitado a tratar de cumplir esos objetivos, sino que sus aplicaciones abarcan distintas áreas del conocimiento, pasando el umbral de “simplemente” elaborar árboles de especies con alta precisión y soporte estadístico, y se dirige hacia estudios cada vez más complejos (Pyron, 2015). Basta con mirar los avances de esta disciplina en unos pocos ejemplos: en el campo de la medicina, el exhaustivo estudio sobre la evolución de la virulencia (severidad o nocividad de un patógeno) en los virus, gracias al marco filogenómico y la obtención de genomas completos de virus, es posible comprender los mecanismos de evolución de la virulencia de distintos tipos de virus que han afectado de manera sustancial poblaciones de animales no humanos y humanos (virus del Nilo, influenza A de aves, virus del mixoma, virus de la enfermedad de Marek, VIH, ébola y virus del Zika). Sin un marco evolutivo, sería muy difícil (si no es que imposible) comprender los mecanismos que generan virulencia en los virus causantes de un gran número de muertes y enfermedades en todo el mundo. Este marco evolutivo es presentado a través de las filogenias de genomas completos de los virus y el análisis de mutaciones en los genomas de las distintas cepas de virus (genómica comparada p. ej. cambios en aminoácidos), lo cual nos permite comprender desde el origen (temporal y espacial) de los virus hasta la dinámica evolutiva (selección, mutación, deriva, adecuación) que permite cambios en el genotipo y/o fenotipo de los virus, incrementando o disminuyendo la virulencia según cada caso particular. Con estos análisis incluso se pueden predecir algunos aspectos de la dinámica de algunos virus y darles seguimiento en tiempo real (Geoghegan & Holmes, 2018). Evidentemente

en el área de estudios sobre sistemática molecular también ha tenido gran relevancia la filogenómica, a través de obtención de datos genómicos con la técnica de captura de exones (>2000 loci) se pudo demostrar la diversidad críptica que existe en el lacertilio australiano *Carlia amax*, además de delimitar especies con distintas líneas de evidencia (conjunto de datos y metodologías), comparar filogenias de DNA mitocondrial y DNA nuclear, así como obtener modelos demográficos de expansión poblacional, e incluso determinar el probable origen geográfico de los linajes (Potter et al., 2016). En el ámbito de la ecología evolutiva, con el uso de elementos ultraconservados (900 UCE loci) se probaron diversas hipótesis evolutivas en los peces mariposa (Chaetodontidae) en los mares de la península Arábiga, demostrando la utilidad de los UCE para resolver filogenias a escalas de divergencia reciente (entre especies) y permitiendo un marco evolutivo filogenético (espacial y temporal), en el cual se rechazaron hipótesis previas de extinciones masivas durante el Pleistoceno en estos mares, persistencia de las especies durante cambios climáticos en el Pleistoceno y se determinó que las condiciones ambientales de las aguas costeras de la Península, fueron las determinantes en la evolución de especies endémicas altamente especializadas a esta región (DiBattista et al., 2018). Los anteriores son solo algunos ejemplos de los estudios cada vez más sofisticados que se realizan actualmente en distintas áreas del conocimiento con ayuda de un marco filogenómico, las aplicaciones abarcan distintas disciplinas científicas (Phillipe & Blanchet, 2007).

Posteriormente me enfocaré en la biogeografía con algunos ejemplos y en el potencial de la filogenómica en el estudio de especies crípticas así como en modelos evolutivos que intentan explicar los mecanismos de generación de diversidad críptica.

FUNDAMENTOS DE LA FILOGENÓMICA

Secuenciación de alta precisión o NGS

Los métodos de secuenciación de segunda generación (secuenciación de alta precisión o conocidos también como *next generation sequencing*; NGS), permitieron la obtención de datos genómicos a relativamente bajos costos (costo por megabase de secuencias crudas o por genomas); hasta el 2017 por ejemplo, secuenciar un genoma podría costar 1,000 dólares, con una tendencia decreciente en los costos (ver: <https://www.genome.gov/27541954/dna-sequencing-costs-data/>). Las plataformas de secuenciación de segunda generación, en los últimos 15 años, han revolucionado la manera de obtener secuencias de DNA, permitiendo el desarrollo de manera conjunta de la tecnología y de las aplicaciones de la información generada para estudios en ciencia básica y aplicada (Chan & Ragan, 2013;

López-de Heredia, 2016; Bleidorn, 2017; Slatko et al., 2018). Las tecnologías de secuenciación de segunda generación con mayor éxito fueron las plataformas de Illumina, mediante la técnica de "amplificación por puente", fragmentos cortos de ADN (~ 500 pb) con adaptadores ligados en cada extremo del fragmento, son utilizados como sustratos para repetidas reacciones de amplificación por síntesis sobre una superficie sólida (celda de flujo), que contiene secuencias de oligonucleótidos complementarios a cada adaptador ligado (López-de Heredia, 2016; Bleidorn, 2017; Slatko et al., 2018). Los nucleótidos son adicionados junto con una polimerasa y los fragmentos de DNA; ligados, cada nucleótido incorporado está bloqueado en su grupo 3' OH y lleva un fluoróforo removible que es identificado por láser, así se lleva a cabo la secuenciación durante varios ciclos (Bleidorn, 2017).

Por otra parte los métodos de tercera generación permiten secuenciar fragmentos largos (hasta 60 kb por lectura) de DNA y RNA. Pacific Bioscience (PacBio), es la empresa líder hasta el momento. De manera breve, a través del método de secuenciación de una sola molécula en tiempo real (SMRT por sus siglas en inglés), fragmentos largos de DNA ligados en los extremos con adaptadores en forma de hebra con una polimerasa en una de ellas, forman lo que se denominan SMRTbell. Estos fragmentos (SMRTbell) son inducidos a entrar en unos micropocillos (*zero-mode waveguide*, ZMW), contenidos en un chip denominado SMRTcell. En cada uno de los micropocillos, la actividad de la polimerasa es monitoreada sin interrupción mientras incorpora los dNTP marcados fluorescentemente, mediante el uso de imágenes en tiempo real, los nucleótidos incorporados son detectados mientras son sintetizados a lo largo de una sencilla molécula de DNA. La detección se lleva a cabo en los micropocillos (ZMW), mediante la detección de longitudes de onda de cada nucleótido marcado (Rhoads & Au, 2015; Bleidorn, 2017; Slatko et al., 2018).

Por otra parte, los métodos de cuarta generación utilizan nanotecnología para hacer pasar largas moléculas de DNA a través de nanoporos (poros de pocos nm de diámetro). Hasta la fecha solo existe un consorcio que desarrolla esta técnica de secuenciación, Oxford Nanopore Technologies (ONT), con tres secuenciadores: MinION, PromethION y SmigdeION. Previo a la secuenciación, adaptadores son ligados a ambas terminaciones de cada fragmento de DNA, estos adaptadores facilitan el acomodo de las hebras de DNA (por consiguiente, incrementan la probabilidad de captura de hebras de DNA en los nanoporos) y la carga de una enzima de actividad progresiva en el extremo 5' de cada cadena. La enzima es necesaria para garantizar el desplazamiento unidireccional de un solo nucleótido a lo

largo de la cadena de ADN en milisegundos, incrementando la precisión del detector en la secuenciación. En estas plataformas, la detección de nucleótidos es a través de la diferencia de cargas, según cada nucleótido que es identificado por un detector, los nucleótidos que pasan por cada poro disminuyen la corriente iónica, a través de esta diferencia de iones cada nucleótido es identificado. Una vez que la molécula de DNA sale por un nanoporo, este queda disponible para ser utilizado por otra molécula de DNA (Jain, et al., 2016; Bleidorn, 2017; Slatko et al., 2018). Una de las ventajas de los métodos de tercera y cuarta generación sobre los de segunda generación, es la identificación de estructura en las secuencias de DNA, por lo que también estas plataformas son pioneras en la generación de información sobre aspectos de epigenómica (estructura), identificación de haplotipos en tiempo real, secuenciación de fragmentos largos de DNA y de secuencias repetitivas (Rhoads & Au, 2015; Jain et al., 2016; Slatko et al., 2018).

La finalidad de utilizar cualquier método de secuenciación de segunda a cuarta generación es la obtención de datos genómicos, a través de alguna estrategia de colección de datos derivados de alguna plataforma de secuenciación. Aunque existen diversas aproximaciones, las estrategias más utilizadas en la filogenómica han sido principalmente dos: secuenciación asociada a sitios de restricción de ADN (familia RADseq) y más recientemente el enriquecimiento híbrido (conocido también como captura de secuencias) (Lemmon & Lemmon, 2013; Bleidorn, 2017).

MÉTODOS PARA LA OBTENCIÓN DE DATOS EN LA ERA DE LA FILOGENÓMICA

Familia RADseq

Esta técnica sencilla de reducción de genoma, ha sido bastante popular y se encuentra en una amplia variedad de estudios que abarcan distintas disciplinas. Una de las principales ventajas de esta técnica es que no requiere información genómica previa, por lo que es ideal para iniciar estudios en organismos-no modelo en prácticamente cualquier taxón. Además permite obtener gran profundidad de cobertura por locus (mejorando la confianza del llamado de genotipos), y la secuenciación de múltiples muestras en paralelo en una misma corrida de secuenciación. Aunque el término RADseq fue acuñado para describir un método en particular (digestión de DNA genómico por una enzima de restricción), actualmente se utiliza para referirse al grupo de técnicas que dependen de enzimas de restricción, para determinar el conjunto de loci a ser secuenciado (genotyping by sequencing o GBS, ezRAD, 2bRAD, ddRAD) (Lemmon & Lemmon, 2013; McCormack et al., 2013; Andrews, 2016; Bleidorn, 2017). Esta técnica, se basa en el

uso de endonucleasas (enzimas) de restricción, que reconocen un patrón específico de bases nucleotídicas en las secuencias de DNA, fragmentando la doble hélice en sitios específicos de cuatro a ocho nucleótidos; posteriormente hay un proceso de fragmentación mecánica y ligamiento de índices y barcodes, después ocurre la secuenciación masiva de lecturas simples o pareadas y finalmente el pipeline bioinformático (Andrews, 2016). Como cualquier otra técnica, el RADseq no está libre de errores; entre ellos destaca el *alele dropout*, que se refiere a las mutaciones en los sitios de restricción, resultando en una omisión en el corte del DNA genómico en esos sitios. Para minimizar este tipo de error, se recomienda utilizar enzimas de restricción cortas (4 bases) y utilizar solo una enzima (RADseq original o GBS), ya que por ejemplo, el ddRAD incrementa al doble la proporción de este tipo de error (Andrews, 2016). Otros errores son las duplicaciones por PCR y errores de secuenciación, éstos son errores estocásticos, que pueden causar que un alelo sea secuenciado más que otro alelo en un locus durante el proceso de clonación. Para minimizar este tipo de error de manera práctica, el uso de lecturas pareadas puede ayudar a reconocer los alelos duplicados por PCR. De manera similar, las diferencias de cobertura entre muestras y loci, pueden variar debido a la generación de la librería o procesos de secuenciación. Este error es comúnmente resuelto mediante la eliminación de muestras que presentan baja cobertura, o antes de iniciar un proyecto RADseq, podemos disminuir el número de muestras por línea de secuenciación. Alternativamente, también existen opciones de filtrado de loci con baja cobertura en los pipelines bioinformáticos (Andrews, 2016). Algunas creencias sobre las limitaciones de la implementación de este método en el campo de la filogenética/filogenómica se han rechazado con el tiempo; una de las más comunes era, que esta técnica no serviría para inferir filogenias a escalas profundas, debido a las mutaciones en los sitios de restricción que reduce el número de loci homólogos conforme incrementa la distancia filogenética (McCormack et al., 2013). Este supuesto fue cuestionado y finalmente rechazado por algunos estudios, destacan las simulaciones *in silico*, con genomas de *Drosophila* (con tiempos de divergencia entre 5 y 63 mda), los cuales recuperaron la filogenia “correcta” entre 12 especies, además de un gran porcentaje de loci homólogos y bajo porcentaje de parálogos (Cariou et al., 2013). También estudios empíricos han comprobado la utilidad de inferir filogenias con datos RADseq a diferentes escalas (superficiales y profundas), ejemplo de ello son: ranas en Etiopía (*Leptopelis*), con divergencia de 40 mda (Reyes-Velasco et al., 2018); las especies crípticas de *Viburnum* en Norte América con edad corona ~50 mda (Eaton et al., 2017; Spriggs et al., 2019) y plantas *Juglans* en China que divergieron de su especie hermana *Castanea mollissima*, hace 65 mda (Zhao et al., 2018).

Otro aspecto que ha sido criticado es la cantidad de datos faltantes (*missing data RADseq*), minuciosamente revisado por Eaton et al. (2017), quienes discuten los datos faltantes principalmente causados por mutación-disrupción y baja cobertura de secuenciación en un contexto filogenético. Los resultados demostraron que cada una de esas fuentes de datos faltantes, lleva a diferentes distribuciones de información filogenética, particularmente con respecto a la cantidad de información filogenética de cuartetos (*quartet informativeness*); es decir, el número de loci en un conjunto de datos multilocus que tienen suficiente muestreo en taxones para ser potencialmente informativos. Los análisis permitieron descubrir la redundancia jerárquica (la cantidad de información filogenética en una división sobre un árbol se ve afectada por su ubicación jerárquica en el árbol) y su influencia sobre las posibles topologías resueltas. Árboles más balanceados (topológicamente) incrementan la cantidad de información de los cuartetos (*quartet informativeness*) a través de nodos más profundos, en comparación con árboles asimétricos. Por otra parte, los patrones recuperados por secuenciación de baja cobertura, es que los datos faltantes disminuirán la información filogenética (*informativeness*) hacia las puntas en un árbol balanceado, o igualmente en todas las divisiones de un árbol no-balanceado, sin embargo, nunca incrementarán la pérdida de información en los bordes más profundos en el tiempo independientemente de la topología. Estas fuentes de datos faltantes, pueden ser resueltas incrementando la cobertura de secuenciación y adicionando más muestras (o taxones), para ayudar a balancear el árbol filogenético y permitir mayor redundancia jerárquica (Eaton et al., 2017)

Enriquecimiento híbrido

El objetivo de este método es la captura y enriquecimiento de secuencias en regiones blanco específicas del genoma (McCormack et al., 2013; López-de Heredia, 2016; Bleidorn, 2017). De manera breve, este método recae en el uso de sondas de captura (DNA o RNA) que son complementarias a regiones blanco del genoma que son hibridizadas en una librería de DNA. El DNA blanco, es enriquecido posteriormente a través de alguna plataforma de secuenciación de alta precisión, mientras que el DNA restante es eliminado (McCormack et al., 2013; Bleidorn, 2017). El enriquecimiento híbrido, fue diseñado para enriquecer loci ortólogos (de una sola copia), evitando la intersección con genes parálogos y con suficiente muestreo a través de los genomas (cientos a miles de loci ortólogos) para análisis filogenéticos/filgenómicos. Los métodos que implementan esta técnica incluyen enriquecimiento híbrido anclado (AHE), enriquecimiento de elementos ultraconservados (UCE) y enriquecimiento de exones (Faircloth et al., 2012;

Lemmon et al., 2012; McCormack et al., 2013; López-de Heredia, 2016; Bleidorn, 2017). Las sondas de captura son secuencias de oligonucleótidos (~60–120 bp), que cubren las regiones blanco del genoma, por lo que deben ser diseñadas y sintetizadas previamente. Aquí encontramos una diferencia fundamental con respecto a los métodos de restricción-digestión (técnicas RAD), ya que se requiere de cierto conocimiento sobre el genoma o transcriptoma del taxón (o taxones), previo al diseño de sondas (McCormack et al. 2013; Bleidorn, 2017). Una de las principales ventajas de estas técnicas, es la capacidad de recuperar loci ortólogos con gran profundidad filogenética, es decir desde 250 hasta los 500 mda (Lemmon et al. 2012; Jones & Good, 2016) lo que resulta en un gran potencial para estudios evolutivos a distintas escalas temporales. Otras ventajas que hacen atractiva la implementación de elementos ultraconservados (UCE), es que este tipo de marcadores se encuentran en amplio número a través del genoma, y muy importante, tienen poco solapamiento con genes parálogos (Lemmon et al. 2012; McCormack et al., 2013; Jones & Good, 2016; Bleidorn, 2017). Adicionalmente, los métodos de captura dirigida, generan datos con mayor calidad en comparación con otros métodos de representación reducida del genoma, incluyendo una menor varianza en la cobertura, mayor precisión en el llamado de SNP, alta reproducibilidad y *contigs* ensamblados de mayor longitud (Jones & Good, 2016).

Todos los estudios que implican la elaboración de una filogenia molecular, requieren la obtención de genes homólogos (copias de genes de la descendencia originadas a partir de una secuencia de DNA ancestral común) para elaborar árboles de genes (y si es el caso, también árboles de especies). Específicamente lo que se desea, es la obtención de genes ortólogos (copias de genes que evolucionaron de un gen ancestral mediante un evento de especiación), evitando la cantidad de genes parálogos (copias de genes originadas por un evento de duplicación) y xenólogos (cuando una copia de un gen homólogo fue transferida de otra especie mediante transferencia genética horizontal) (Escalona-Fermín, 2018).

Es durante el ensamble de *reads* provenientes de la secuenciación de alta precisión, que estamos llevando a cabo la inferencia de ortología, generalmente hay dos aproximaciones: ensamble con genoma de referencia o ensamble *de novo*. Cuando utilizamos un genoma de referencia, asumimos que estamos “mapeando” loci ortólogos con respecto al genoma de referencia, cuando utilizamos un ensamble *de novo*, la inferencia de ortología es mediante un porcentaje de similitud. Aunque por ejemplo Pyrad (y su predecesor ipyrad) también cuenta con mecanismos para ensamblar *de novo* de alta calidad, maximizando la probabilidad de ortología y minimizando la probabilidad de

recuperar loci parálogos. Además, este programa fue diseñado específicamente para estudios filogenéticos/filogenómicos con datos obtenidos de algún método de la familia RADseq (Eaton, 2014; Reyes-Velasco et al., 2018). Evidentemente el uso de métodos de enriquecimiento híbrido (UCE, AHE, captura de exones), incrementa la probabilidad de obtener exclusivamente loci ortólogos y minimizan la cantidad de *missing data* durante el pipeline bioinformático, pues desde el diseño de las sondas, ya se sabe de antemano qué regiones genómicas estamos buscando para la inferencia filogenética (regiones homólogas únicas) (Jones & Good, 2016).

Reconstrucción filogenética/filogenómica

Durante años se utilizaron los árboles de genes de un solo marcador como aproximaciones exactas de los árboles de especies, sin embargo, con el paso del tiempo nos percatamos de que los árboles de genes y los árboles de especies no necesariamente presentan la misma topología (o resultan equivalentes) y por lo tanto, no deben ser interpretados como resultado de los mismos procesos evolutivos (Mallo & Posada, 2016; Escalona-Fermín, 2018). A pesar de que la distinción entre árboles de especies y de genes se conocía desde hace varias décadas (Maddison, 1997) la dificultad para obtener múltiples marcadores moleculares, retrasó su reconocimiento empírico hasta hace pocos años (Mallo & Posada, 2016). Esta discordancia entre árboles de genes y de especies puede ser debida a errores sistemáticos (incorrecta especificación del modelo, cantidad finita de datos y proceso de muestreo); pero también debido a procesos evolutivos como *incomplete lineage sorting* (ILS), duplicación y pérdida de genes, transferencia horizontal de genes, especiación híbrida y flujo génico (Maddison, 1997; Escalona-Fermín, 2018, Mallo & Posada, 2016). Los árboles de especies representan la historia evolutiva de los organismos muestreados, los nodos representan eventos de especiación, las ramas reflejan la historia poblacional entre especiaciones, el ancho de las ramas representan los tamaños efectivos poblacionales (N_e), mientras que la longitud de ramas representa el tiempo en años o generaciones. Por otra parte, los árboles de genes representan la historia evolutiva de las copias de genes muestreados, los nodos indican eventos de coalescencia, los cuales corresponden al proceso de replicación del DNA y divergencia, la longitud de ramas representa el conjunto de sustituciones por sitio (Mallo & Posada, 2016).

Existen varios métodos para reconstruir árboles de especies, los más utilizados pueden ser clasificados de acuerdo a los datos de entrada: supermatriz (concatenación), superárbol y co-estimación o de evidencia total (Mallo & Posada, 2016; Escalona-Fermín, 2018).

El método de supermatriz o concatenación, se basa en la unión de todos los loci en una matriz de secuencias alineadas y estima un árbol (de un “supergen”), utilizando las mismas herramientas para reconstruir filogenias con un solo gen (parsimonia, máxima verosimilitud, inferencia bayesiana, métodos de distancia). El supuesto de este método es que todos los loci comparten una misma historia evolutiva o que las diferentes historias genéticas se anulan (Mallo & Posada, 2016; Escalona-Fermín, 2018).

El método de superárbol consiste en dos pasos: 1) los árboles de genes son estimados independientemente y 2) los árboles de genes son combinados en un solo árbol de especies. Aquí existen tres estrategias principalmente, de reducción de discordancia, proceso evolutivo simple y proceso evolutivo múltiple. El primero no utiliza ningún modelo evolutivo, únicamente intenta encontrar el árbol que minimiza la discordancia entre árboles de genes, aquí encontramos métodos consenso y concordancia tales como BUCKy (por factores de concordancia) y ASTRAL (minimiza la distancia de cuartetos entre árboles de genes y especies) (Liu et al., 2015; Mallo & Posada, 2016). Los métodos de proceso evolutivo simple se basan en la optimización de reconciliación de árboles de genes y árboles de especies, estos métodos calculan las coalescencias profundas, duplicaciones-pérdidas de genes o transferencias horizontales de genes necesarias para explicar la discordancia entre los árboles y a partir de ello genera el árbol de especies; algunos ejemplos de estos métodos son STEM, STEAC, STAR, MP-EST (Mallo & Posada, 2016). Por otra parte, los métodos de reconstrucción de árboles de especies basados en múltiples procesos evolutivos son escasos, un ejemplo es Guenomu, el cual considera ILS, duplicación y pérdida de genes, transferencia horizontal de genes y discordancia entre árboles de genes y de especies. Éste método se basa en modelos jerárquicos bayesianos, calculando la probabilidad posterior del árbol de especies dado los árboles de genes.

Otros métodos de reconstrucción de árboles de especies se basan en la evidencia total de las secuencias, están los que modelan ILS como SVDquartets. Éste método intenta estimar la mejor topología de cuartetos de taxones, basado en el *singular value decomposition* (SVD), de una matriz de frecuencias de caracteres; posteriormente los árboles de cuartetos válidos son ensamblados en un árbol de especies. En el caso de SVDquartets los datos de entrada son una matriz de SNP (Chifman & Kubatko, 2014; 2015; Mallo & Posada, 2016). Otro método que modela ILS y emplea marcadores bialélicos (SNP y AFLP) es SNAPP, implementado en BEAST2, el cual estima árboles de especies, tiempos de divergencia y tamaños poblacionales.

Este programa se basa en dos supuestos: cada marcador es un sencillo carácter bialélico y las genealogías de marcadores no ligados son condicionalmente independientes dado el árbol de especies (por ejemplo si los SNP están bien espaciados a lo largo del genoma). Si los supuestos son cumplidos, el algoritmo intentará obtener la densidad de probabilidades conjunta para los árboles de genes y de especies dado el conjunto de datos (Bryant et al., 2012).

Cada uno de los métodos de reconstrucción de árboles filogenéticos tiene sus limitaciones y debe ser seleccionado de acuerdo al tipo de datos a analizar, así como los supuestos de cada aproximación y al modelo de estudio. Aunque existen tendencias generales y recomendaciones que podemos seguir para la elaboración de árboles de especies, por ejemplo: los árboles de genes pueden ser robustos a pesar de cierta cantidad de datos faltantes (según simulaciones hasta el 95% en un marco bayesiano) (Wiens & Moen, 2008), siempre y cuando la señal filogenética de los datos sean suficientes para recuperar árboles confiables (Wiens & Moen, 2008; Mallo & Posada, 2016). Asimismo debemos comprender la naturaleza de datos faltantes y tener ciertas precauciones con los *missing data*, pero no desecharlos completamente ni creer que un porcentaje “alto” de *missing data* necesariamente tiene implicaciones negativas, esto ha sido observado en experimentos empíricos que utilizan métodos de reducción del genoma (Huang & Knowles, 2016; Eaton et al., 2017). En cuanto a los métodos de reconstrucción de árboles de genes y de especies, el que recupera con mayor precisión las topologías “verdaderas” (según simulaciones y datos empíricos) de los árboles de genes es *BEAST (Chou et al., 2015; Mallo & Posada, 2016); sin embargo, debido a la gran demanda computacional, se puede recurrir a otros métodos basados en coalescencia de reconstrucción de árboles de especies (SVDquartets, ASTRAL, NJst), los cuales presentan rendimientos similares bajo ciertas condiciones (Chou et al., 2015). Por otra parte el método de la supermatriz es el más robusto para reconstruir árboles de especies siempre y cuando los efectos del ILS o transferencia horizontal de genes sean bajos y/o si utilizamos secuencias cortas (Chou et al., 2015; Mallo & Posada, 2016). Finalmente, incrementar la cobertura de secuenciación en métodos de reducción del genoma y tratar de incrementar lo más posible el muestreo, ayuda a generar árboles de genes balanceados, lo que a su vez ayudará a incrementar la probabilidad de obtener el árbol de especies correcto (Eaton et al., 2017).

FILOGENÓMICA Y BIOGEOGRAFÍA

La biogeografía es la rama de la biología que se encarga del estudio de patrones y procesos (históricos y ecológicos) que determinan la distribución espacial de los organismos. Convencionalmente se divide en dos: biogeografía ecológica (distribución espacial a escala local) y biogeografía histórica (distribución de los seres vivos de acuerdo a su historia evolutiva). La biogeografía histórica se encuentra estrechamente vinculada a la biología evolutiva y a la sistemática, pues busca inferir los probables procesos que explican los patrones de distribución de la diversidad biológica, generalmente utilizando como base las relaciones de parentesco (filogenias) entre los organismos e información geográfica. Por lo tanto esta es una disciplina que unifica conocimiento de la sistemática, geografía, paleontología, geología (Sanmartín, 2012; Sánchez-Ramírez et al., 2017) y recientemente también incorpora aspectos de genómica y filogenómica (Grover et al., 2017; Barby et al., 2018; Grab et al., 2019). La incorporación de la filogenómica en la biogeografía, permite ahora realizar estudios más precisos en la evolución de cualquier rama del

árbol de la vida, incluso estudiar taxones de amplia distribución geográfica, patrones de distribución disyunta y otros más complejos, que antes quedaban limitados por la cantidad de información obtenida (pocos loci o un solo locus). Conforme se van integrando la filogenómica y la genómica en biogeografía, la precisión en la reconstrucción de filogenias, permite conocer a detalle procesos evolutivos involucrados en la diversificación de algunos taxones. Asimismo, permite generar hipótesis más precisas sobre los probables eventos geográficos y climáticos (a distintas escalas de tiempo) que han delineado los patrones de distribución de dichos taxones.

En esta parte me centraré en casos de estudio de reciente literatura que han examinado patrones biogeográficos particularmente en taxones de amplia distribución. Intentaré ejemplificar que los análisis biogeográficos, se benefician de la obtención de datos genómicos; en el sentido que las filogenias más robustas, generalmente conducirán a una evidencia más sólida, cuando distintas hipótesis son evaluadas en un mismo estudio (Sánchez-Ramírez et al., 2017)

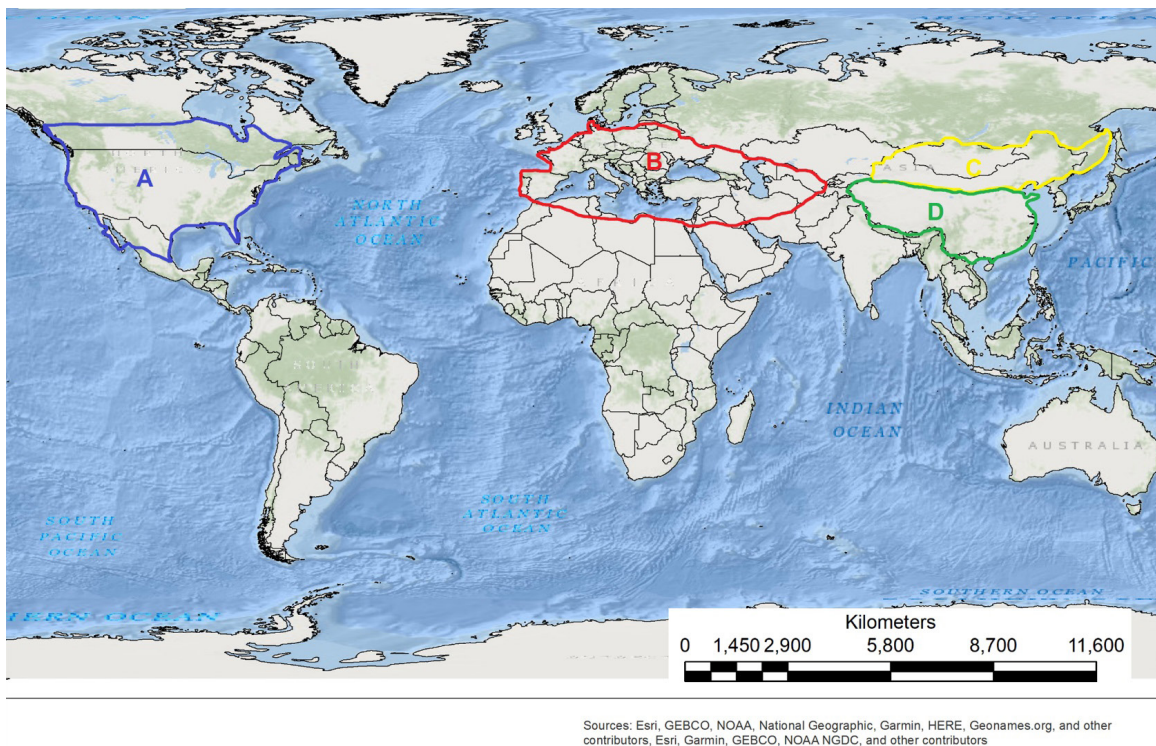


Figure 1. Four areas of endemism identified on the evolution of the *Vitis* subgenus. (A) North America, (B) Europe and Western Asia, (C) Asiatic Northeast, North of China and Himalayan Northeast, (D) Center and South of China. Adapted from Ma et al. (2018), for further information, consult the original source.

Figura 1. Cuatro áreas de endemismo identificadas en la evolución del subgénero *Vitis*. (A) Norteamérica, (B) Europa y Oeste de Asia, (C) Noreste Asiático, Norte de China y Noreste de los Himalayas, (D) Centro y Sur de China. Adaptado de Ma et al. (2018), para mayor información, ver la fuente original.

Subgénero *Vitis* (Ma et al., 2018)

A pesar de que la uva es uno de los cultivos de domesticación temprana y uno de los mayores cultivos de importancia económica, gran controversia aún existía hasta hace poco tiempo sobre aspectos básicos de su biología, como las relaciones filogenéticas entre las especies (sobre todo en las ramas profundas) y debido a su amplia distribución intercontinental (Holártica), también existían varias hipótesis sobre aspectos biogeográficos dentro del subgénero *Vitis*, que intentaban explicar el origen y evolución de su patrón de distribución actual (Fig. 1). Como mencionan claramente los autores del estudio, una filogenia bien resuelta es clave para comprender la historia evolutiva y los patrones de diversificación en especies de este subgénero (y en cualquier taxón). Mediante la técnica de resecuenciación de 40 genomas, demostraron las relaciones filogenéticas existentes entre las especies y los clados recuperados, además demostraron el probable centro de origen de las uvas en Norteamérica. Durante el Eoceno se originaron las divergencias de los clados de Norteamérica y Europa y durante este periodo también diversificaron los clados americanos. Esta hipótesis sugiere que existieron eventos de migración de Norteamérica a Europa, aunque el mecanismo no se sabe con certeza. Por otra parte cambios climáticos durante la transición Eoceno-Oligoceno, permitieron una rápida diversificación en el sureste de Asia, este supuesto fue confirmado por el análisis de algunas estructuras morfológicas en las plantas actuales de aquella región (Ma et al., 2018).

Género *Citrus* (Wu et al., 2018)

El género *Citrus* se distribuye ampliamente a través de la región del monzón, desde el Oeste de Pakistán hacia el Norte en la parte central de China, al Sur a través del archipiélago de las indias Orientales, Nueva Guinea, noreste de Australia, Nueva Caledonia, Melanesia y las islas Polinesias Occidentales. Aunque los autores del estudio sabían que el origen de varias especies de cítricos, se encontraba inmersa dentro de esa amplia área de distribución, el origen geográfico y el tiempo de diversificación de los cítricos eran dos aspectos prácticamente desconocidos, a pesar de la gran importancia económica y alimenticia que representan los cultivos de cítricos. A partir del análisis de 58 genomas completos, recuperaron la historia evolutiva de los cítricos y la diversidad que incluye este grupo de plantas comestibles. Una de las dificultades a la que se enfrentaron los investigadores fue a la taxonomía y a las propuestas filogenéticas controversiales en este grupo de plantas, principalmente debido a la dificultad de identificar especies progenitoras puras, ocasionado por los altos niveles de hibridación interespecífica y propagación de estos híbridos para cultivo. A través de la identificación de las variedades híbridas, pudieron determinar la monofilia de los cítricos con gran apoyo. El centro de origen de los cítricos fue ubicado en una región entre el noreste de India, norte de Myanmar y noroeste de Yunnan, con una rápida radiación en Asia durante el Mioceno tardío. Posteriormente durante el Plioceno ocurrieron eventos de migración del sur de Asia a través de la línea de Wallace hacia Australia originando la divergencia de las especies de cítricos en esa región (*E. glauca*, *Microcitrus australasica* y *Microcitrus australis*) (Wu et al. 2018).

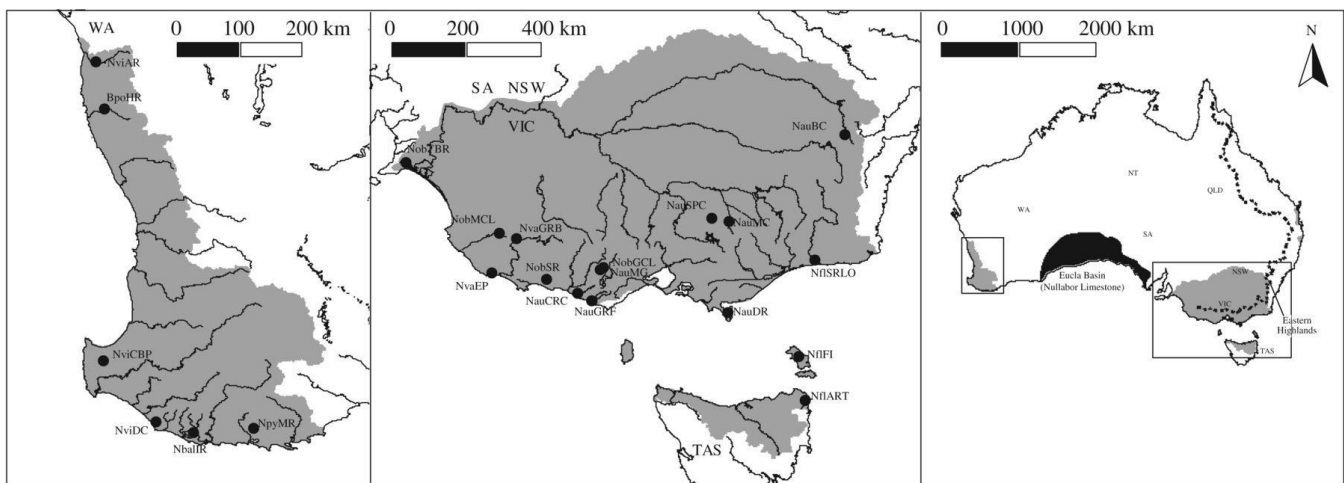


Figure 2. Current distribution of the perca pigmeas species studied by Buckley et al. (2018). Note the disjunctive distribution pattern on the right box. The Nullarbor plain (geographic barrier) is identified in black on the right box.

Figura 2. Distribución actual de las especies de perca pigmeas estudiadas por Buckley et al. (2018). Nótese el patrón de distribución disyunta en el recuadro derecho. La planicie de Nullarbor (barrera geográfica) es identificada en negro en el recuadro derecho.

Género *Nannoperca* (Buckley et al., 2018)

Las percas pigmeas australianas presentan un patrón de distribución bastante interesante (distribución disyunta continental Este-Oeste), este aspecto las define como un buen modelo para estudiar aspectos de diversidad, taxonomía, evolución y biogeografía mediante un enfoque filogenómico (Fig. 2). A diferencia de los ejemplos anteriores (que analizan genomas completos), en este estudio los autores implementaron la estrategia de reducción del genoma mediante la técnica ddRADseq. Los tiempos de divergencia obtenidos demostraron que las percas pigmeas son un grupo relativamente antiguo (20 mda), mientras que el modelo biogeográfico más probable ubicó a los ancestros en el Este de Australia, además de identificar un efecto fundador no muy común (a larga distancia) desde poblaciones ancestrales del Este, dicho evento permitió la colonización del Oeste de Australia y posterior diversificación. Resulta interesante que el análisis de reconstrucción de áreas

ancestrales, no apoya el modelo simple de divergencia por vicarianza, a partir del surgimiento de la barrera biogeográfica en el Centro-Sur de Australia (Planicie Nullarbor), en cambio apoya múltiples migraciones desde el Este hacia el Oeste, implicando eventos de diferenciación conforme surgía la principal barrera biogeográfica (Planicie Nullarbor) y principalmente debido a los cambios climáticos, los cuales, aislaron periódicamente a algunas poblaciones de perca pigmea. Con el paso del tiempo los linajes diferenciados quedaron aislados partidos por la mitad formando la actual distribución disyunta (Buckley et al., 2018).

Parásitos del género *Physconelloides* (Sweet et al., 2018)

Unos de los estudios más interesantes son aquellos que estudian la coevolución de algunos parásitos con sus hospederos, sobre todo cuando se trata de parásitos obligados. Un modelo interesante es el estudio de los piojos parásitos (Insecta: Phthiraptera: Ischnocera), de las palomas del nuevo mundo (Aves: Columbidae:



Figure 3. Geographic distribution of lice (*Physconelloides*) studied by Sweet et al. (2017), colors represent biogeographic regions where the taxa are distributed recovered over the phylogeny solved with 1058 nuclear genes. Adapted from Sweet et al. (2018), for further information, consult the original source.

Figura 3. Distribución geográfica de piojos (*Physconelloides*) estudiadas por Sweet et al. (2017), los colores representan las regiones biogeográficas donde se distribuyen los taxones recuperados en la filogenia resuelta con 1058 genes nucleares. Adaptado de Sweet et al. (2017), para mayor información revise la fuente original.

Claravinae), los cuales son ectoparásitos obligados permanentes que se alimentan de las plumas suaves de sus hospederos. Aunque existen diversos estudios de coespeciación entre algunas especies de piojos y las aves que parasitan, no es claro como algunas especies como las de *Physconelloides* se dispersan ampliamente y diversifican, ya que estos parásitos presentan escasas capacidades de dispersión. Los autores del estudio estaban particularmente interesados en las interacciones de los piojos con relación a la estructura filogenética del hospedero y a su distribución geográfica, específicamente ¿cómo están estructuradas las especies y poblaciones de piojos con relación a la evolución de su hospedero?

Las especies evaluadas presentan una distribución bastante amplia, la cual, abarca prácticamente todo el continente americano (Fig. 3). Mediante la obtención de genomas completos de los piojos por la técnica *shotgun*, los autores obtuvieron múltiples marcadores para abordar diferentes cuestiones evolutivas a distintas escalas espaciales y temporales. Los hallazgos sugieren que las poblaciones de piojos están inicialmente estructurados de acuerdo a los patrones biogeográficos de las propias especies de parásitos (*Physconelloides*), sin embargo, con el paso del tiempo eventualmente su distribución geográfica se acomodará paulatinamente de acuerdo a la especie hospedera, perdiendo la estructura biogeográfica con respecto a la filogenia propia de

los parásitos. Esta característica es distintiva de los ejemplos anteriores, donde las filogenias resueltas otorgan pistas de estructura biogeográfica, en cambio este estudio, demostró que existen interacciones complejas entre el parásito y el hospedero, por lo tanto, la estructura biogeográfica actual no es visualizada claramente en la filogenia del parásito, sino más bien en relación con la estructura poblacional del hospedero, en este caso, las palomas que parasitan (Sweet et al., 2018).

Estos pocos ejemplos demuestran la gran utilidad de reconstruir árboles de genes y especies, confiables y robustos estadísticamente, para posteriormente ubicarlos en algún programa que intente reconstruir la historia biogeográfica de los linajes. En algunos casos la filogenia por sí misma otorga información sobre la diversificación de los taxones, aunque en ocasiones, resultan más complejos, debido principalmente a las interacciones intra e interespecíficas de algunos modelos como los parásito-hospedero. Estos ejemplos, también subrayan la importancia de obtener miles de marcadores de DNA genómico, toda la evidencia disponible puede ser utilizada para responder preguntas evolutivas a diferentes escalas (profundas o recientes). Además, es indiscutible para todos los casos que se requiere una filogenia robusta para probar distintas hipótesis, por lo que la filogenómica resulta indispensable en estos estudios (Tabla 1).

Table 1. Summary of the genomic evidence and some findings of the studied cases.

Tabla 1. Resumen de la evidencia genómica y algunos hallazgos de los casos de estudio.

Modelo de estudio	Referencia	Recursos genómicos	Análisis biogeográficos	Patrones actuales de distribución y hallazgos
Uvas (<i>Vitis</i>)	Ma et al., 2018	41 genomas completos resecuenciados	Análisis binario bayesiano BMM en RASP	Distribución Holártica. Origen Norteamericano del ancestro de <i>Vitis</i> y subsecuente migración y diversificación en Europa y Asia.
Cítricos (<i>Citrus</i>)	Wu et al., 2018	60 genomas completos	Análisis de árboles filogenéticos, tiempos de divergencia y uso de fósiles.	Distribución en el Sureste de Asia y norte de Australia. Centro de origen a los pies del Himalaya, con periodos de diversificación en el Mioceno, con las últimas divergencias durante el Plioceno y el Pleistoceno en Australia y Japón respectivamente.
Percas pigmeas (<i>Nannoperca</i>)	Buckley et al., 2018	Representación reducida del genoma (ddRad) de 45 muestras individuales	Paquete de R BioGeoBEARS	Distribución disyunta en Australia (Este- Oeste). Múltiples migraciones desde el Este hacia el Oeste promovieron el establecimiento y posterior diversificación de especies de percas pigmeas. Posteriormente se estableció una barrera infranqueable originando la distribución disyunta actual.
Parásitos (<i>Physconelloides</i>)	Sweet et al., 2018	12 genomas completos	Método Madisson-Slatkin	Distribución en todo el continente americano. La diversificación en parásitos es multifactorial. La biogeografía es un pobre predictor de los posibles procesos de diversificación en este grupo de parásitos.

FILOGENÓMICA Y ESPECIES CRÍPTICAS

Una vez que hemos notado las aplicaciones y el potencial de la filogenómica en distintas ramas de la ciencia, podemos integrar esta poderosa herramienta en el estudio de las denominadas especies crípticas. La taxonomía es la ciencia (y la práctica) de la delimitación y clasificación de unidades biológicas, se encarga principalmente de identificar, nombrar, clasificar y describir organismos (Dantas-Torres, 2018; Struck & Cerca, 2019). El desafío que plantean las especies crípticas ha sido reconocido por más de 300 años, sin embargo, con la llegada de métodos de secuenciación de DNA relativamente a bajo costo y en tiempos cortos han dado una herramienta poderosa para detectar y diferenciar especies morfológicamente muy similares, pero que al menos en el aspecto genético resultan muy distintas entre sí (Bickford et al., 2007; Poulin & Pérez-Ponce de León, 2017; Dantas-Torres, 2018; Struck & Cerca, 2019). A partir de la década de los 90, las especies crípticas han sido detectadas a un ritmo cada vez mayor a través de todos los hábitats de la Tierra y en prácticamente todas las ramas del árbol de la vida incluyendo una amplia diversidad de organismos: hongos, algas, protistas, plantas, reptiles, anfibios, invertebrados, crustáceos y primates (Bickford et al., 2007; Pfenninger & Schwenk, 2007; Pérez Ponce-de León & Poulin; Poulin & Pérez-Ponce de León, 2017; Fišer et al., 2018; Struck & Cerca, 2019). El descubrimiento de especies crípticas en todas las ramas del árbol de la vida, sugiere una independencia del taxón y el bioma o hábitat y por lo tanto, podría tener profundas implicaciones en la teoría evolutiva, biogeografía, macroecología, conservación y manejo de la biota (Bickford et al. 2007, Pfenninger & Schwenk, 2007). El primer intento de unificar un marco conceptual para el estudio de especies crípticas fue el de Bickford et al. (2007), quienes definieron especie críptica como: “dos o más especies distintas que son erróneamente clasificadas (por lo tanto ocultas) bajo un solo taxón específico” (Struck & Cerca, 2019). Otra definición similar fue dada por Pfenninger y Schwenk (2007), quienes definen a las especies crípticas como “dos o más especies que fueron clasificadas y nombradas como una sola especie debido a su gran similitud morfológica”. Más recientemente Dantas-Torres (2018), definió a las especies crípticas como: “dos o más especies distintas que son morfológicamente similares o idénticas”. Este último concepto parece ser el más apropiado para designar linajes metapoblacionales con historias evolutivas independientes, pero que comparten gran similitud en varios caracteres morfológicos entre las especies comparadas. Existen algunos patrones generales reconocidos hasta la fecha para estudios que han abordado el tema de especies crípticas, el primero (y quizá bastante obvio), es que mientras más secuencias de DNA e individuos incorporan, estos

estudios tienden a recuperar una mayor diversidad críptica; segundo, aparentemente las especies de agua dulce tienen mayor probabilidad de ser crípticas en comparación con linajes terrestres o marinos y tercero, que los parásitos no presentan más especies crípticas en comparación con organismos de vida libre (Poulin & Pérez-Ponce de León, 2017). Consecuentemente con el advenimiento de la genómica y la filogenómica, en los próximos años podremos ver si estos patrones obtenidos con datos de secuenciación Sanger se mantienen o cambiarán en la era de la filogenómica.

Actualmente se está incorporando la filogenómica para descubrir, y en algunos casos describir formalmente especies crípticas en reptiles (Singhal et al., 2018), anfibios (Dufresnes et al., 2018), algas marinas (Boo & Hughey, 2019) y plantas con flor (Spriggs et al., 2019), por mencionar algunos ejemplos. Sin embargo, si deseamos saber cuáles son los procesos involucrados en los patrones reconocidos en la formación de especies crípticas (especiación críptica), podemos seguir el criterio recientemente propuesto por Struck et al. (2018). Estos autores establecen una metodología clara para designar especies crípticas y determinar los posibles procesos involucrados en la formación de estas especies. Primero, debemos establecer que las especies bajo estudio son especies verdaderas, dado el concepto de especie aplicado (este paso no es distinto de cualquier otro proceso de delimitación de especies, independientemente si las entidades son crípticas o no); y segundo, debemos demostrar que las especies son fenotípicamente más similares entre sí de lo esperado, dado el tiempo que ha pasado desde su último ancestro común (o el grado de divergencia genética como aproximación de tiempo). Por lo tanto, las especies crípticas deben ser llamadas así, solo si el grado de disparidad fenotípica es significativamente menor de lo esperado (Struck et al., 2018; Struck & Cerca, 2019). Ahora bien, ¿cómo delimitar especies con escasa o nula información fenotípica? A diferencia de especies que reflejan los procesos de especiación en la morfología, para descubrir y describir a las crípticas, se deben seguir otras estrategias en el proceso de delimitación de especies. Generalmente estas especies son abordadas con métodos de delimitación, a partir de análisis de información genética/genómica, y posteriormente validadas estadísticamente por algún programa.

Actualmente se siguen desarrollando y mejorando programas que ayudan a delimitar especies a partir de caracteres moleculares, la ventaja de estos métodos es que evita la necesidad de tener muchos criterios subjetivos en la delimitación de especies (en cualquiera, sea críptica o no). Por lo tanto, la delimitación de especies con datos moleculares ha sido implementada en algunos estudios como método único

(no recomendado), o como parte de un enfoque de taxonomía integrativa (recomendado) (Sukumaran & Knowles, 2017; Luo et al., 2018). Los primeros intentos fueron dirigidos a evaluar un simple locus, aquí destacan los programas GMYC y PTP ampliamente utilizados, sin embargo, en la actualidad estos dos programas también son incorporados en aproximaciones multilocus (incluso genómicos), para delimitar especies. El GMYC intenta diferenciar estadísticamente procesos de coalescencia intraespecífica, contra eventos de especiación a través de un árbol ultramétrico, mientras que el PTP intenta diferenciar el número de sustituciones nucleotídicas por comparación de secuencias, bajo el supuesto de que las secuencias que reflejan eventos de especiación mostrarán mayores sustituciones nucleotídicas en comparación con las secuencias de una sola especie, las cuales tendrán menor número de sustituciones (Leavitt et al. 2015; Luo et al., 2018). Estos dos métodos pueden combinarse incluso con otros programas más sofisticados, que integran estadísticos y modelos complejos, como el popular BP&P o SpeDeSTEM, y BFD*, estos tres métodos, incorporan el modelo multiespecies-coalescencia para delimitar y validar estadísticamente especies putativas con datos genómicos (multilocus los dos primeros y el tercero incorpora SNP). Hasta la fecha, el programa BP&P es el que ha demostrado mayor precisión y menores errores en cuanto a la delimitación de especies “verdaderas”, incluso en escenarios complejos de especiación (Luo et al. 2018). Evidentemente este programa no está libre de errores (es particularmente sensible a modelos que incluyen flujo génico y a errores en la definición de los parámetros θ y τ), actualmente también sabemos que este programa delimita estructura tanto poblacional (intraespecífica) como debida a especiación (interespecífica), por lo que existe incertidumbre sobre como diferenciar estas estructuras, ya que solo las segundas delimitan especies en sentido estricto (Sukumaran & Knowles, 2017). Estas debilidades del sofisticado BP&P pueden subsanarse incorporando otras líneas de evidencia, en el caso de especies crípticas serán diferentes a la morfología, como aspectos en ecología, distribución, etología y cualquier atributo que ayude a delimitar especies putativas (Sukumaran & Knowles, 2017; Luo et al., 2018).

Aunque también existen otras propuestas, que sugieren que el potencial de los recursos genómicos puede detallar a niveles muy precisos la diversidad biológica, a pesar de ello, la mayoría de los trabajos han fallado en demostrar este potencial. Por ejemplo, las distintas líneas de evidencia para delimitar especies pueden ser derivadas directamente de aspectos genómicos (al menos en teoría), de acuerdo con Stanton et al. (2019): 1) la evidencia morfológica puede ser identificada vía diferenciación en el desarrollo y genes estructurales, 2) evidencia biogeográfica puede ser derivada utilizando modelos complejos

y datos a escala genómica, 3) diferencias en el comportamiento pueden ser inferidas mediante la identificación de genes asociados con el comportamiento, selección de pareja, y otras características demográficas, 4) la evidencia ecológica puede ser derivada de las firmas genómicas de selección a ciertas características ambientales y 5) el aislamiento reproductivo puede ser observado en la compatibilidad/incompatibilidad de cromosomas sexuales, estructura cromosómica y transmisión epigenómica. Si efectivamente utilizamos todo el potencial de información derivada del estudio de los genomas, incluso tendríamos evidencias para abordar los dos conceptos de especie más utilizados sin inconvenientes en un mismo estudio, el filogenético y el biológico.

Una vez que tenemos certeza que las especies crípticas que hemos descubierto son un resultado inherente de la evolución, podemos tratar de inferir los procesos involucrados en la generación de esta diversidad. La escasa disparidad fenotípica (morfologías conservadas), ha recibido considerablemente menor atención en biología evolutiva, contrario a su fenómeno opuesto, la radiación adaptativa (Struck & Cerca, 2019). Mientras que las especies crípticas pueden ser sistemas ideales para informarnos sobre las causas de la reducida disparidad fenotípica, lo primero que se requiere es identificar el proceso evolutivo (o los procesos) que es/son causante(s) de la formación de especies crípticas, para lo cual, se han sugerido cuatro procesos (no excluyentes) que dan lugar a la *crípsis*: divergencia reciente, convergencia, paralelismo y estasis (Fišer et al., 2018; Struck et al., 2018; Struck & Cerca, 2019). La idea que subyace en reconocer estos procesos, es la utilización de las filogenias robustas (de preferencia con datos genómicos), como base para poder identificar estos procesos en marcos de tiempo evolutivos o divergencias genómicas, evidentemente a la luz de similitud morfológica.

En el caso de la divergencia reciente, las especies crípticas no son distintas de las no-crípticas, en el sentido de que debido a su origen reciente, poco o nulo cambio fenotípico puede observarse (particularmente en la morfología). En el caso de la similitud fenotípica debida a la convergencia o paralelismo, las especies crípticas evolucionan independientemente a fenotipos similares. En el paralelismo, las especies crípticas no necesariamente deben ser taxones hermanos, aquí lo importante es que las especies crípticas evolucionaron de ancestros similares fenotípicamente. En comparación con taxones hermanos no-crípticos y más recientes, la disparidad morfológica cambia menos a medida que las especies crípticas evolucionan de un morfotipo similar (en el ancestro común) a otro similar (en la descendencia). Por otra parte, en la convergencia, las especies crípticas no están

estrechamente relacionadas y su similitud morfológica (en la descendencia) resulta de una evolución independiente de ancestros morfológicamente distintos. Finalmente, la estasis, se refiere a grupos de taxones hermanos o grupos monofiléticos que retienen un alto grado de similitud morfológica sobre periodos extensos de tiempo. La propuesta de abordar a las especies crípticas bajo este enfoque, permite inferir probables procesos evolutivos involucrados en la similitud de caracteres fenotípicos en algunos taxones, más que “simplemente” reportar patrones de diversidad críptica en ciertos grupos taxonómicos (Struck & Cerca, 2019; Struck et al., 2018).

CONCLUSIONES FINALES Y PERSPECTIVAS

Es indudable que ya estamos inmersos en la era de la genómica, las tendencias en los próximos años serán la obtención de genomas completos de organismos no-modelo con alta precisión, conforme las tecnologías de secuenciación incrementan la precisión en la secuenciación de largos fragmentos de DNA. Mientras obtenemos genomas completos, las técnicas de reducción del genoma irán cada vez más dirigidas hacia métodos de enriquecimiento híbrido (captura de secuencias), aunque las técnicas RAD seguirán siendo una opción como primeras aproximaciones genómicas a los grupos de estudio. El desarrollo tecnológico en secuenciación, deberá ir de la mano con la actualización de herramientas bioinformáticas que permitan la integración de la información genómica en estudios ecológicos, evolutivos, taxonómicos, etc. Actualmente, aún estamos en el proceso de optimizar al máximo la información derivada de los datos genómicos, en este sentido, son evidentes las aportaciones de la filogenómica al conocimiento de la diversidad biológica, desde aspectos básicos (sistemática y en ocasiones en taxonomía), hasta la formación de un marco evolutivo, que permita la integración de diversas disciplinas biológicas. La finalidad es responder distintas preguntas que ayudarán a la conservación y mejor manejo de los recursos naturales, así como en el desarrollo de terapias y fármacos. Conforme las limitaciones tecnológicas en la obtención de datos genómicos parecen estar desapareciendo, los retos aparecen en otras áreas (pérdida de la biodiversidad a gran velocidad, reducción y destrucción de hábitats y ecosistemas, sobrepoblación, desigualdad).

Aunque las especies crípticas han sido reconocidas desde hace muchos años, ahora podemos estudiar los procesos involucrados en la escasa disparidad morfológica de algunas especies. El reconocimiento de la diversidad críptica, permitirá tener mayor conocimiento sobre los mecanismos de evolución y especiación, así como establecer mejores estrategias de conservación.

Debido a que las especies crípticas tienen una función en el ecosistema como cualquier otra, tienen relaciones intra e interespecíficas en las redes complejas que mantienen la evolución de la biodiversidad. Si no somos capaces de reconocer la diversidad críptica, estaremos sesgando el conocimiento y la conservación a únicamente especies morfológicamente diferenciadas, por lo que no tendremos el rompecabezas completo, desconociendo así la biodiversidad total y el papel que desempeña cada una de las especies que componen el árbol de la vida.

Agradecimientos.— Al Posgrado en Ciencias Biológicas de la UNAM, al CONACYT por la beca de doctorado otorgada (CVU 586418) y a los revisores anónimos por sus comentarios, los cuales, ayudaron a mejorar el manuscrito. Agradezco al PAPIIT IN216218 y a la Dra. Leticia Ochoa Ochoa por la invitación a someter el manuscrito a la RLH. Agradezco al comité editorial de la Royal Society Open Science, a Lianne Parkhouse y a Sean Buckley por permitir la reproducción de la Figura 2.

LITERATURA CITADA

- Andrews, K.R. 2016. Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics* 17:81-92.
- Barby, F.F., P. Ráb, S. Lavoue, T. Ezaz, L.A.C. Bertollo, A. Kilian, S.R. Maruyama, E.A. de Oliveira, R.F. Artoni, M.H. Santos, O.I. Jegede, T. Hatanaka, A. Tanomtong, T. Liher & M. de Bello Cioffi. 2018. From chromosomes to genome: insights into the evolutionary relationships and biogeography of old world knifefishes (Notopteridae; Osteoglossiformes). *Genes* 306:1-21.
- Bickford, D., D.J. Lohman, N.J. Sodhi, P.K.L. Ng, R. Meier, K. Winker, K.K. Ingram & I. Das. 2007. Cryptic species as a window on diversity and conservation. *Trends in Ecology and Evolution* 22(3):148-155.
- Bleidorn, C. 2017. *Phylogenomics an introduction*. Springer International Publishing, Cham, Switzerland.
- Boo, G.H. & J.R. Hughey. 2019. Phylogenomics and multigene phylogenies decipher two new cryptic marine algae from California, *Gelidium gabrielsonii* and *G. kathyanniae* (Gelidiales, Rhodophyta). *Journal of Phycology* 55:160-172.
- Bryant, D., R. Bouckaert, J. Felsenstein, N.A. Rosenberg & A. RoyChoudhury. 2012. Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. *Molecular Biology and Evolution* 29:1917-1932.

- Buckley, S.J., F.M.C.B. Domingos, C.R.M. Attard, C.J. Brauer, J.D. Sandoval-Castillo, R. Lodge, P.J. Unmack & L.B. Behegaray. 2018. Phylogenomic history of enigmatic pygmy perches: implications for biogeography, taxonomy and conservation. *Royal Society Open Science* 5:1-17.
- Cariou, M., L. Duret & S. Charlat. 2013. Is RAD-seq suitable for phylogenetic inference? An in silico assessment and optimization. *Ecology and Evolution* 3:846-852.
- Chan, C.X. & M.A. Ragan. 2013. Next-generation phylogenomics. *Biology Direct* 8:1-6.
- Chou, J., A. Gupta, S. Yaduvanshi, R. Davidson, M. Nute, S. Mirarab & T. Warnow. 2015. A comparative study of SVDquartets and other coalescent-based species tree estimation methods. *BMC Genomics*, 16:1-11.
- Chifman, J. & L. Kubatko. 2014. Quartet Inference from SNP Data Under the Coalescent Model. *Bioinformatics*. 30:3317-3324.
- Chifman J. & L. Kubatko. 2015. Identifiability of the unrooted species tree topology under the coalescent model with time-reversible substitution processes, site-specific rate variation, and invariable sites. *Journal of Theoretical Biology* 374:35-47.
- Da Fonseca, R.R., A. Albrechtsen, G. E. Themudo, J. Ramos-Madrigal, J. A. Sibbesen, L. Maretty, M. L. Zepeda-Mendoza, P. F. Campos, R. Heller & R. J. Pereira. 2016. Next-generation biology: Sequencing and data analysis approaches for non-model organisms. *Marine Genomics* 30:3-13.
- Dantas-Torres, F. 2018. Species concepts: what about ticks? *Trends Parasitology* 34(12):1017-1026.
- Delsuc, F., H. Brinkmann & H. Phillipe. 2005. Phylogenomics and the reconstruction of the tree of life. *Nature Reviews Genetics* 6:361-375.
- DiBattista, J.D., M.E. Alfaro, L. Sorenson, J.H. Choat, J.P. Hobbs, T. Sinclair-Taylor, L.A. Rocha, J. Chang, O.J. Luiz, P.F. Cowman, M. Friedman & M.L. Berumen. 2018. Ice ages and butterflyfishes: Phylogenomics elucidates the ecological and evolutionary history of reef fishes in an endemism hotspot. *Ecology and Evolution* 8:10989-11008.
- Dufresnes, C., G. Mazepa, N. Rodrigues, A. Brelsford, S.N. Litvinchuk, R. Sermier, G. Lavanchy, C. Betto-Colliard, O. Blaser, A. Borzée, E. Cavoto, G. Fabre, K. Ghali, C. Grossen, A. Horn, J. Leuenberger, B.C. Phillips, P.A. Saunders, R. Savary, T. Maddalena, M. Stöck, S. Dubey, D. Canestrelli & D.L. Jeffries. 2018. Genomic evidence for cryptic speciation in tree frogs from the Apennine Peninsula, with description of *Hyla perrini* sp. nov. *Frontiers in Ecology and Evolution* 6(144):1-18.
- Eaton, D.A.R. 2014. PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics* 30: 1844-1849.
- Eaton, D.A.R., E.L. Spriggs, B. Park & M.J. Donoghue. 2017. Misconceptions on missing data in RAD-seq phylogenetics with a deep-scale example from flowering plants. *Systematic Biology* 66: 399-412.
- Escalona-Fermín, M.M. 2018. Sensitivity of phylogenomic inference to the design of NGS target enrichment in non-model organisms. PhD Thesis. Universidad de Vigo, España.
- Faircloth, B.C., J.E. McCormack, N.G. Crawford, M.G. Harvey, R.T. Brumfield & T.C. Glenn. 2012. Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. *Systematic Biology* 61:717-726.
- Fišer, C., C.T. Robinson & F. Malaré. 2018. Cryptic species as a window into the paradigm shift of the species concept. *Molecular Ecology* 1:1-23.
- Geoghegan, J.L. & E.C. Holmes. 2018. The phylogenomics of evolving virus virulence. *Nature Reviews Genetics* 19:756-769.
- Grab, H., M.G. Branstetter, N. Amon, K.R. Urban-Mead, M.G. Park, J. Gibbs, E.J. Blitzer, K. Poveda, G. Loeb & B.N. Danfort. 2019. Agriculturally dominated landscapes reduce bee phylogenetic diversity and pollination services. *Science* 363:282-284.
- Grover, C.E., M.A. Arick II, J.L. Conover, A. Thrash, G. Hu, W.S. Sanders, C.Y. Hsu, R.Z. Naqvi, M. Farooq, X. Li, L. Gong, J. Mudge, T. Ramaraj, J. Udall, D.G. Peterson & J.F. Wendel. 2017. Comparative genomics of an unusual biogeographic disjunction in the cotton tribe (Gossypieae) yields insights into genome downsizing. *Genome Biology and Evolution* 9:3328-3344.
- Huang, H. & L.L. Knowles. 2016. Unforeseen consequences of excluding missing data from Next-Generation Sequences: simulation study of RAD sequences. *Systematic Biology* 65:357-365.
- Jain, M., H.E. Olsen, B. Paten & M. Akeson. 2016. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biology* 17:1-11.

- Jones, M.R. & J.M. Good. 2016. Targeted capture in evolutionary and ecological genomics. *Molecular Ecology* 25:185-202.
- Leavitt, S.D., C.S. Moreau & H.T. Lumbsch. 2015. The dynamic discipline of species delimitation: progress toward effectively recognizing species boundaries in natural populations. Pp. 11-44. En Kumar, D., P. K. Divakar, V. Shukla, & R. Bajpai (Eds.). *Recent advances in lichenology modern methods and approaches in Lichen systematics and culture techniques*. Springer New Delhi.
- Lemmon, A.R., S.A. Emme & E.M. Lemmon. 2012. Anchored hybrid enrichment for massively high-throughput phylogenomics. *Systematic Biology* 61:727-744.
- Lemmon, E.M. & A.R. Lemmon. 2013. High-Throughput genomic data in systematics and phylogenetics. *Annual Review of Ecology, Evolution, and Systematics* 44:99-121.
- Liu, L., S. Wu & L. Yu. 2015. Coalescent methods for estimating species trees from phylogenomic data. *Journal of Systematics and Evolution* 53:380-390.
- López-de Heredia, U. 2016. Las técnicas de secuenciación masiva en el estudio de la diversidad biológica. *Munibe Ciencias. Naturales* 64:1-25.
- Luo, A., C. Ling, S.Y.W. Ho & C. Zhu. 2018. Comparison of methods for molecular species delimitation across a range of speciation scenarios. *Systematic Biology* 67(5):830-846.
- Ma, Z.Y., J. Wen, S. Ickert-Bond, Z.L. Nie, L.Q. Chen & X.Q. Liu. 2018. Phylogenomics, biogeography, and adaptive radiation of grapes. *Molecular Phylogenetics and Evolution* 129:258-267.
- Maddison, W.P. 1997. Gene trees in species trees. *Systematic Biology* 46(3):523-536.
- Mallo, D. & D. Posada. 2016. Multilocus inference of species trees and DNA barcoding. Multilocus inference of species trees and DNA barcoding. *Philosophical Transactions of the Royal Society B* 371:1-11.
- McCormack, J.E., S.M. Hird, A.J. Zellmer, B.C. Carstens & R.T. Brumfield. 2013. Applications of next-generation sequencing to phylogeography and phylogenetics. *Molecular Phylogenetics and Evolution* 66:526-538.
- Pérez-Ponce de León, G. & R. Poulin. 2016. Taxonomic distribution of cryptic diversity among metazoans: not so homogeneous after all. *Biology Letters* 12:1-5.
- Pfenninger, M. & K. Schwenk. 2007. Cryptic animal species are homogeneously distributed among taxa and biogeographical regions. *BMC Evolutionary Biology* 7(121):1-6.
- Philippe, H., F. Delsuc, H. Brinkmann & N. Lartillot. 2005. Phylogenomics. *Annual Review of Ecology, Evolution, and Systematics* 36:541-562.
- Phillipe, H. & M. Blanchette. 2007. Overview of the first phylogenomics conference. *BMC Evolutionary Biology* 7:1-4.
- Posada, D. 2016. Phylogenomics for systematic biology. *Systematic Biology* 65(3):353-356.
- Potter, S., J.G. Bragg, B.M. Peter, K. Bi & C. Moritz. 2016. Phylogenomics at the tips: inferring lineages and their demographic history in a tropical lizard, *Carlia amax*. *Molecular Ecology* 25:1367-1380.
- Poulin, R. & G. Pérez-Ponce de León. 2017. Global analysis reveals that cryptic diversity is linked with habitat but not mode of life. *Journal of Evolutionary Biology* 30:641-649.
- Pyron, A. 2015. Post-molecular systematics and the future of phylogenetics. *Trends in Ecology and Evolution* 30:384-389.
- Reyes-Velasco, J., J.D. Manthey, X. Freilich & S. Boissinot. 2018. Diversification of African tree frogs (genus *Leptopelis*) in the highlands of Ethiopia. *Molecular Ecology* 27:2256-2270.
- Rhoads, A. & K.F. Au. 2015. PacBio sequencing and its applications. *Genomics Proteomics Bioinformatics* 13:278-289.
- Sanmartín, I. 2012. Biogeografía. Pp. 457-474. En P. Vargas & R. Zardoya (Eds.). *El Árbol de la Vida: sistemática y evolución de los seres vivos*. Ediciones Reverté, España.
- Sánchez-Ramírez, S., A.W. Wilson & M. Ryberg. 2017. Overview of phylogenetic approaches to mycorrhizal biogeography, diversity and evolution. Pp 1-37. En L. Tedersoo (Ed.), *Biogeography of Mycorrhizal Symbiosis*. Springer Nature, Switzerland.
- Singhal, S., C.J. Hoskin, P. Couper, S. Potter & C. Moritz. 2018. A framework for resolving cryptic species: a case study from the lizards of the Australian wet tropics. *Systematic Biology* 67(6):1061-1075.
- Slatko, B. E., A. F. Gardner & F. M. Ausubel. 2018. Overview of next-generation sequencing technologies. *Current Protocols in Molecular Biology* 122:1-11.

- Spriggs, E.L., D.A.R. Eaton., P.W. Sweeney, C. Schlutius, E.J. Edwards & M.J. Donoghue. 2019. Restriction-Site-Associated DNA sequencing reveals a cryptic *Viburnum* species on the North American coastal plain. *Systematic Biology* 68:187-203.
- Stanton, D.W.G., P. Frandsen, R.K. Waples, R. Heller, I.M. Russo, P.A. Orozco-terWengel, C.E.T. Pedersen, H.R. Siegismund & M.W. Bruford. 2019. More grist for the mill? Species delimitation in the genomic era and its implications for conservation. *Conservations Genetics* 20:101-113.
- Struck, T.H. & J. Cerca. 2019. Cryptic species and their evolutionary significance. *eLS* 1:1-9.
- Struck, T.H., J.L. Feder, M. Bendiksby, S. Birkeland, J. Cerca, V.I. Gusarov, S. Kistenich, K.H. Larsson, L.H. Liow, M.D. Nowak, B. Stedje, L. Bachmann & D. Dimitrov. 2018. Finding evolutionary processes hidden in cryptic species. *Trends in Ecology and Evolution* 33(3):153-163.
- Sukumaran, J. & L. Knowles. 2017. Multispecies coalescent delimits structure, not species. *Proceedings of the National Academy of Science of the United States of America* 114(7):1607-1612.
- Sweet, A.D., B.M. Boyd, J.M. Allen, S.M. Villa, M.P. Valim, J.L. Rivera-Parra, R.E. Wilson & K.P. Johnson. 2018. Integrating phylogenomic and population genomic patterns in avian lice provides a more complete picture of parasite evolution. *Evolution*. 72:95-112.
- Wiens, J.J. & D.S. Moen. 2008. Missing data and the accuracy of Bayesian phylogenetics. *Journal of Systematics and Evolution* 46:307-314.
- Wu, G.A., J. Terol, V. Ibanez, A. López-García, E. Pérez-Roman, C. Borredá, C. Domingo, F.R. Tadeo, J. Carbonell-Caballero, R. Alonso, F. Curk, D. Du, P. Ollittraut, M.L. Roose, J. Dopazo, F.G. Gmitter Jr, D. Rokhsar & M. Talon. 2018. Genomics of the origin and evolution of Citrus. *Nature*. 554, 311-316.
- Zhao, P., H.J. Zhou, D. Potter, Y.H. Hu, X.J. Feng, M. Dang, L. Feng, S. Zulfigar, W.Z. Liu, G.F. Zhao & K. Woste. 2018. Population genetics, phylogenomics and hybrid speciation of Juglans in China determined from whole chloroplast genomes, transcriptomes, and genotyping-by-sequencing (GBS). *Molecular Phylogenetics and Evolution* 126:250-265.

