



**UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO**  
**POSGRADO EN CIENCIAS BIOLÓGICAS**  
INSTITUTO DE BIOLOGÍA  
**SISTEMÁTICA**

**SISTEMÁTICA Y EVOLUCIÓN DE LAS AVISPAS PARASITOIDES DEL GÉNERO**  
***STENOCORSE* (HYMENOPTERA: BRACONIDAE: DORYCTINAE)**  
**EMPLEANDO DATOS DE 3RAD Y MITOCONDRIALES.**

**TESIS**

**POR ARTÍCULO CIENTÍFICO**

**GENETIC STRUCTURE, SPECIES LIMITS AND EVOLUTION OF THE**  
**PARASITOID WASP GENUS *STENOCORSE* (BRACONIDAE: DORYCTINAE)**  
**BASED ON NUCLEAR 3RAD AND MITOCHONDRIAL DATA**

QUE PARA OPTAR POR EL GRADO DE:  
**MAESTRA EN CIENCIAS BIOLÓGICAS**

PRESENTA:

**NATALIA DELGADO MACHUCA**

**TUTOR PRINCIPAL: DR. ALEJANDRO ZALDÍVAR RIVERÓN**  
INSTITUTO DE BIOLOGÍA, UNAM  
**COMITÉ TUTOR: DRA. CLAUDIA PATRICIA ORNELAS GARCÍA**  
INSTITUTO DE BIOLOGÍA, UNAM  
**DR. ALEJANDRO VALDEZ MONDRAGÓN**  
INSTITUTO DE BIOLOGÍA (SEDE TLAXCALA), UNAM

**MÉXICO, CD. MX.**

**JUNIO 2019**



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.

OFICIO CPCB/660/2019

Asunto: Oficio de Jurado para Examen de Grado.

**M. en C. Ivonne Ramírez Wence**  
**Directora General de Administración Escolar, UNAM**  
**Presente**

Me permito informar a usted que en la reunión del Subcomité por Campo de Conocimiento de Sistemática del Posgrado en Ciencias Biológicas, celebrada el día 11 de marzo de 2019, se aprobó el siguiente jurado para el examen de grado de **MAESTRA EN CIENCIAS BIOLÓGICAS** de la alumna **DELGADO MACHUCA NATALIA** con número de cuenta **308532740** con la tesis titulada **"Sistemática y evolución de las avispas parasitoides del género *Stenocorse* (Hymenoptera: Braconidae: Doryctinae) empleando datos de 3RAD y mitocondriales"**, realizada bajo la dirección del **DR. ALEJANDRO ZALDÍVAR RIVERÓN**:

Presidente: DR. ALEJANDRO CÓRDOBA AGUILAR  
Vocal: DR. LUIS ENRIQUE EGUIARTE FRUNS  
Secretario: DR. ALEJANDRO VALDÉZ MONDRAGÓN  
Suplente: DR. JAIRO ANDRÉS ARROYAVE GUTIÉRREZ  
Suplente: DRA. CLAUDIA PATRICIA ORNELAS GARCÍA

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

**ATENTAMENTE**  
**"POR MI RAZA HABLARA EL ESPIRITU"**  
Cd. Universitaria, Cd. Mx., a 4 de junio de 2019.



**DR. ADOLFO GERARDO NAVARRO SIGÜENZA**  
**COORDINADOR DEL PROGRAMA**



c.c.p. Expediente del (la) interesado (a).

### **Agradecimientos Institucionales**

Extiendo un cordial agradecimiento al Posgrado en Ciencias Biológicas de la Universidad Nacional Autónoma de México (UNAM) por el apoyo a lo largo del desarrollo de este estudio.

Al Consejo Nacional de Ciencia y Tecnología (CONACyT) por la beca otorgada durante el desarrollo mi maestría.

Al Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica (PAPIIT) por el apoyo otorgado a este proyecto de investigación. No. de Proyecto: IN207016

Al Programa de Apoyo a los Estudios de Posgrado (PAEP) de la UNAM por el apoyo económico otorgado durante mi estancia de investigación.

Agradezco ampliamente al Dr. Alejandro Zaldívar Riverón por coordinar y supervisar este trabajo, así como a la Dra. Claudia Patricia Ornelas García y al Dr. Alejandro Valdez Mondragón por el apoyo y las revisiones que hicieron posible este estudio.

## **Agradecimientos Personales**

Agradezco personalmente a todos mis compañeros del posgrado que están o han estado presentes a lo largo del desarrollo de este trabajo: Jovana, Ernesto, Rubén, Sofía, Jorge, Karen, Vladimir, David y particularmente a Rubí porque sin ella esto no habría sido posible. A mi asesor Alejandro Zaldívar por su guía y apoyo.

A Paul Hanson y Adina Chaín Guadarrama por su apoyo y guía durante el muestreo en Costa Rica. A Valentina Battistino por su apoyo durante el muestreo en Argentina. A Natalia Bayona Vásquez por su apoyo y enseñanza durante la estancia de investigación en Estados Unidos.

A todos los profesores que impartieron cursos en el programa de posgrado, gracias por compartir su conocimiento.

Agradezco también a mi familia y en especial a mis papás Sandra y Rodolfo por apoyarme incondicionalmente y siempre creer en mi.

Agradezco a Javier Gutiérrez Nova por toda la inspiración y el apoyo que hizo posible este trabajo.

A todas mis amigas y amigos biólogos que siempre me motivan a hacer las cosas mejor.

*Le dedico esta tesis a los insectos del mundo.*

## Índice

<b>RESUMEN</b>	<b>1</b>
<b>ABSTRACT</b>	<b>2</b>
<b>INTRODUCCIÓN</b>	<b>3</b>
<i>GRUPO DE ESTUDIO: HYMENOPTERA Y AVISPAS PARASITOIDES</i>	8
<i>ICHNEUMOIDEA: BRACONIDAE: DORYCTINAE</i>	10
<i>EL GÉNERO STENOCORSE</i>	12
<i>INTERACCIONES PARASITOIDE HOSPEDERO Y PROCESOS BIOGEOGRÁFICOS</i>	13
<b>JUSTIFICACIÓN</b>	<b>15</b>
<b>OBJETIVOS</b>	<b>16</b>
<i>GENERAL</i>	<b>16</b>
<i>PARTICULARES</i>	<b>16</b>
<b>MANUSCRITO DEL ARTÍCULO</b>	<b>17</b>
<b>DISCUSIÓN</b>	<b>60</b>
TIPOS DE DATOS Y SU DESEMPEÑO	60
DELIMITACIÓN DE ESPECIES	64
ESTIMACIÓN DE TIEMPOS DE DIVERGENCIA Y RELACIONES FILOGENÉTICAS CON DATOS MITOCONDRIALES	66
ESTRUCTURA Y RELACIONES FILOGENÉTICAS CON DATOS GENÓMICOS	68
BIOGEOGRAFÍA E INTERACCIONES	71
<b>CONCLUSIONES</b>	<b>73</b>
<b>REFERENCIAS</b>	<b>75</b>

## Lista De Figuras

- Figura 1.** A) Esquema general de la clasificación del orden Hymenoptera. B) Vista dorsal de las principales divisiones morfológicas de una avispa del suborden Apocrita, en rojo se muestra el propodeo y el primer segmento abdominal (Fuente: Goulet & Huber, 1993).....9
- Figura 2.** Braconidae: Doryctinae. Dibujo que muestra las principales características del grupo, así como la longitud que puede alcanzar el ovipositor de la hembra (Fuente: Goulet & Huber, 1993).....11
- Figura 3.** Ejemplar macho de *Stenocorse bruchivora* colectado en Cuajiniquil, Costa Rica.....13
- Figura 4.** Árbol ultramétrico obtenido con el programa BEAST. Con barras negras se muestran las especies delimitadas con base en el marcador mitocondrial.....67
- Figura 5.** Filograma Bayesiano obtenido con base en la matriz M1 usando el programa Exabayes. Los círculos negros corresponden a probabilidades posteriores < 95%. En las cajas de color se indican las especies delimitadas con datos de 3RAD y mitocondriales.....70



## RESUMEN

*Stenocorse* (Braconidae: Doryctinae) es un género de avispas parasitoides ampliamente distribuido desde el sur de Estados Unidos hasta el centro de Argentina. *Stenocorse* ataca a más de 60 especies de escarabajos de la subfamilia Bruchinae que se alimentan de semillas de diferentes legumbres. El género está compuesto por una sola especie descrita, *S. bruchivora* (Crawford, 1910), aunque se sospecha que está compuesto por varias especies similares morfológicamente. En este estudio investigamos la estructura genética, límites de especies y las relaciones evolutivas de las poblaciones de *Stenocorse* muestreadas a lo largo de su área de distribución, con especial énfasis en el territorio mexicano. Generamos secuencias de ADN del locus mitocondrial del código de barras, así como secuencias genómicas nucleares obtenidas mediante la técnica de 3RAD. Las reconstrucciones filogenéticas derivadas de los análisis Bayesianos de cada genoma fueron altamente congruentes, presentando estructura geográfica. Se detectaron cuatro grupos genéticos con los datos nucleares, lo cual es congruente con la distribución geográfica de las muestras y la delimitación de especies. Delimitamos seis linajes evolutivos con los dos conjuntos de datos, así como tres especies adicionales delimitadas únicamente con el locus del código de barras. *Stenocorse* pudo originarse del Mioceno temprano al Mioceno medio, y los eventos de diversificación probablemente ocurrieron entre el Plioceno temprano y el Pleistoceno tardío. Nuestro análisis de delimitación de especies, reconstrucciones filogenéticas y datos de recolección sugieren que las especies de *Stenocorse* poseen hábitos alimenticios generalistas. Este resultado es contrario a lo esperado debido a la especialización huésped-planta que presentan las especies de brúquinos y el efecto en cascada que puede tener sobre la diversificación de sus parasitoides.

## **ABSTRACT**

*Stenocorse* (Braconidae: Doryctinae) is an ectoparasitoid wasp genus widely distributed from southern USA to central Argentina, which is known to attack over 60 bruchine beetle species that feed on a variety of legumes. This genus currently comprises a single described species, *S. bruchivora* (Crawford, 1910), though it is probably composed of additional, morphologically similar species. In this study we investigated genetic structure, species limits and evolutionary relationships of sampled populations of *Stenocorse* obtained along its known geographic distribution, with an emphasis on the Mexican territory. For this, we generated DNA sequences from both the mitochondrial (mt) barcoding locus and nuclear genome data obtained with the 3RAD technique. The Bayesian phylogenetic reconstructions based on the two data sets were mostly congruent, both showing clear geographic structure. Four clusters were recovered with nuclear data, which were also congruent with the geographic distribution of the samples and the species delimitation analyses. We delimited six evolutionary lineages with the two data sources, and three additional species were also delimited with the COI locus alone. *Stenocorse* probably originated during the mid to early Miocene, and the diversification events that led to its current species distribution likely occurred between the early Pliocene and late Pleistocene. Our species delineation analyses, phylogenetic reconstructions and rearing records suggest that the species of *Stenocorse* have generalist feeding habits, contrary to what was expected due to host plant specialization of the bruchine species and the cascading effect that this could have in the species diversification of their parasitoids.

## INTRODUCCIÓN

La historia de los seres vivos que habitan la tierra ha sido forjada por diversos procesos evolutivos, los cuales, a través de millones de años de cambios heredables, han dado lugar a la maravillosa diversidad de morfologías, comportamientos, estrategias de vida, fisiologías y por supuesto, a la información genética que constituye la biodiversidad de nuestro planeta. El estudio de la biodiversidad, su clasificación y sistematización ha permitido comprender que existen similitudes entre los organismos que se deben a una historia evolutiva compartida, es decir, debido a que cualquier par de organismos de nuestro planeta posee un ancestro en común. Las características que permiten establecer esta conexión son conocidas como “homologías táxicas” (Delsuc *et al.*, 2005), y la reconstrucción histórica con base en estas características se conoce como filogenética, del griego *phyle*, tribu + *genesis*, nacimiento (Delsuc *et al.*, 2005; Gregory, 2008).

Tradicionalmente la reconstrucción filogenética está basada en la identificación de homologías táxicas usando caracteres morfológicos (p. ej., Belokobylskij *et al.*, 2004; Liljeblad y Ronquist, 1998). El descubrimiento del ADN y de sus mecanismos de replicación y traducción llevaron a la idea que las características de un organismo, o su fenotipo, son producto del ADN (genotipo). El ADN, presenta variaciones particulares a cada organismo entre las cuales es posible identificar homologías y es posible mapear la historia de su herencia. A estos caracteres se les denominó caracteres moleculares o marcadores puntuales (Avice, 2009; Peterson, *et al.*, 2012).

Los caracteres moleculares han permitido resolver problemas filogenéticos donde los caracteres morfológicos presentaban ambigüedades (Hebert, *et al.*, 2004), y aportan nueva información útil para la genética de poblaciones, filogeografía, la delimitación de especies, y en general revolucionaron la forma de hacer sistemática y taxonomía (Yang y Rannala, 2010).

La exploración y aplicación de los caracteres moleculares en todo tipo de organismos también ha permitido apreciar sus limitaciones. Hoy en día se sabe que el fenotipo de los organismos no es el producto directo ni completo de la expresión de su genotipo (West-Eberhard, 2005), y que la complejidad genética puede oscurecer, en lugar de aclarar, algunas preguntas relacionadas con la historia evolutiva de las especies (Maddison, 1997). Así, actualmente se considera que los marcadores moleculares son una herramienta poderosa pero no única ni incuestionable.

Ante la complejidad y vastedad de la biodiversidad de la Tierra, explorar la realidad de las especies consiste en recopilar la mayor cantidad de evidencia posible para respaldar las hipótesis y usar, en medida de lo posible, todo tipo de caracteres, incluso distintos tipos de caracteres moleculares. Los marcadores moleculares más utilizados actualmente corresponden a genes que se

encuentran conservados a lo largo de diferentes linajes, como por ejemplo los genes de la mitocondria, algunos genes nucleares y los microsatélites (Davey y Blaxter, 2010).

La forma en la que se hereda la información genética, y su cambio de generación en generación está influenciada por diversos factores, tales como el tamaño de las poblaciones, las presiones de selección y el tiempo (John C. Avise, 2009). El resultado de este proceso de herencia es que en algunos casos el ancestro común entre dos genes se extiende más allá en el tiempo que el evento de especiación que observamos en el grupo, lo cual se conoce como “persistencia de polimorfismos ancestrales” o “sorteo incompleto de linajes” y es considerada como la principal causa de la incongruencia entre árboles de genes y árboles de especies (Maddison, 1997). La principal forma de enfrentar el problema que representa esta incongruencia, consiste en utilizar la mayor cantidad posible de genes o loci (análisis multilocus) para reconstruir árboles de especies, modelando el proceso de coalescencia que genera la discordancia entre árboles de genes y árboles de especies debido principalmente a la presencia de polimorfismos ancestrales (Leaché y Rannala, 2011; Yang y Rannala, 2010).

Los análisis multilocus han impactado tanto los estudios de filogeografía como los de genética de poblaciones y de filogenética. En un punto medio entre las inferencias derivadas de estas disciplinas es posible delimitar especies como linajes con historias evolutivas independientes (Carstens *et al.*, 2013; De Queiroz, 2007). La categoría de especie conforma una de las unidades fundamentales de los niveles de organización biológica, y por lo tanto resulta primordial el desarrollo de conceptos y métodos de delimitación de especies que sean robustos y reproducibles (De Queiroz, 2007; Fujita, *et al.*, 2012).

La definición del concepto de especie es uno de los principales problemas a los que se enfrenta la práctica de delimitación. Dependiendo de la aproximación del estudio, es posible adoptar la definición de un concepto de entre más de veinte, los cuales incluso pueden llegar a ser parcialmente incompatibles y resultar en distintos límites y cantidad de especies (Mayden, 1997; De Queiroz, 2007). Por ejemplo, para los sistématas, el concepto de especie más empleado es el filogenético (Henning, 1966), el cual indica que las especies deben estar compuestas de grupos exclusivamente monofiléticos. Para los ecólogos, por otra parte, las especies están conformadas por grupos que ocupan nichos ecológicos diferentes (Van Valen, 1976), mientras que el concepto biológico de especie indica que los organismos deben estar aislados reproductivamente (Mayr, 1942 y 1963).

Visto de esta forma, todas las definiciones de los conceptos de especie están basadas en propiedades de los organismos que son de interés para cada investigación, por lo que son aportaciones valiosas para una mejor comprensión de la naturaleza. Por lo tanto, identificar los

elementos compartidos entre los distintos conceptos de especie, en lugar de las diferencias, ha permitido reconocer que de manera generalizada las especies están conformadas por segmentos de linajes de poblaciones que evolucionan de manera independiente (De Queiroz, 2007), en donde un “linaje” está definido como una secuencia de ancestros y descendientes (Simpson, 1961; Hull, 1980). Bajo este contexto, la definición asociada a cada concepto de especie puede ser vista como una propiedad secundaria de la especie. Estas propiedades secundarias pueden emerger o no a lo largo del proceso de especiación, en momentos diferentes y no necesariamente en el mismo orden, dependiendo de las características biológicas de cada grupo (De Queiroz, 1988 y 2007). El descubrimiento de alguna de estas propiedades en un conjunto de poblaciones, a través de alguna metodología de aproximación a la delimitación de especies, tales como la evaluación de caracteres morfológicos o la secuenciación de genes, representa la evidencia de que existe un proceso de especiación. Por lo tanto, la presencia de cualquiera de estas propiedades, no es garantía, pero es evidencia que respalda la hipótesis de que estas representan linajes con evolución independiente. De tal forma, una hipótesis de delimitación de especies puede ser planteada con base en una de sus propiedades secundarias y reforzada con base en diferentes líneas de evidencia que proporcionan información sobre las propiedades que las delimitan (De Queiroz, 2007).

El objetivo principal de este trabajo es la obtención de información a nivel molecular que permita evaluar la riqueza de especies dentro de un grupo de avispas parasitoides altamente diverso y poco estudiado de la familia Braconidae. En este caso, la congruencia entre los resultados de diferentes análisis con base en la información molecular permitirá determinar si existe evidencia para respaldar la hipótesis de una mayor diversidad en un grupo de avispas. Este trabajo sienta las bases para estudios posteriores que permitan integrar la evidencia de diferentes áreas de la biología para corroborar la existencia de linajes independientes.

La combinación de herramientas y conceptos de diferentes ramas de la biología a la luz de la era genómica ha revolucionado la manera en que entendemos los procesos evolutivos y de especiación. La tecnología que permite secuenciar genomas completos dio lugar desarrollo de tecnologías de secuenciación masiva paralela (SMP) que permiten secuenciar partes representativas del genoma de manera práctica y a un costo accesible (Glenn *et al.*, 2017; Goodwin *et al.*, 2016), promoviendo el uso de varios miles de loci para estudios evolutivos.

Las técnicas de SMP permiten recuperar fragmentos de ADN distribuidos a lo largo del genoma usando diversos protocolos de laboratorio tales como la secuenciación por genotipado (GBS), la secuenciación asociada a sitios de restricción (RADseq) y sus variantes ddRAD, RADcap, RADTags así como la obtención de secuencias asociadas a elementos ultra conservados

(Andrews *et al.*, 2016; Goodwin *et al.*, 2016). Entre ellos destaca el RADseq (Restriction Associated DNA sequencing) (Baird *et al.*, 2008; Davey y Blaxter, 2010) y la captura de secuencias asociadas a elementos ultra conservados (Faircloth *et al.*, 2012; Lemmon *et al.*, 2012). Por sus características y el tipo de datos que obtiene, la técnica de RADSeq ha sido utilizada mayormente para estudiar organismos no modelo y particularmente en estudios filogeográficos y de genética de poblaciones en animales (Harvey *et al.*, 2016). Por otro lado, se ha visto que la captura de secuencias de regiones ultrconservadas permite resolver relaciones a un nivel de tiempo más profundo, ya que los datos obtenidos corresponden a secuencias ultra conservadas flanqueadas por fragmentos variables (Faircloth *et al.*, 2015).

Dependiendo de la técnica usada, las muestras biológicas deben prepararse de manera particular, a este proceso se le conoce como “construcción de librerías”. En el protocolo general de la técnica de RADseq, esta construcción está fundamentada en la existencia de secuencias cortas de ADN correspondientes a sitios de restricción distribuidos de manera representativa a lo largo del genoma (Glenn *et al.*, 2017; Harvey *et al.*, 2016). Estos sitios son reconocidos y cortados por enzimas de restricción específicas, lo cual permite obtener una gran cantidad de fragmentos cortos correspondientes a las secuencias de ADN adyacente a los sitios de restricción, y que por su amplia distribución a lo largo del genoma son representativos del mismo (Baird *et al.*, 2008; Davey y Blaxter, 2010; Hoffberg *et al.*, 2016).

El protocolo implementado en las técnicas de SMP, incluye la preparación de una solución única donde se encuentran todos los fragmentos pertenecientes a diferentes genomas (individuos) que se desea secuenciar simultáneamente. Por lo tanto, los fragmentos correspondientes a cada muestra deben marcarse molecularmente de manera única durante su procesamiento en el laboratorio (Glenn *et al.*, 2017; Graham *et al.*, 2015). Dado que las enzimas de restricción cortan el ADN en sitios específicos, los fragmentos genómicos poseen extremos adhesivos con secuencias conocidas; a estos fragmentos se les pueden unir etiquetas o adaptadores especiales que permitirán reconocer las muestras individualmente durante el procesamiento bioinformático. Los fragmentos resultantes son sometidos a una selección por tamaños para tener un mayor control sobre las regiones particulares que estarán representadas en la librería final (Peterson *et al.*, 2012). De esta manera, al comparar secuencias del mismo tamaño y correspondientes al mismo sitio de restricción, esta técnica permite identificar polimorfismos en loci homólogos, a bajo costo y sin necesariamente contar con un genoma de referencia (Baird *et al.*, 2008; Davey y Blaxter, 2010; Glenn *et al.*, 2017).

En el presente trabajo, se implementó el protocolo de 3RADseq que busca mejorar el aprovechamiento de ADN mediante la digestión de fragmentos usando tres enzimas de restricción

(Glenn *et al.*, 2017). La principal variación presentada en este protocolo es que la tercera enzima impide la formación de dímeros y quimeras. Estos productos indeseados son el resultado de la ligación de fragmentos genómicos que poseen extremos adhesivos complementarios, pero que no son continuos, y de la ligación de adaptadores con extremos adhesivos complementarios (Glenn *et al.*, 2017). La ventaja principal de este protocolo es que la digestión enzimática y la ligación de adaptadores se lleva a cabo al mismo tiempo, permitiendo la destrucción de productos indeseados, disminuyendo el número de pasos necesarios para la construcción de librerías y el número de limpiezas necesarias, resultando en un mejor aprovechamiento del ADN de las muestras.

La inferencia de filogenias a diferentes escalas de tiempo evolutivo con base en información de SMP es un tema controvertido (Eaton, *et al.*, 2017; Leaché *et al.*, 2015). Algunos estudios que comparan distintos protocolos de secuenciación reducida del genoma han encontrado que para especies cercanamente relacionadas o poblaciones de divergencia reciente, la técnica de RADSeq recupera genealogías de manera robusta, por lo que ha sido recomendada para estudios a nivel de genética de poblaciones y filogeografía (Davey y Blaxter, 2010; Leaché *et al.*, 2015). No obstante, es preferible recurrir a otras técnicas de secuenciación, como la secuenciación de elementos ultra conservados cuando se requiere una mayor cantidad de información por locus para estimar filogenias a escalas de tiempo más profundas (Harvey *et al.*, 2016; McCormack *et al.*, 2012).

En el presente trabajo se utilizó la información genética del gen mitocondrial COI, así como información de genes nucleares obtenida mediante 3RADseq para delimitar especies dentro de un género de avispa parasitoide. La delimitación de especies con base en este tipo de información se realizó utilizando, primero, una aproximación de descubrimiento aplicada a la información del gen COI. Se realizó un análisis de reconstrucción filogenética mediante inferencia bayesiana que aportó una primera aproximación a la estructura geográfica de la variación. Posteriormente se realizó un análisis de distancias genéticas. Se ha observado que la variación entre secuencias del gen mtCOI permite delimitar especies de manera robusta en himenópteros y otros grupos de insectos (Hajibabaei *et al.*, 2006; Hebert, Ratnasingham, *et al.*, 2003). No obstante, una de las principales limitaciones de este método es que no existe un umbral determinado de diferenciación que funcione para todos los grupos. Comúnmente se aplica un criterio de 1 a 3% de diferenciación, y particularmente en himenópteros se ha visto que un valor del 2% obtiene buenos resultados (Fernández-Flores *et al.*, 2013; Schwarzfeld y Sperling, 2015; Stahlhut *et al.*, 2013). Aunado a esto, se realizó un análisis de delimitación de especies usando el método General Mixed Yule Coalescent model (GMYC; Pons *et al.*, 2006), el cual ha sido usando exitosamente para delimitar especies usando el locus del código de barras, aunque se ha visto que en algunos casos suele sobre estimar el número de especies (Talavera *et al.*, 2013).

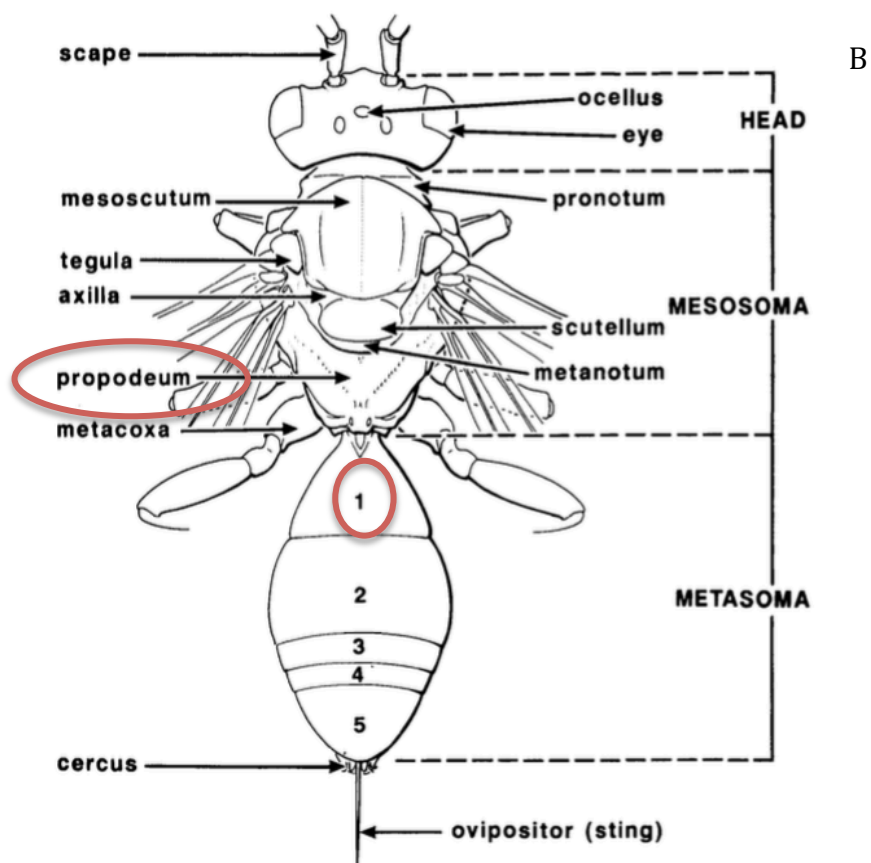
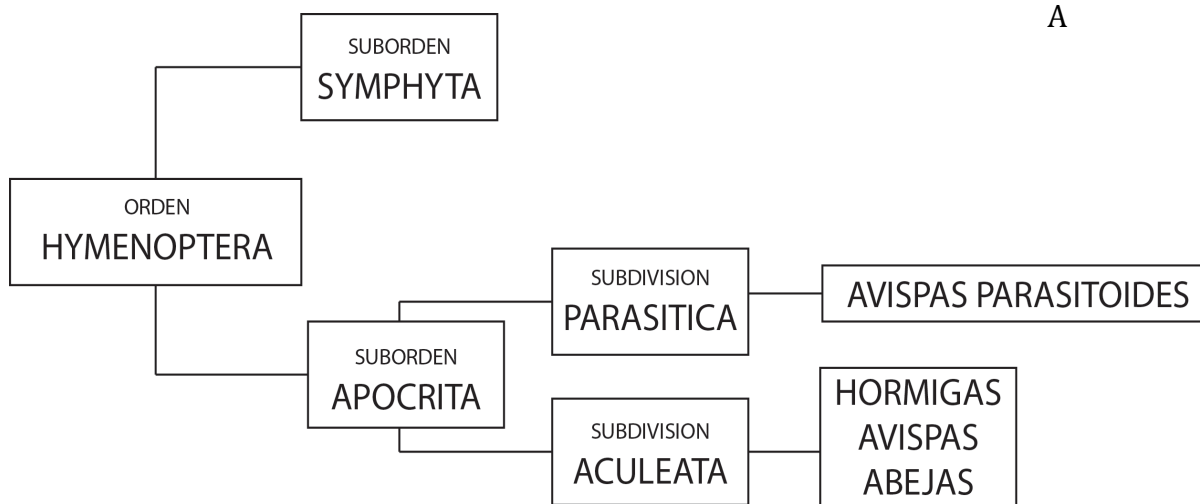
Por otro lado, se realizaron análisis de descubrimiento usando el grupo de datos nucleares. Se llevó a cabo una reconstrucción filogenética mediante inferencia bayesiana, así como una evaluación de la estructura genética usando el programa Structure, el cual es sensible a la cantidad de datos faltantes y al número de loci analizados, por lo cual resulta conveniente usar información obtenida mediante técnicas de RAD (Pritchard *et al.*, 2000). No obstante, este análisis debe ser realizado con un subconjunto de datos nucleares debido al alto poder computacional que requiere su algoritmo. Finalmente se realizó un análisis de validación de especies con base en los datos nucleares usando el programa BPP (Rannala *et al.*, 2017), el cual utiliza el modelo coalescente multiespecies para estimar la probabilidad posterior de diferentes modelos de delimitación, reconociendo como especies a los grupos sin flujo genético reciente (Yang *et al.*, 2010). Una de las principales críticas a estos métodos de validación, es que están basados en el supuesto de poco flujo genético y la ausencia de polimorfismos ancestrales, por lo tanto, se ha visto que delimitan eficientemente tanto estructura genética a nivel poblacional como estructura entre linajes con evolución independiente (Sukumaran *et al.*, 2017). En este contexto, el presente estudio busca integrar la evidencia de los métodos de descubrimiento y de validación de especies usando tanto datos mitocondriales como nucleares, para investigar la variación genética y la riqueza taxonómica en un grupo de avispas parasitoides poco estudiado y de muy alta diversidad.

#### *Grupo de estudio: Hymenoptera y avispas parasitoides*

El orden Hymenoptera es un grupo altamente diverso y relativamente poco estudiado, conformado por abejas, hormigas, avispas y entre ellas las avispas parasitoides. Los parasitoides se caracterizan porque invariablemente matan a su huésped, y en el caso de las avispas conforman un grupo particularmente diverso (Forbes *et al.*, 2018; Quicke, 2015).

Dentro del orden Hymenoptera, las avispas parasitoides se agrupan bajo el suborden Apocrita (Fig. 1A), caracterizado morfológicamente por la constricción del propodeo y el primer segmento abdominal (Fig. 1B) (Quick, 1997). Dentro de este suborden, las avispas parasitoides están agrupadas en la división Parasitica, la cual está conformada por un conjunto heterogéneo de superfamilias sin aguijón, y de hábitos principalmente parasíticos y parasitoides (Fig. 1A) (Hanson y Gould., 2006).





**Figura 1.** A) Esquema general de la clasificación del orden Hymenoptera. B) Vista dorsal de las principales divisiones morfológicas de una avispa del suborden Apocrita, en rojo se muestra el propodeo y el primer segmento abdominal (Fuente: Goulet & Huber, 1993).

La primera aparición de avispas con hábitos parasitoides en el registro fósil data del Jurásico medio, durante la primera diversificación del orden, hace aproximadamente 160 millones de años (Nieves-Aldrey *et al.*, 2006). Aunque el parasitoidismo es una estrategia que evolucionó independientemente en siete órdenes de insectos (Diptera, Coleoptera, Lepidoptera, Trichoptera, Neuroptera, Strepsiptera e Hymenoptera), el orden Hymenoptera es el que incluye más especies de parasitoides, los cuales conforman el grupo más grande dentro del orden e incluso se ha sugerido que componen el grupo de insectos más diverso del planeta (Forbes *et al.*, 2018).

Este grupo presenta dos estrategias principales de parasitoidismo: idiobiosis o koinobiosis. La idiobiosis se caracteriza porque el hospedero deja de moverse y desarrollarse cuando es invadido, mientras que en la koinobiosis el hospedero continúa su desarrollo durante un periodo de tiempo (Quicke, 1997). Generalmente, la estrategia koinobionte implica endoparasitoidismo y conlleva modificaciones fisiológicas que permiten contrarrestar las defensas inmunológicas del hospedero, mientras que la estrategia idiobionte generalmente implica ectoparasitoidismo (Quicke, 2015). La variedad de hospederos asociados a avispas parasitoides es muy alta. Se han registrado interacciones en todos los órdenes de insectos en la etapa de larva, pupa o imago, e incluso en otros grupos de artrópodos como arañas (Forbes *et al.*, 2018; Korenko *et al.*, 2018; Quicke, 2015). La mega-diversidad que presentan las avispas parasitoides y el hecho de que invariablemente matan a su hospedero, les confiere un papel ecológico clave en la mayoría de los ecosistemas que habitan, ya que tienen un efecto importante en la dinámica de las poblaciones de insectos y por eso han sido ampliamente utilizados en programas de control biológico (Quicke, 2015; Smith *et al.*, 2008; Veijalainen *et al.*, 2012).

#### *Ichneumoidea: Braconidae: Doryctinae*

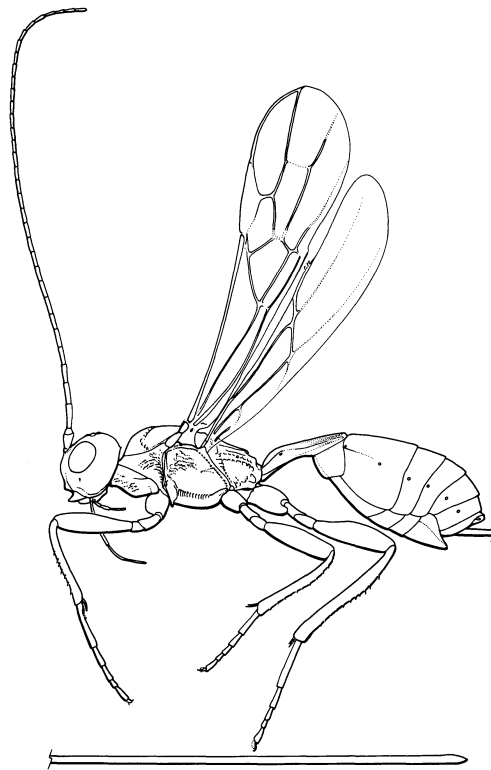
El orden Hymenoptera agrupa a aproximadamente 20 superfamilias (Sharkey *et al.*, 2012). Entre ellas, la superfamilia Ichneumonoidea agrupa a las dos familias más diversas del orden: Braconidae e Ichneumonidae. Braconidae es un grupo monofilético compuesto por aproximadamente 46 subfamilias, 1,057 géneros y 19,434 especies descritas, aunque se estima que la diversidad real es de aproximadamente 40,000 especies de las cuales, la mayoría se encuentran en los trópicos del planeta (Sharkey *et al.*, 1992; Jones *et al.*, 2009).

Dentro de Braconidae se encuentra la subfamilia Doryctinae, la cual agrupa aproximadamente 165 géneros y 1,300 especies descritas, aunque se estima que hay aproximadamente 2,000 especies más por describir (Jones *et al.*, 2009). Las avispas de la subfamilia

Doryctinae se distinguen por tener la cabeza particularmente robusta y una hilera de espinas a lo largo de la tibia anterior, así como por la inserción separada de dos ductos secundarios de veneno en el conducto primario, el ápice del ovopositor altamente esclerotizado con un doble nodo en la valva superior, y una forma aserrada en la valva inferior (Zaldivar-Riverón *et al.*, 2008). Su distribución es principalmente tropical y subtropical, siendo especialmente diversas en el continente Americano (Quicke, 2015).

Los dorictinos son avispas ectoparasitoides idiobiontes de insectos que se encuentran resguardados de alguna manera, por ejemplo, coleópteros barrenadores de la madera. La localización del hospedero la realizan mediante la vibración que este produce al alimentarse, aunque la biología de la mayoría de los géneros tropicales de dorictinos se desconoce (Belokobylskij *et al.*, 2004; Quicke, 2015).

Se ha estimado que la radiación de especies de dorictinos en el continente Americano ocurrió después del Eoceno tardío hace aproximadamente 30 millones de años, y que posteriormente hubieron eventos de dispersión hacia el continente Europeo y Asiático, que dan cuenta de la amplia distribución de ciertos géneros como *Ecphylus* y *Heterospilus* (Zaldivar-Riverón *et al.*, 2008).



**Figura 2.** Braconidae: Doryctinae. Dibujo que muestra las principales características del grupo, así como la longitud que puede alcanzar el ovipositor de la hembra (Fuente: Goulet & Huber, 1993).

### *El género Stenocorse*

Dentro de la subfamilia Doryctinae se encuentra el género monotípico *Stenocorse* Marsh con su única especie reconocida, *Stenocorse bruchivora* (Crawford, 1910). Esta especie se distribuye ampliamente en América, desde el sur de Estados Unidos hasta el centro de Argentina (Shenefelt y Marsh, 1976). El holotipo de la especie corresponde a una localidad en Texas, pero la descripción menciona que también ocurre en Arizona (E.U.), México, Perú y Hawaii (Marsh, 1968), además de varias especies sin describir en el Neotrópico.

El grupo está compuesto de avispa que miden aproximadamente 5 mm de longitud. Estas se distinguen fácilmente de otros géneros por tener un cuerpo particularmente robusto. Los machos no poseen un estigma en el ala anterior, y la vena 2SR se encuentra bien esclerotizada. Se ha reportado que son avispa ectoparasitoides idiobiontes de hábitos generalistas, ya que a la fecha se conoce que parasitan larvas de 62 especies diferentes de brúquinos (Coleoptera: Chrysomelidae: Bruchinae), los cuales, durante el estado larval se alimentan de semillas de diferentes especies de leguminosas (López-Martínez, *et al.*, 2003).

Hasta la fecha se considera que *Stenocorse* es un género monotípico de hábitos generalistas en cuanto a la elección de especie de hospedero, aunque se ha sugerido que el género está compuesto por un complejo de varias especies (Marsh, 1968), con preferencias más particulares, como se ha visto en otros grupos de artrópodos (Hambäck *et al.*, 2013; Smith *et al.*, 2008, 2006, 2017). Sin embargo, la diversidad del grupo no ha sido explorada debido a la falta de conocimiento sobre la biología, hábitos y relaciones filogenéticas.

Las avispa del género *Stenocorse* son haplo-diploides, es decir que los machos son haploides y nacen a partir de huevos sin fertilizar, mientras que las hembras diploides son producto de huevos fertilizados (Pérez y Bonet, 1984). El ciclo de vida está estrechamente relacionado con el del brúquino, el cual en etapa adulta generalmente coloca un huevo sobre el fruto de una leguminosa y al nacer la larva hace un agujero en la vaina para alimentarse del endospermo de las semillas en el interior (Romero Gómez *et al.*, 2009). La avispa adulta utiliza el agujero hecho por el brúquino para ingresar a la vaina, y utiliza las antenas para detectar a la larva en las semillas y poner un huevo en su superficie (Pérez y Bonet, 1984). El periodo larval de la avispa dura aproximadamente 15 días y el adulto vive aproximadamente otros 15 días (David Schneider, comentario personal).



**Figura 3.** Ejemplar macho de *Stenocorse* colectado durante este trabajo en Cuajiniquil, Costa Rica.

#### *Interacciones parasitoide hospedero y procesos biogeográficos*

Entre los animales más diversos y mejor adaptados a los diferentes ecosistemas terrestres y acuáticos se encuentran los insectos. Los procesos que han dado lugar a la riqueza de especies de insectos que apreciamos hoy en día han sido motivo de estudio desde los comienzos de la biología y continúan siendo fuente de preguntas de investigación en diferentes áreas de esta disciplina.

En particular, los procesos de especiación en insectos fitófagos tales como los escarabajos brúquinos, pueden estar promovidos por la incorporación de nuevas plantas huéspedes a la dieta o por la diversificación de las poblaciones de plantas, lo cual promueve la especiación paralela de sus depredadores herbívoros (Yu *et al.*, 2018). Estos procesos pueden estar acompañados por la interrupción parcial o total del flujo génico entre las poblaciones de fitófagos, dando lugar al desarrollo de aislamiento reproductivo y a la formación de nuevas especies mediante especialización ecológica (Forbes *et al.*, 2017; Stireman *et al.*, 2006). No obstante, la intensidad del flujo génico y la cantidad de tiempo transcurrido desde la separación de dos poblaciones, son los factores determinantes en la formación de especies diferentes (Hey y Pinho, 2012), lo cual está fuertemente influenciado por las características biológicas y ecológicas de los organismos.

Las variaciones en las poblaciones de fitófagos causadas por cambios de huésped o por especiación paralela a las plantas hospederas, pueden tener un efecto en cascada sobre los procesos

de especiación de las poblaciones de los parasitoides asociados (Forbes *et al.*, 2017). El proceso de especiación acoplada debido a este efecto en cascada resultaría en patrones filogenéticos similares en ambos grupos (Hamerlinck *et al.*, 2016). No obstante, los mecanismos de especiación ecológica que pueden producir efectos en cascada sobre las poblaciones de sus depredadores tienen un mayor efecto en el caso de las avispas con una estrategia de vida endoparasitoide (Quicke, 2015). Las larvas de las avispas que presentan esta estrategia han desarrollado adaptaciones, principalmente fisiológicas, como respuesta al sistema inmunológico del hospedero o a elementos como bacterias endosimbiontes o toxinas obtenidas durante la ingesta de plantas (Oliver *et al.*, 2003; Quicke, 2015). Por otra parte, la amplia distribución geográfica de algunos parasitoides, también puede promover la especiación a través de vicarianza, debido al surgimiento de barreras geográficas tales como cultivos controlados con pesticidas, cuerpos de agua o elevaciones altas, afectando a organismos tanto endo como exoparasitoides (Lozier *et al.*, 2009)

Los escarabajos de la subfamilia Bruchinae se distribuyen en todo el mundo, aunque la mayor riqueza de especies se encuentra en las regiones tropicales y subtropicales (Romero Gómez *et al.*, 2009). Las larvas de brúquinos se alimentan de semillas de plantas de aproximadamente 34 familias diferentes, aunque un 80% se concentra en plantas de la familia Fabaceae y la mayoría poseen un rango de hospederos restringido (Romero-Nápoles *et al.*, 2002; Tuda *et al.*, 2006). Dada la estrecha interacción brúquino-planta, se ha planteado la posibilidad de un escenario de evolución conjunta (Center y Johnson, 1974; Jermy y Szentesi, 2003). Sin embargo, aún se requiere de mayor evidencia en cuanto a la bioquímica de las plantas, la biogeografía y ciclo de vida del hospedero, así como mayor investigación en cuanto a sus relaciones filogenéticas para descubrir los patrones de la interacción y los procesos subyacentes (Romero-Nápoles *et al.*, 2002).

Mesoamérica y el Neotrópico en general son territorios caracterizados por importantes procesos geológicos que han marcado y delineado su biodiversidad. Por ejemplo la unión de Centro y Sudamérica o la formación de las principales cadenas montañosas en México y Centro América han sido críticos para determinar la estructura geográfica de la diversidad genética de varios grupos de organismos, principalmente vertebrados y plantas (Bryson, *et al.*, 2012; Mastretta-Yanes *et al.*, 2015; Morrone, 2015; Ornelas *et al.*, 2013). No obstante, puede resultar inadecuado aplicar esta generalización a los patrones de diversificación en insectos, ya que en esto están implicadas características particulares de su historia de vida, como el tipo de reproducción, la vagilidad o el ciclo de vida, además de las particularidades de su historia demográfica, tales como cuellos de botella o deriva génica. Estos son factores determinantes en el flujo génico y la historia evolutiva, por lo que deben ser tomados en cuenta para proponer una hipótesis robusta sobre la historia evolutiva de un grupo (Mendoza-Ramírez *et al.*, 2018; Pedraza-Lara *et al.*, 2015).

## JUSTIFICACIÓN

A pesar de la enorme diversidad y complejidad de la familia Braconidae, su estudio ha estado restringido a caracteres morfológicos y un número limitado de loci mitocondriales y nucleares (p. ej., Samacá-Sáenz, Belokobylskij, Quicke, & Zaldívar-Riverón, 2016; Zaldívar-Riverón et al., 2018), lo cual ha limitado la resolución de las relaciones evolutivas en algunos grupos, en parte por falta de señal filogenética, especialmente en presencia de divergencia reciente o morfología conservada.

En el presente trabajo, se empleó el método de SMP, 3RADseq para caracterizar la diversidad genética de un género de avispa estrechamente relacionada con los brúquinos que plagan plantas de importancia económica, como el frijol (Rodríguez, 2018). La técnica seleccionada está diseñada para recuperar una gran cantidad de loci de organismos pequeños, sin un genoma de referencia e incluso de muestras antiguas.

El género *Stenocorse* está compuesto por avispa parasitoides muy pequeñas, con morfología compleja, y en la literatura se considera como un género de amplia distribución y hábitos generalistas en cuanto a la selección de hospedero. En este trabajo se caracterizó la diversidad genética del grupo para evaluar su composición taxonómica, ya que se sospecha la existencia de una mayor riqueza específica. Esto permitió inferir algunos de los procesos que dieron lugar a la estructura poblacional del grupo y su diversificación. Este trabajo sienta las bases para estudiar con mayor profundidad la diversidad del género *Stenocorse*, las interacciones ecológicas entre parasitoide, huésped herbívoro y planta y su evolución e impacto económico y en los ecosistemas.

## **OBJETIVOS**

### *General*

Investigar la sistemática y evolución de las avispas parasitoides del género *Stenocorse* empleando secuencias de ADN nuclear y mitocondrial con énfasis en el territorio Mexicano.

### *Particulares*

- Evaluar la presencia de estructura genética en las poblaciones muestreadas.
- Reconstruir las relaciones genealógicas entre los ejemplares de *Stenocorse* examinados a lo largo de su área de distribución, desde el sur de Estados Unidos hasta el norte de Argentina, con énfasis en el territorio Mexicano.
- Aplicar herramientas de delimitación de especies para determinar la presencia de linajes con evolución independiente.
- Inferir qué eventos pudieron dar origen a la variación genética observada en el contexto geográfico de la distribución del género.





**GENETIC STRUCTURE, SPECIES LIMITS AND EVOLUTION OF THE PARASITOID WASP GENUS STENOCORSE (BRACONIDAE: DORYCTINAE) BASED ON NUCLEAR 3RAD AND MITOCHONDRIAL DATA**

Journal:	<i>Systematic Entomology</i>
Manuscript ID	Draft
Wiley - Manuscript type:	Original Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Delgado-Machuca, Natalia; Universidad Nacional autónoma de México, Instituto de Biología Meza-Lázaro, Rubi; Universidad Nacional autónoma de México, Instituto de Biología Sarmiento, Carlos; Universidad Nacional de Colombia, Instituto de Ciencias Naturales Laboratorio de Sistemática y Biología Comparada de Insectos Romero-Nápoles, Jesús; Colegio de Posgraduados Amarillo-Suárez, Ángela; Pontificia Universidad Javeriana, Departamento de Ecología y Territorio Bayona-Vásquez, Natalia; University System of Georgia Zaldívar-Riverón, Alejandro; Universidad Nacional autónoma de México, Instituto de Biología
Taxon:	Hymenoptera
Topic:	Barcoding, Molecular, Phylogenomics
Abstract:	Stenocorse (Braconidae: Doryctinae) is an ectoparasitoid wasp genus widely distributed from southern USA to central Argentina, and it is known to attack over 60 bruchine beetle species that feed on a variety of legumes. Currently, this genus only comprises a single described species, <i>S. bruchivora</i> (Crawford, 1910), though it is probably composed of various additional, morphologically similar species. In this study we investigated the genetic structure, species limits and evolutionary relationships of sampled populations of <i>Stenocorse</i> obtained along its known geographic distribution, with an emphasis on the Mexican territory. For this, we generated DNA sequences from both the mitochondrial (mt) barcoding locus and nuclear genome data obtained with the 3RAD technique. The Bayesian phylogenetic reconstructions derived from the two data sets were mostly congruent, both showing a marked geographic structure. Four genetic clusters were recovered with nuclear data, which were also congruent with the geographic distribution of the samples and the species delimitation analyses. We delimited six evolutionary lineages with the two data sources, and three additional species were also delimited with the barcoding locus alone. <i>Stenocorse</i> could have originated during the mid to early Miocene, and the

diversification events that led to its current species distribution probably occurred during the early Pliocene to late Pleistocene. Our species delineation analyses, phylogenetic reconstructions and rearing records suggest that species of *Stenocorse* have generalist feeding habits, contrary to what was expected due host plant specialization of the bruchine species and the cascading effect that this would have in the species diversification of their parasitoids.

Note: The following files were submitted by the author for peer review, but cannot be converted to PDF. You must view these files (e.g. movies) online.

Sup.mat\_7.tiff  
Sup.mat\_8.tiff

SCHOLARONE™  
Manuscripts

**GENETIC STRUCTURE, SPECIES LIMITS AND EVOLUTION OF THE PARASITOID  
WASP GENUS *STENOCORSE* (BRACONIDAE: DORYCTINAE) BASED ON NUCLEAR  
3RAD AND MITOCHONDRIAL DATA**

Natalia Delgado-Machuca<sup>1</sup>, Rubi N. Meza-Lázaro<sup>1</sup>, Jesús Romero-Nápoles<sup>2</sup>, Carlos E. Sarmiento-  
Monroy<sup>3</sup>, Ángela R. Amarillo-Suárez<sup>4</sup>, Natalia J. Bayona-Vásquez<sup>5</sup> and Zaldívar-Riverón  
Alejandro<sup>1,\*</sup>

<sup>1</sup> Colección Nacional de Insectos, Instituto de Biología, Universidad Nacional Autónoma de México,  
3er. circuito exterior s/n, Cd. Universitaria, Copilco, Coyoacán, A. P. 70-233, C. P. 04510, Cdmx.,

México; <sup>2</sup> Laboratorio de Sistemática y Biología Comparada de Insectos, Instituto de Ciencias

Naturales, Universidad Nacional de Colombia, A. A. 7495, Bogotá, Colombia; <sup>3</sup> Instituto de  
Fitosanidad, Colegio de Postgraduados, Km 36.5 carr. México- Texcoco, Montecillo, Edo. de

México, México; <sup>4</sup> Departamento de Ecología y Territorio, Facultad de Estudios Ambientales y

Rurales, Pontificia Universidad Javeriana, Bogotá, Colombia; <sup>5</sup> Department of Environmental  
Health Science, University of Georgia, Athens, GA, USA.

**Corresponding author:** AZR (azaldivar@ib.unam.mx).

**Running head.** Genetic structure, species limits and evolution of *Stenocorse*.

## **Abstract**

*Stenocorse* (Braconidae: Doryctinae) is an ectoparasitoid wasp genus widely distributed from southern USA to central Argentina, which is known to attack over 60 bruchine beetle species that feed on a variety of legumes. This genus currently comprises a single described species, *S. bruchivora* (Crawford, 1910), though it is probably composed of additional, morphologically similar species. In this study we investigated genetic structure, species limits and evolutionary relationships of sampled populations of *Stenocorse* obtained along its known geographic distribution, with an emphasis on the Mexican territory. For this, we generated DNA sequences from both the mitochondrial (mt) barcoding locus and nuclear genome data obtained with the 3RAD technique. The Bayesian phylogenetic reconstructions based on the two data sets were mostly congruent, both showing clear geographic structure. Four clusters were recovered with nuclear data, which were also congruent with the geographic distribution of the samples and the species delimitation analyses. We delimited six evolutionary lineages with the two data sources, and three additional species were also delimited with the COI locus alone. *Stenocorse* probably originated during the mid to early Miocene, and the diversification events that led to its current species distribution likely occurred between the early Pliocene and late Pleistocene. Our species delineation analyses, phylogenetic reconstructions and rearing records suggest that the species of *Stenocorse* have generalist feeding habits, contrary to what was expected due to host plant specialization of the bruchine species and the cascading effect that this could have in the species diversification of their parasitoids.

## **Keywords**

Species delimitation, Hymenoptera, DNA barcoding, coalescent model, Next Generation Sequencing, RADseq.

## Introduction

Unveiling the species richness of megadiverse invertebrate taxa is one of the most challenging topics in systematics. Besides the vast number of species that remain to be described in these groups, a complete knowledge of their species richness is hampered by their high morphological similarity during different developmental stages (*e.g.* Che *et al.*, 2017; Eyer *et al.*, 2017).

Moreover, disagreements regarding conceptual and methodological approaches for species delimitation recurrently arise among authors (De Queiroz, 2007; Naciri *et al.*, 2015), making it difficult to establish systems for species delimitation in groups with extremely high diversity.

Parasitoid wasps are perhaps the most species rich group of insects in the planet, being characterised by having a wide range of life history traits (Forbes *et al.*, 2018; Quicke, 1997). These include interactions with hosts of several insect orders as well as other groups of arthropods, where they can lay their eggs inside or outside of the hosts' different developmental stages. (Korenko *et al.*, 2018; Quicke, 1997, 2015). The extraordinary species richness of the parasitoid wasps and the fact that they almost invariably kill their hosts make them a key element of most terrestrial ecosystems, since they have important effects on the population dynamics of their hosts (Quicke, 1997). For this reason, various species of parasitoid wasps have been used for almost a century in a number of biological control programs (Quicke, 2015; Smith *et al.*, 2008; Veijalainen *et al.*, 2012).

*Stenocorse* (Braconidae: Doryctinae) is a monotypic ectoparasitoid wasp genus with a single recognised species, *S. bruchivora* (Crawford, 1910), described from Texas, USA. In this description, however, the author mentioned that the species also occurs in Arizona, Mexico, 'the Canal Zone' (in Panama), Peru and Hawaii, and that the genus actually is composed of various additional overlooked species, though he did not give further details (Marsh, 1968). Currently, *Stenocorse* is reported to be widely distributed along the Nearctic and Neotropical regions, from southern USA to central Argentina (Shenefelt & Marsh, 1976; this study), though no taxonomic studies have attempted to assess its actual number of species.

*Stenocorse* has been reported to attack over 60 different bruchine (Chrysomelidae) beetle species that feed on a variety of legumes (López-Martínez *et al.*, 2003), some of which

are of great economic importance. Most of these host records, however, have been assigned to *S. bruchivora* without carrying out any species identification assessment, mainly due to the conserved external morphology displayed by members of the genus. Thus, confirmation of additional species of *Stenocorse* with associated biological records is essential to assess with more certainty the host range preferences among members of the genus.

The mitochondrial (mt) DNA genes are widely used markers in molecular phylogenetic and species delineation studies because of their ease of use due to their high number of copies compared to protein coding nuclear genes, high mutation rate and low exposure to recombination because of its purely maternal way of inheritance (Hebert *et al.*, 2003a,b). In particular, the DNA barcoding locus (Hebert *et al.*, 2003b), represented by a 658 bp fragment of the mtDNA cytochrome oxidase I (COI) gene, has been extensively employed over the last two decades to delimit a vast number of animal species, including insects (Hajibabaei *et al.*, 2006; Hebert *et al.*, 2003a; Ratnasingham & Hebert, 2013). The use of this locus alone, however, has been mainly criticised due to its problems for delimiting recently diverged species in the presence of incomplete lineage sorting or by the lack of a ‘barcode gap’, i.e. an overlapping of intra and interspecific genetic distances (van Velzen *et al.*, 2012).

The development of next generation sequencing (NGS) methods has made it possible to obtain a vast amount of genetic information for non-model organisms across their nuclear genome (Baird *et al.*, 2008; Goodwin *et al.*, 2016; Cruaud *et al.*, 2018). Among the techniques that generate a reduced representation of the nuclear genome are the RADseq approaches, which produce thousands of short fragments of DNA associated to restriction sites throughout the genome (Baird *et al.*, 2008). Different RADseq techniques are used to address population genetics and phylogeographic studies (Andrews *et al.*, 2016; Glenn *et al.*, 2017; Salas-Lizana & Oono, 2018), but also to carry out phylogenetic studies of closely related taxa (*e.g.* Eaton *et al.*, 2017) and to assess species limits and speciation processes (Battey & Klicka, 2017; Cruaud *et al.*, 2018).

In this study, we generated both mt sequence and nuclear 3RAD (Glenn *et al.*, 2017) data for specimens of *Stenocorse* that were collected across various parts of its geographic distribution,

with a focus on the Mexican territory. This region is characterised by having a complex geological and paleoclimatic history and high levels of endemism (Ferrusquía-Villafranca, 1998; Marshall *et al.*, 2000). We employed the sequencing data to: a) assess the presence of genetic structure to identify the number of evolutionary lineages (i.e. species) in the sampled populations, and b) to perform phylogenetic reconstruction and relaxed molecular clock analyses to make inferences about the evolutionary history and host range preferences in the genus.

## **Methods**

### *Taxon sampling*

We examined a total of 119 specimens assigned to *Stenocorse*. Most of the specimens were reared from legumes together with their potential bruchid hosts, though few were collected with sweeping nets. The examined specimens were collected in several localities in the USA and six Neotropical countries, with an emphasis in central and southeast Mexico. All specimens were stored in 96% alcohol until they were processed in the laboratory. Both 3RAD and COI data (86 and 79 specimens, respectively) were generated for up to five specimens per locality. 3RAD data could not be obtained for specimens from localities situated in Quintana Roo (Mexico), Venezuela, Argentina and Brazil due to the low amount of extracted genomic DNA. COI sequences on the other hand could not be obtained for few specimens because all DNA template had to be used for the 3RAD technique. Six specimens of the doryctine genus *Heterospilus* Haliday were also sequenced to root the trees.

The samples from Mexico were grouped according to their morphotectonic provinces and subprovinces of provenance (Ferrusquía-Villafranca, 1998): Central Plateau province (CP: Queretaro, Mexico), Trans-Mexican Volcanic Belt province (TMVB: Estado de Mexico, Morelos and Puebla, Mexico), Pacific Coastal Plains subprovince, Sierra Madre del Sur province (PCP-SMS; Jalisco, Guerrero and Oaxaca, Mexico), Gulf Coastal Plain province (GCP; Veracruz, Mexico); Pacific Coastal Plain subprovince, Sierra Madre de Chiapas province (PCP-Ch; Chiapas, Mexico) and Yucatan Platform province (YP; Quintana Roo, Mexico).

We recorded the Fabaceae plant species from which the specimens of *Stenocorse* were reared, as well as their associated bruchine species to explore the host range preferences and host plant association of the delimited species. All Fabaceae plant and bruchine species were identified using the relevant literature (Kingsolver & Johnson, 1978; Johnson, 1983, 1990; Kingsolver, 1986, 1988; Johnson & Kingsolver, 1989; Romero & Johnson, 2000). A list with the examined specimens, collection localities, DNA voucher and GenBank accession numbers and the associated bruchine and plant species is given in Supplementary material 1.

### *Laboratory procedures*

Genomic DNA was extracted by gently puncturing the ventral area of the mesosoma of each specimen with a sterile pin under a ZEISS™ StemiDV4 stereoscopic microscope, and then using the EZ-10 Spin Column Genomic DNA Minipreps kit (BIO BASIC®, Toronto, Canada) following the manufacturer's protocol. DNA quality was examined on a 1.5% agarose gel and DNA concentration was quantified with the Life Technologies Qubit fluorometry system and the High Sensitivity DNA kit (Invitrogen, Carlsbad, USA).

MtDNA Sanger sequencing. We generated sequences of a fragment of 640 bp belonging to the COI mtDNA gene. Amplification was performed using the primers LCO (5' – GTCAACAAATCATAAAGATATTGG- 3') and HCO (5' - TAAACTTCAGGGTGACCAAAAAATCA 3') (Folmer *et al.*, 1994). The total PCR volume was of 15 µl, which consisted of 3 µl of 10xPCR Buffer, 0.1 µl of MgCl<sub>2</sub> 50 mM, 0.2 µl of each primer 10 µM, 0.2 µl of MyTaq™ DNA Polymerase (BIOLINE®, London, U.K.), 1 µl of DNA and 10.3 µl of ddH<sub>2</sub>O. PCR profiles were: 1 min at 94°C followed by 35 cycles of 94°C for 30 s, 46°C for 45 s and 72°C for 1min and a final extension at 72°C for 7 min. PCR products were sequenced in the Laboratory of Genomic Sequencing and Health of the Instituto de Biología, Universidad Nacional Autónoma de México (IB UNAM). COI sequences were edited in the program Geneious v.10.2.3 (Kearse *et al.*, 2012) and were manually aligned.

3RAD library construction and loci assembly. RADseq libraries were obtained following the Adapterama III protocol, which employs three enzymes to digest the genomic



DNA to prevent the formation of adapter dimers and chimeras during the reaction (Glenn *et al.*, 2017). Before digestion, each sample was purified using Sera-Mag™ SpeedBeads at a ratio of 1.8:1 Speed-Beads to DNA, respectively, following the protocol given by Rohland & Reich (2012). For each sample, between 80 to 200ng of genomic DNA were double digested for 1 hour at 37°C in a solution containing 1.5µl of 10x CutSmart® Buffer, 0.5µl of a rare cutter (XbaI 20 U/µl; NEB®; New England, USA), 0.5µl of a common cutter (EcoRI 20 U/µl; NEB®; New England, USA), 0.3µl of a third cutter to enhance the reaction (NheI 20 U/µl; NEB®, New England, USA), 2µl of left adapter at 5µM, 2µl of right adapter at 5µM, 0.2µl of ddH2O and 8µl of DNA template.

Ligation of digested fragments and custom adaptors with indexes was performed by adding the following onto the digestion solution: 2.75 µl of H<sub>2</sub>O, 1.5 µl of ATP at 10 µM, 0.5 µl of 10x Ligase Buffer (NEB® New England, USA) and 0.25 µl of T4 DNA Ligase at 400 U/µl. The solution was incubated at 22°C for 20 min and at 37°C for 10 min during two cycles and a final cycle of 80°C for 20 min. Each sample was then cleaned using Sera-Mag™ SpeedBeads at a ratio of 1.5:1 Speed-Beads to DNA, respectively.

PCR enrichment of the cleaned ligation products was performed on a Thermo Scientific Arktik™ thermal cycler using 5 µl of Kapa Long Range Buffer (Kapa Biosystems Wilmington, USA), 0.125 µl of KAPA LongRange DNA Polymerase at 5U/µl, 0.75 µl of dNTPs mix (10mM each dNTP), 1.75 µl of MgCL<sub>2</sub> at 25mM, 2.5 µl of iTru5 primer at 5 µM, 2.5 µl of iTru7 primer at 5 µM and 10 µl of ligation product. Adapter and primer combinations are listed in Supplementary Material 2. PCR temperature conditions were 98°C for 2 min of initial denaturation, followed by 18 cycles of 98°C for 20 s, 55°C for 15 s, 72°C for 1 min, and a final cycle of 72°C for 5 min. PCR products from each sample were cleaned with Sera-Mag™ at a ratio of 1.8:1 Speed-Beads to DNA, respectively. PCR products concentration was quantified with a Life Technologies Qubit® fluorometry system and High Sensitivity DNA kit (Invitrogen, Carlsbad, USA).

Final pools were constructed with equimolar volumes of each sample for sequencing. Libraries were sent to the Environmental Health Science Department, University of Georgia,

Athens, Georgia, USA, for size selection of fragments with an automated size-selection technology Pippin Prep (Sage Science<sup>®</sup>, Beverly, USA) in a range of  $550 \pm 15\%$  (470pb-630pb). Sequencing was performed on an Illumina platform HiSeq 2000 for paired-end reads of 150bp.

Sequence reads were demultiplexed using the program `process_radtags` implemented in the software STACKS v. 2.0 (Catchen *et al.*, 2013; Rochette *et al.*, 2017). The demultiplexed data were quality filtered and clustered into putative loci using only read one from the paired end data on the software pipeline iPyrad v.0.7.28 (Eaton, 2014) implemented on Jupyter Notebooks (Kluyver *et al.*, 2016). We excised R2 reads since it has been shown that single end data recover more phylogenetic information in the absence of a reference genome (Eaton *et al.*, 2017).

The filtering and clustering of reads were performed on the iPyrad pipeline with the default parameters. The clustering threshold parameter, which is the level of sequence similarity at which two fragments are considered homologous (Eaton, 2014), was set to 0.9 based on the observed clustering behaviour of reads using the scripts developed by Ilut *et al.* (2014). This approach seeks to minimize the number of false homozygosity (the splitting of a single locus into two) and false heterozygosity (clustering of paralogs) to find the optimum value of the clustering threshold parameter.

The joint estimation of heterozygosity, error rate, and the consensus base calling and filtering, were completed on two different branches of the assembly composed by the female (diploid) and male (haploid) samples, respectively. The parameter that varied on each branch was the maximum number of alleles allowed in individual consensus reads after accounting for sequencing errors, which for diploid and haploids samples was of 2 and 1, respectively. The resulting files from these steps were then merged to apply the final clustering, filtering and file formatting steps of the pipeline on a single dataset.

A considerably low number of loci ( $< 2\%$ ) were shared between the ingroup and outgroup taxa. We therefore employed the following strategy to assess the effect of including/excluding the outgroup in our phylogenomic analyses. Two matrices were built

without the outgroups based on different values of the parameter minimum sample per locus to evaluate the influence of missing data on the analyses and to detect samples that yielded a low number of loci (Eaton *et al.*, 2017; Tripp *et al.*, 2017). Matrix M1 was built with a minimum sample per locus of eight, i.e. that a locus would have to be shared across at least eight samples (ca. 10% of the data) to be included in the matrix. Matrix M2 was built using 40 as minimum sample per locus (ca. 50% of the data). We also generated two additional matrices (M3 and M4) using the above parameters but including the outgroup taxa. The trees derived from the M1 and M2 matrices were rooted according to the relationships obtained with the matrices that included the outgroups.

An additional matrix (M2.5) was generated to conduct analyses with the program Bayesian Phylogenetics and Phylogeography v.3.3 (BPP; Rannala & Yang, 2017) and STRUCTURE v.2.3.4 (Pritchard *et al.*, 2000) with the same parameters as matrix M2 but excluding the sample from Texas, USA, which yielded a considerably lower number of loci compared to the remaining samples.

### *COI analyses*

Genetic variation and haplotype diversity. We estimated the following population genetic metrics for each population group with the program DNASp v.6 (Rosas *et al.*, 2017): haplotype (h) and nucleotide ( $\pi$ ) diversity, number of polymorphic sites (S) and average number of nucleotide differences between haplotypes (k). We could not carry out neutrality tests among the delineated species due to their low number of specimens. We obtained a COI haplotype genealogy and geographical distribution of the *Stenocorse* haplotypes with a haplotype network using the TCS method (Templeton, 1992) implemented in the program PopART v.1.7.2 (Leigh & Bryant, 2015).

Species delimitation. We carried out single locus-based species delimitation analyses with the COI gene fragment using two different methods, the General Mixed Yule Coalescent model (GMYC; Pons *et al.*, 2006) and the 2% genetic divergence criterion (Hebert *et al.*, 2003b). The GMYC method models the transition point between cladogenesis and allele

coalescence, taking into account that the former occurs at a lower rate than the latter, causing a shift in the rate of branching in the genealogy (Pons *et al.*, 2006). The transition between species and population level processes then are delimited by a threshold in the topology (Pons *et al.*, 2006; Carstens *et al.*, 2013). For this, we obtained a maximum credibility ultrametric tree derived from a BEAST analysis that was ran for 500 million generations, sampling trees every 10,000 generations. Burn-in was set to occur after 25% of the sampled trees based on the observed Effective Sample Size with the program TRACER v.1.6 (Rambaut *et al.*, 2009). A GMYC topology was obtained implementing the single-threshold approach with the SPLITS package (<http://r-forge.r-project.org/projects/splits>) of the R environment v.3.4.2 (Development Core team, 2008).

For the 2% genetic divergence criterion, we calculated the uncorrected pairwise genetic distances among the examined specimens with the program MEGA v.7.0 (Kumar *et al.*, 2016). This criterion has proved to be a fast, generally reliable tool to explore species delimitation of animal taxa, including insects (Hajibabaei *et al.*, 2006; but see Meyer *et al.*, 2005; Wiemers *et al.*, 2007 for criticisms).

Phylogenetic relationships. We performed a partitioned Bayesian phylogenetic analysis for the COI data set with the program MrBayes v.3.2.6 (Ronquist *et al.*, 2003). The number of partitions and their substitution models were obtained with the program PARTITIONFINDER v.2.1.1 (Lanfear *et al.*, 2012). A single COI partition with the GTR + I + G substitution model was selected. The analysis was performed with two independent searches of 100 million generations each, with four MCMC chains and sampling trees every 5,000 generations. Burn-in was evaluated using the observed Effective Sample Size using the program TRACER v.1.6 and 25% of the sampled trees were discarded. A tree topology with branch lengths was summarized using the program TreeAnnotator v.1.10 (<http://beast.bio.ed.ac.uk/TreeAnnotator>). We considered the resulting clades from all Bayesian analyses to be significantly supported if they had a posterior probability  $\geq 0.95$ .

Divergence times. We estimated the times of divergence of selected clades using a relaxed molecular clock analysis with the program BEAST v.1.8.2 (Drummond *et al.*, 2002). In

the absence of known fossils for our group of study, we obtained rough estimation of divergence times calibrating the tree with the COI substitution rate ( $0.0168 \pm 0.0018$  substitutions per site per My) reported by Papadopoulou *et al.* (2010).

We used a single partition with the GTR+I+ G substitution model, a birth-death with incomplete sampling tree prior and a relaxed log-normal molecular clock. MCMC chains were run for 500 million generations, sampling trees every 10,000 generations. Burn-in was set to occur after 25% of the sampled trees based on the observed Effective Sample Size with the program TRACER v.1.6. We used the program TreeAnnotator v.1.8.2 available from the BEAST package to obtain the maximum clade probability tree with posterior probability of clades using the 'median heights' option.

Phylogenetic analyses for the COI data set were carried out in the CIPRES Science Gateway (Miller *et al.*, 2010) online service. All COI and 3RAD trees were edited using the program FigTree v.1.4.3 (Rambaut, 2009).

### *3RAD Analyses*

Genetic structure. We used the M2.5 matrix to perform an individual-based Bayesian clustering approach with the program STRUCTURE v.2.3.4. We used the u.str output file from the iPyrad pipeline, which is a SNP based format that includes only one randomly selected variable site per locus to evaluate the number of genetic clusters in the examined populations. We used an admixture model with correlated frequencies and assessed values of population differentiation (K) in six independent runs for K between 1 and 7 according to the groups that were formed in previous analyses. All runs were conducted for 1 million MCMC iterations with a burn-in of 100,000 and figures were generated based on the iterations with the highest posterior probability. We selected the K value based on the highest average value of the likelihood [LnP(D)] obtained.

Species delimitation. We performed a species delimitation analysis using the M2.5 matrix with the program BPP implemented in jupyter notebooks (Kluyver *et al.*, 2016). This program analyses DNA sequence alignments under the multispecies coalescent model

employing a reversible-jump MCMC to estimate the posterior probability of different delimitation models, recognising groups without recent gene flow as potential species (Rannala *et al.*, 2013; Yang *et al.*, 2010).

We applied the species delimitation algorithm, which requires a fixed guide species tree. We generated a topology with the samples separated into eight evolutionary lineages based on the estimated 3RAD Bayesian phylogenies and on the geographic distribution of the clades. We used a random subsample of 50 loci from the M2.5 matrix on each repetition of the analysis. The analysis was run for 200,000 MCMC generations, with a sampling interval of two and discarding trees from the first 20,000 generations as burn-in. We applied a diffuse gamma prior of  $\Theta \sim G(2, 200)$  and  $\tau \sim G(2, 2000)$  and applied 10 repetitions to verify convergence in our results.

Phylogenomic relationships. We reconstructed the evolutionary relationships among the examined taxa with the four 3RAD matrices using the SVDQuartets method (Chifman *et al.*, 2014, 2015) implemented in the program PAUP\* v.4.0a161 (Swofford, 2002). This method uses the coalescent model to reconstruct relationships among all subsets of four taxa based on multi-locus (Chifman *et al.*, 2014). The analyses were performed evaluating all possible quartets under a multispecies coalescent model with 500 bootstrap replicates (Reaz *et al.*, 2014).

We also performed Bayesian analyses for the four matrices with the program Exabayes v.1.5 (Aberer *et al.*, 2014). Two independent runs were performed for 500 thousand generations with the GTR substitution model and four MCMC chains. Trees were sampled every 500 generations and we discarded 25% of the sampled trees as burn-in. We used the remaining trees to build a majority rule consensus tree with posterior probabilities of clades.

The 3RAD analyses were performed on the Miztli supercomputer owned by the Dirección General de Cómputo y de Tecnologías de Información y Comunicación, National Autonomous University of Mexico (DGTIC, UNAM).

## Results

### *COI analyses*

Genetic variation and haplotype diversity. The results of the genetic diversity and neutrality tests for the examined populations groups of *Stenocorse* are shown in Supplementary Material 3. The COI data set had 121 polymorphic sites (*S*) and 31 haplotypes (*h*). The TMVB and PCP-SMS population groups, which had the highest number of examined specimens, contained the highest number of polymorphic sites (6 and 50, respectively) and haplotypes (6 and 6, respectively). The COI haplotype network (Supplementary Material 4) shows nine haplogroups that are separated by a large number of mutational steps and that correspond to the lineages recovered in the COI Bayesian phylogeny (see below).

Species Delimitation. The GMYC model delimited nine GMYC species, which were grouped in three main clades containing three GMYC species each (Supplementary Material 5). One of these clades was composed of a species with samples from the PCP-Ch and YP provinces and Nicaragua (*Stenocorse* sp. 1); a second species with samples from Nicaragua, the CP and TMVB provinces, Costa Rica and Texas (*Stenocorse* sp. 2); and a third species with the singleton from Argentina (*Stenocorse* sp. 3). The second main clade had one species with two specimens from the PCP-SMS (Jalisco) (*Stenocorse* sp. 4), a second with samples from Brazil and Colombia (*Stenocorse* sp. 5), and a third represented by a singleton from the CP in Querétaro, central Mexico (*Stenocorse* sp. 6). The third main clade had two species with samples from the GCP and PCP-SMS provinces (*Stenocorse* spp. 7 and 8), respectively, and a third represented by the two specimens from Venezuela (*Stenocorse* sp. 9).

The 2% divergence criterion applied to the uncorrected COI distances (Supplementary Material 6) discriminated the same nine molecular taxonomic units (MOTUs). However, some of specimens belonging to *Stenocorse* sp. 2 from the TMVB and CP provinces, Texas and Central America (Nicaragua and Costa Rica) had pairwise distances of 2.1%, though the mean values between these regions were lower than 2%.

Phylogenetic analysis. The Bayesian phylogram derived from the COI data set recovered two main clades that were geographically structured (Supplementary Material 7). One of these clades (PP= 0.65) was composed of two subclades with the samples from the GCP and PCP-SMS, respectively. The remaining main clade (PP= 0.69) had the specimens from Venezuela as sister to the other two subclades, one with specimens from Brazil, Colombia and one sample from the CP province, and the other one with samples from Argentina, Central America (Nicaragua and Costa Rica), central and southern Mexico (CP, TMVB, PCP-Ch, YP provinces) and Texas.

Divergence times. The divergence time estimates derived from the relaxed molecular clock analysis (Fig. 1) showed that *Stenocorse* originated during the mid to late Miocene, (15.1–8.14 Mya; mean = 11.23 Mya). The split between the two main clades in the genus was estimated to occur during the late Miocene to mid Pliocene, (6.84–3.96 Mya; mean = 5.29 Mya), whereas diversification of the various species occurred during the early Pliocene to late Pleistocene (5.71–0.72 Mya). The clade containing the lineage from the PCP-SMS province diverged from the lineage from the GCP province during the mid-Pliocene to late Pleistocene, (3.82-1.58 Mya; mean= 2.66 Mya). The subclade with the samples from the TMVB on the other hand diverged from the ones from the CP province between 0.8 and 0.28 Mya (mean = 0.54 Mya).

### *3RAD data*

Genetic structure. The five matrixes that were generated yielded between 434 and 7059 loci (Table 1). Sequence rate of similarity converged at a value of 0.90 after using Ilut *et al.*'s (2014) approach for optimizing the clustering threshold. The M1 and M3 matrices yielded the highest number of loci and informative sites with 7059 and 6368 and 30301 and 28507, respectively.

We recovered the highest average likelihood value [LnP(D)] for K= 4 in the STRUCTURE analysis (Fig. 2B). The specimens from the CP, TMVB and PCP-Ch provinces, Texas, Nicaragua and Costa Rica formed one genetic cluster, whereas the specimens from



Colombia and two from the CP formed a second one (Fig. 2A). The remaining two clusters on the other hand were each represented by the specimens from the GCP and PCP-SMS provinces, respectively. The GCP cluster showed a considerable admixture with samples from the PCP-SMS province, and in a lower degree with samples from the TMVB and PCP-Ch provinces, Nicaragua and Colombia (Fig. 2A, B).

Species delimitation. The species delineation analysis carried out with the BPP approach strongly supports an eight-species model (Supplementary Material 8). The recovered BPP species are highly congruent with the species delineation analysis performed with the GMYC model using the COI data set, except for the terminal taxa missing in the 3RAD analyses and the samples from the CP and TMVB provinces, Central America and Texas, which were each regarded as separate BPP species with high posterior probability support but were fused as a single GMYC species.

Phylogenetic analyses. The topologies derived from all matrices using SVDQuartets and Exabayes (Supplementary Materials 9 and 10) were highly congruent in most of their relationships except for those involving the specimen from Texas in the analyses with the M3 and M4 matrices, which were distinct but with low support, and the placement of two samples from the CP province, Queretaro, which were nested within a clade with the specimens from Colombia in some of the Exabayes analysis (M1, M2.5 and M3).

The topologies derived from the SVDQuartets and Exabayes analyses with the M1 matrix had the highest support values and obtained the most congruent relationships with respect to the COI analyses (Figs 2B and 3). These analyses recovered the same two major clades with significant support in all their relationships. One of the specimens from the PCP-Ch that was not sequenced for COI was recovered within the GCP clade.

We reared 18 different bruchine species belonging to six genera, which emerged from seeds of six different Fabaceae genera (Supplementary Material 1). The 3RAD Bayesian phylogram obtained with Exabayes shows our delimited species of *Stenocorse* together with their associated bruchine species and host plant genera (Fig. 3). Our rearing records showed that up to nine bruchine species from up to five genera were reared together with each species of

*Stenocorse*. We did not observe an evident association among *Stenocorse* bruchine species and their host plants.

## **Discussion**

### *Congruence between mt and nuclear 3RAD data*

In this study, we successfully employed both mt and genomic nuclear data to assess the species boundaries among populations of the parasitoid wasp genus *Stenocorse*. Our analyses revealed a high congruence between the above markers, consistently showing that the sampled populations actually consist of various species, with *S. bruchivora* having a marked genetic structure and a wide geographic distribution. Our gathered information also helped to reconstruct a strongly supported phylogeny for *Stenocorse*, providing a framework to study the association of its species to their bruchine hosts and the plants where they occur.

The COI data set proved to be highly useful in recovering the evolutionary lineages present in our examined samples. The use of mtDNA sequence data has been widely employed in systematics for almost three decades, since it provides phylogenetic information that is frequently concordant with phenotypic and nuclear data (Dedeine *et al.*, 2016; Suchan *et al.*, 2017; Toews *et al.*, 2012). However, the reconstruction of genealogical relationships using this locus alone may differ from species phylogenies due to incomplete lineage sorting, horizontal transfer and/or gene duplication (Avice, 2000). Combining mitochondrial and nuclear markers has therefore been shown to improve phylogenetic studies at different taxonomic levels, including at intraspecific level or among recently diverged taxa (Funk *et al.*, 2003; Toews *et al.*, 2012).

The nuclear 3RAD data set recovered an almost fully supported topology and species boundaries that were largely congruent with the COI analyses, even in the presence of a high percentage of missing data. Similarly to previous studies (Eaton *et al.*, 2017; Tripp *et al.*, 2017), our 3RAD phylogenetic analyses recovered robust phylogenies, even using matrices with up to 90% of missing data. The 3RAD and mt phylogenies showed congruent geographic structure,

particularly in the lineages from Mexico and Central America. This congruence between the two examined markers provide robustness to the proposed number of species diversification events, as well as to the inferences that we make about the evolutionary history of the genus.

#### *Genetic structure and admixture*

We investigated the patterns of nuclear genetic structure and mt haplotype diversity in *Stenocorse* to assess whether this would confirm the geographical clusters recovered by the phylogenetic analyses. The STRUCTURE analysis with 3RAD recovered four major genetic clusters that are partially congruent with the geographical distribution of the samples involved and with the clades obtained in our estimates of phylogeny.

Two of the genetic clusters recovered by STRUCTURE were congruent with the clades comprising the samples from the GCP and PCP-SMS provinces, which were delimited as *Stenocorse* spp. 7 and 8, respectively. The latter split appears to have been influenced by climatic oscillations that elevated the isometric sea levels during the mid Pliocene to Pleistocene (Haq *et al.*, 1987). This, together with the presence of high elevation areas in the SMS province, separated the Mexican Pacific and Gulf coastal plains along southeast Mexico (Ferrusquía-Villafranca, 1998). However, we found admixture between these two clusters, which could be explained by incomplete lineage sorting due to recent divergence, historical introgression or low levels of gene flow. The existence of limited gene flow and/or historical introgression could exist between these two taxa along the Isthmus of Tehuantepec in Oaxaca, since this is a narrow continental area with a central hill that does not exceed the 200 m.a.s.l. (Ferrusquía-Villafranca, 1998). A similar species diversification pattern was also observed in species of the grasshopper genus *Sphenarium* (Orthoptera: Phrygomonidae) from the PCP-SMS and GCP provinces (Pedraza-Lara *et al.*, 2015).

A third genetic cluster fused the members of two clades consistently recovered in our phylogenies, one restricted to Central America (*Stenocorse* sp. 1) and a second one, delimited as *Stenocorse* sp. 2 and assigned here to *S. bruchivora* (see below), widely distributed from southern USA to Costa Rica. Specimens assigned to the latter species were divided into various

putative species in the BPP analysis. This could be because the BPP algorithm often detects fine population structure and not species boundaries (Sukumaran *et al.*, 2017), and in our study this probably occurred due to the marked geographical structure observed in this taxon.

The samples from Colombia in northern South America and one from the CP province in Queretaro, Mexico, comprised the remaining cluster. The species delimitation analyses carried out both with COI and 3RAD data consistently showed that the specimens from the CP province in Central Mexico represent a separate evolutionary lineage from the ones from South America.

### *Species delimitation*

One of the main difficulties in delimiting insect species is the absence of a clear morphological differentiation (Cruaud *et al.*, 2018). Here we explored the species boundaries in the morphologically conserved genus *Stenocorse* based on mt and nuclear 3RAD data, which were analysed using three species delimitation methods with different properties.

The 2% genetic divergence criterion, which is widely employed for the Barcoding locus, faces the problem of determining a particular threshold of DNA differentiation above which a new species can be delimited (Hajibabaei *et al.*, 2006; Schwarzfeld *et al.*, 2015). This 2% cut-off value, however, has been successfully used to delineate species for a number of parasitic hymenopteran groups (*e.g.* Fernández-Flores *et al.*, 2013; Stahlhut *et al.*, 2013). The GMYC model, on the other hand, has proved to be a powerful tool for species delimitation using the barcoding locus, though it has been shown that it might overestimate species boundaries under some circumstances (Talavera *et al.*, 2013).

The multispecies coalescent model has shown to be robust in the presence of low levels of gene flow and to work well with a high number of informative loci (Hime *et al.*, 2016; Luo *et al.*, 2018; Zhang *et al.*, 2011). It represents a powerful tool for species delimitation of understudied groups since it is relatively robust in the face of unbalanced sampling, mutation rate heterogeneity and unequal population size (Barley *et al.*, 2018; Luo *et al.*, 2018). However,

it has been suggested that it does not distinguish species limits but genetic structure, and thus could also overestimate species diversification events (Sukumaran *et al.*, 2017).

Based on the observed genetic structure and the general congruence between the three species delineation analyses performed for the mt and nuclear 3RAD data, a total of six species were consistently delimited. Moreover, three additional species whose specimens could not be sequenced for 3RAD were delimited by the two approaches carried out for the COI data set.

The species delimited by the mt and nuclear data sets are as follows:

(1) *Stenocorse* sp. 1: species with records from the PCP-Ch subprovince along the coastal plain of Chiapas, Mexico, as well as Nicaragua. According to the COI analysis, the samples from the YP, Quintana Roo, belong to this species though they could not be included in the 3RAD data set.

(2) *Stenocorse* sp. 2 (*S. bruchivora*): this is a widely distributed species represented in our study by specimens from the TMVB and CP provinces, Costa Rica, Nicaragua and Texas, USA. The locality from Texas is near to the type locality of the only described species of the genus, *S. bruchivora*. Therefore, based on this and on examination of the morphological features displayed by holotype specimen and our gathered material, we assign *Stenocorse* sp. 2 to *S. bruchivora*.

The 2% COI divergence criterion and the BPP analyses revealed a marked geographical structure in the sampled populations of this species, suggesting the existence of several lineages. However, the genetic divergence supporting the split of these populations is inconclusive and the BPP analysis may be delimiting lineages of populations instead of species (Sukumaran *et al.*, 2017). Thus, we consider them as a single evolutionary unit pending further studies to confirm their taxonomic status.

(3) *Stenocorse* sp. 5: this species is composed of specimens from the department of Cundinamarca, central Colombia, and according to the COI analysis it also comprises the samples from the state of Minas Gerais, southeast Brazil.

(4) *Stenocorse* sp. 6: species represented by two specimens from the CP province in Queretaro, central Mexico.

(5) *Stenocorse* sp. 7: species occurring in the GCP province in the coast plain of Veracruz, Mexico, and by one specimen from the PCP-Ch subprovince in the coastal plain of Chiapas.

(6) *Stenocorse* sp. 8: species from the PCP subprovince along the coastal plains of Jalisco, Guerrero and Oaxaca, Mexico.

We also delimited the following three additional species based on samples that only had the COI information:

(1) *Stenocorse* sp. 3: represented by a singleton collected in the province of Cordoba, central Argentina.

(2) *Stenocorse* sp. 4: represented by two specimens collected in the PCP-SMS province in Jalisco, Mexico.

(3) *Stenocorse* sp. 9: species occurring in the state of Lara, northwest Venezuela in northern South America.

We found three cases of sympatry, one involving a pair of species co-occurring in the CP province in Querétaro, Mexico (*S. bruchivora* and *Stenocorse* sp. 6), a second with two species co-occurring in the GCP province in Veracruz (*Stenocorse* spp. 1 and 7) and a third with two species collected in the locality from Nicaragua in Central America. These cases of sympatry confirm the presence of intrinsic barriers to gene flow for the taxa involved. We recognise that our taxon sampling was limited, and thus various additional species of *Stenocorse* probably remain to be discovered, particularly in South America. Further collecting efforts are therefore necessary to know the actual species richness of this genus.

#### *Evolutionary relationships and biogeographic inferences*

Our phylogenetic analyses carried out with the mt and nuclear 3RAD data sets recovered topologies with a marked geographic structure, particularly for the clades composed by specimens whose populations are distributed along the Mexican territory, where most of our taxon sampling was focused. An undetermined number of *Stenocorse* species still remain to be

discovered; however, here we present some preliminary inferences that can be drawn based on the relationships obtained in this study.

The split of the two main *Stenocorse* clades recovered in our phylogenetic analyses could have occurred during the late Miocene to Pliocene, 3.96–6.84 Mya. This time period is subsequent to the connection between South and Central America, which was proposed to occur about 9 Mya ago (Montes *et al.*, 2012). The above two clades contain members both from north and south of the Isthmus of Panama, suggesting that the South and Central America connection could have promoted dispersal events along these two regions.

The complex geological history of the Mexican territory has led to its great altitude and climatic variation (Ferrusquía-Villafranca, 1998). This intricate orography is composed of various mountain ranges of distinct origins, which, together with the paleoclimatic events that occurred in the region, have shaped the evolutionary history and species diversification of many groups of organisms (*e.g.* Mastretta-Yanes *et al.*, 2015; Maldonado-Sánchez *et al.*, 2016; Rodríguez-Gómez *et al.*, 2013).

Two main clades were recovered in our phylogenetic analyses for the specimens assigned to *S. bruchivora*, one with samples collected in high altitude localities along the CP and MTBV provinces in central Mexico, and a second composed of samples obtained in low altitude areas from Texas, USA, and Central America. This evidence suggests the probable presence of genetic structure between lowland and highland areas in *S. bruchivora*, which was also found in the mainly Mexican endemic grasshopper genus *Sphenarium* (Pedraza-Lara *et al.*, 2015).

The estimated time for the split between the populations assigned here to *S. bruchivora* from the CP and MTBV provinces (0.8-0.3 Mya) is after the completion of the latter mountain range along central Mexico, which occurred during the early Pleistocene, 1.8 Mya (Ferrari *et al.*, 2012). Thus, this geological event could have caused the recent separation of these two high altitude lineages. Our phylogenetic and molecular clock analyses also show a relatively recent separation between the species of *Stenocorse* from the GCP and PCP-SMS provinces during the mid Pliocene to late Pleistocene. This period was characterised by having a series of climatic

oscillations, which caused significant variation in the isometric sea levels through that time (Haq *et al.*, 1987). Further studies are needed to assess in detail whether these climatic fluctuations caused speciation between the above two species.

#### *Host and plant association*

It has been suggested that one of the main factors that has triggered the extraordinary species diversification in hymenopteran parasitoids is their host specificity (Hall *et al.*, 2017; Stireman *et al.*, 2006). In the case of parasitoids of phytophagous insects, speciation is often influenced in a cascading way by specialisation on hosts that feed on a particular group of plants (Abrahamson & Blair, 2008; Yu *et al.*, 2018). Previous molecular studies have revealed that some groups of parasitoid wasps previously considered as generalists of herbivore insects actually represent various morphologically conserved species with a specialized host preference (Hambäck *et al.*, 2013; Smith *et al.*, 2008, 2006).

It has been shown that plant-specificity of the host or its shift from one plant or plant organ to another can promote speciation of the associated parasitoids (Deng *et al.*, 2013; Futuyma *et al.*, 1988). However, due to the complexity of their life histories, the response to ecological shifts of the hosts could vary considerably (Cronin *et al.*, 2001; Yu *et al.*, 2018). *Stenocorse bruchivora* has been reported to attack larvae of a number of bruchine beetle species (López-Martínez *et al.*, 2003). However, a number of these records belong to undescribed species that were erroneously assigned to it. Most of these bruchid species are known to have host specialisation that has been largely attributed to herbivore-plant interactions that promote adaptation to toxic compounds present in the plant seeds (Jermy & Szentesi, 2003; Romero-Nápoles *et al.*, 2002; Tuda *et al.*, 2006).

Direct host association could not be determined with our rearing, since different bruchine species emerged from the same pod and even the same seed from which the *Stenocorse* specimens were obtained. Taking this into account, our phylogenetic analyses and rearing records did not suggest host specialisation, as various bruchid species from up to five genera were found to be associated to each parasitoid species. The apparent association



observed between some species of *Stenocorse* and their bruchine and host plant species (*e.g.* for species from the PCP-SMS and GCP provinces) could be due to collecting bias, as only one species of legume was obtained in these localities. Thus, opposite to what should be expected due to the frequent host-plant specialization of bruchids, species of *Stenocorse* might possess generalist-feeding habits.

Our findings reinforce the commonly accepted hypothesis that ectoparasitoids tend to have more generalist feeding habits than endoparasitoids, which can be explained by a lack of selection over traits associated with the coping of the wasp's larva with the immune system of the host since it does not affect the overall fitness of the wasp (Quicke, 2015). Further studies using techniques that help to confirm host-ectoparasitoid associations are therefore needed in order to confirm the wide host range in species of *Stenocorse*.

## **Acknowledgments**

We thank Jean Michel Maes, Paul E. Hanson, Manuel Elías-Guitérrez and Juan José Martínez for their donation/loan of specimens; Mariana Camacho, Valeria Castro and Ana María Sanabria for their help during the collecting trips in Colombia; Cristina Mayorga and Guillermina Ortega for their help at the Colección Nacional de Insectos (CNIN), IB UNAM; Laura Márquez for her help with Sanger sequencing; and Travis Glenn for hosting NDM for a research stay at the UGA EHS DNA Lab. We also thank the Dirección General de Cómputo y de Tecnologías de Información y Comunicación (DGTIC, UNAM) for the services and resources provided to run the 3RAD analyses in the Miztli supercomputer (Proyecto de investigación regular código: LANCAD-UNAM-DGTIC-339). The specimens were collected in Nicaragua under the permit given by the Ministerio de Ambiente y Los Recursos Naturales (DGNP/DB-004-2016) to AZR. This study was supported by grants given by the PAPIIT-DGAPA-UNAM (Convocatoria PAPIIT 2016, Proyecto IN207016) to AZR; a grant given by the Pontificia Universidad Javeriana (Project 7210 “Estudios macro y microevolutivos en la familia Braconidae) to AA-S; and grants given by the National Science Foundation (grant no.

DEB-02-71929 to Dr. Charles Fox from University of Kentucky) and a grant given by the Fundación para la Promoción de la Investigación y la Tecnología, Banco de la República, Colombia (Grant nos 1568 and 620), to AA-S. This study is part of NDM's MSc thesis, who received a MSc scholarship given by CONACyT. NDM also thanks the Posgrado en Ciencias Biológicas, UNAM, for its support during her studies.

## References

- Aberer, A. J., Kobert, K. & Stamatakis, A. (2014). ExaBayes: Massively Parallel Bayesian Tree Inference for the Whole-Genome Era. *Molecular Biology and Evolution*, **31**, 2553–2556.
- Abrahamson, W. G. & Blair, C. P. (2008). Sequential radiation through host-plant formation: herbivore diversity leads to diversity in natural enemies. In *Specialization, Speciation, and Radiation: the Evolutionary Biology of Herbivorous Insects* (pp. 188–202).
- Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G. & Hohenlohe, P. A. (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics*, **17**, 81–92.
- Avice, J. C. (2000). *Phylogeography: The History and Formation of Species*. Cambridge, MA: Harvard University Press.
- Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., ... Johnson, E. A. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE*, **3**, 1–7.
- Barley, A. J., Brown, J. M. & Thomson, R. C. (2018). Impact of Model Violations on the Inference of Species Boundaries under the Multispecies Coalescent. *Systematic Biology*, **67**, 269–284.
- Bathey, C. J. & Klicka, J. (2017). Cryptic speciation and gene flow in a migratory songbird Species Complex: Insights from the Red-Eyed Vireo (*Vireo olivaceus*). *Molecular Phylogenetics and Evolution*, **113**, 67–75.

- Carstens, B. C., Pelletier, T. A., Reid, N. M. & Satler, J. D. (2013). How to fail at species delimitation. *Molecular Ecology*, **22**, 4369–4383.
- Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A. & Cresko, W. A. (2013). Stacks: An analysis tool set for population genomics. *Molecular Ecology*, **22**.
- Che, Y., Gui, S., Lo, N., Ritchie, A. & Wang, Z. (2017). Species delimitation and phylogenetic relationships in ectobiid cockroaches (Dictyoptera, Blattodea) from China. *PLoS ONE*, **12**.
- Chifman, J. & Kubatko, L. (2014). Quartet inference from SNP data under the coalescent model. *Bioinformatics*, **30**, 3317–3324.
- Chifman, J. & Kubatko, L. (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.
- Crawford, J. C. (1910). New parasitic Hymenoptera. *Proceedings of the Entomological Society of Washington.*, **11**, 203–207.
- Cronin, J. T. & Abrahamson, W. G. (2001). Do parasitoids diversify in response to host-plant shifts by herbivorous insects? *Ecological Entomology*, **26**, 347–355.
- Cruaud, A., Groussier, G., Genson, G., Sauné, L. & Rasplus, J.-Y. (2018). Pushing the limits of whole genome amplification : Successful sequencing of RADseq libraries from single micro- hymenoptera ( Chalcidoidea , Trichogramma ). *PeerJ*, **6** : e5640, 1–30.
- De Queiroz, K. (2007). Species concepts and species delimitation. *Systematic Biology*, **56**, 879–886.
- Dedeine, F., Dupont, S., Guyot, S., Matsuura, K., Wang, C., Habibpour, B., ... Luchetti, A. (2016). Historical biogeography of Reticulitermes termites (Isoptera: Rhinotermitidae) inferred from analyses of mitochondrial and nuclear loci. *Molecular Phylogenetics and Evolution*, **94**, 778–790.
- Deng, J., Yu, F., Li, H. Bin, Gebiola, M., Desdevises, Y., Wu, S. A. & Zhang, Y. Z. (2013). Cophylogenetic relationships between Anicetus parasitoids (Hymenoptera: Encyrtidae) and their scale insect hosts (Hemiptera: Coccidae). *BMC Evolutionary Biology*, **13**, 1–11.
- Development Core team, R. (2008). R: A language and environment for statistical computing.

- Vienna, Austria: R Foundation for Statistical Computing.
- Drummond, A. J., Nicholls, G. K., Rodrigo, A. G. & Solomon, W. (2002). Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. *Genetics*, **161**, 1307–1320.
- Eaton, D. A. R. (2014). PyRAD: Assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics*, **30**, 1844–1849.
- Eaton, D. A. R., Spriggs, E. L., Park, B. & Donoghue, M. J. (2017). Misconceptions on missing data in RAD-seq phylogenetics with a deep-scale example from flowering plants. *Systematic Biology*, **66**, 399–412.
- Eyer, P. A., Seltzer, R., Reiner-Brodetzki, T. & Hefetz, A. (2017). An integrative approach to untangling species delimitation in the *Cataglyphis bicolor* desert ant complex in Israel. *Molecular Phylogenetics and Evolution*, **115**, 128–139.
- Fernández-Flores, S., Fernández-Triana, J. L., Martínez, J. J. & Zaldívar-Riverón, A. (2013). DNA barcoding species inventory of Microgastrinae wasps (Hymenoptera, Braconidae) from a Mexican tropical dry forest. *Molecular Ecology Resources*, **13**, 1146–1150.
- Ferrari, L., Orozco-Esquivel, T., Manea, V. & Manea, M. (2012). The dynamic history of the Trans-Mexican Volcanic Belt and the Mexico subduction zone. *Tectonophysics*, **522–523**, 122–149.
- Ferrusquía-Villafranca, I. (1998). Geología de México: Una sinopsis. In T. P. Ramamoorthy, R. Bye, A. Lot, & J. Fa (Eds.), *Diversidad biológica de México: orígenes y distribución* (pp. 3–108). Distrito Federal, Mexico: Instituto de Biología, Universidad Nacional Autónoma de México, México.
- Folmer, O., Black, M., Hoeh, W., Lutz, R. & Vrijenhoek, R. (1994). DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*, **3**, 294–299.
- Forbes, A. A., Bagley, R. K., Beer, M. A., Hippee, A. C. & Widmayer, H. A. (2018). Quantifying the unquantifiable: why Hymenoptera — not Coleoptera — is the most speciose animal order. *bioRxiv*, 274431.

- Funk, D. J. & Omland, K. E. (2003). Species-Level Paraphyly and Polyphyly: Frequency, Causes, and Consequences, with Insights from Animal Mitochondrial DNA. *Annual Review of Ecology, Evolution, and Systematics*, **34**, 397–423.
- Futuyma, D. J. & Moreno, G. (1988). The Evolution of Ecological Specialization. *Annual Review of Ecology and Systematics*, **19**, 207–233.
- Glenn, T. C., Bayona-Vásquez, N. J., Kieran, T. J., Pierson, T. W., Hoffberg, S. L., Scott, P. A., ... Faircloth, B. C. (2017). Adapterama III: Quadruple-indexed, triple-enzyme RADseq libraries for about \$1USD per Sample (3RAD). *BioRxiv*, 1–34.
- Goodwin, S., McPherson, J. D. & McCombie, W. R. (2016). Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics*, **17**, 333–351.
- Hajibabaei, M., Janzen, D. H., Burns, J. M., Hallwachs, W. & Hebert, P. D. N. (2006). DNA barcodes distinguish species of tropical Lepidoptera. *Proceedings of the National Academy of Sciences*, **103**, 968–971.
- Hall, A. A. G., Steinbauer, M. J., Taylor, G. S., Johnson, S. N., Cook, J. M. & Riegler, M. (2017). Unravelling mummies: cryptic diversity, host specificity, trophic and coevolutionary interactions in psyllid - parasitoid food webs. *BMC Evolutionary Biology*, **17**, 1–15.
- Hambäck, P. A., Weingartner, E., Ericson, L., Fors, L., Cassel-Lundhagen, A., Stenberg, J. A. & Bergsten, J. (2013). Bayesian species delimitation reveals generalist and specialist parasitic wasps on Galerucella beetles (Chrysomelidae): Sorting by herbivore or plant host. *BMC Evolutionary Biology*, **13**, 1–14.
- Haq, B., Hardenbohl, J. & Vail, P. R. (1987). Chronology of fluctuating sea levels since the Triassic (250 million years ago to present). *Sciences-New York*, **23**, 1156–1167.
- Hebert, P. D. N., Cywinska, A., Ball, S. L. & DeWaard, J. R. (2003a). Biological identifications through DNA barcodes. *Proceedings of the Royal Society B: Biological Sciences*, **270**, 313–321.
- Hebert, P. D. N., Ratnasingham, S. & de Waard, J. R. (2003b). Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of*

*the Royal Society B: Biological Sciences*, **270**, S96–S99.

- Hime, P. M., Hotaling, S., Grewelle, R. E., O’Neill, E. M., Voss, S. R., Shaffer, H. B. & Weisrock, D. W. (2016). The influence of locus number and information content on species delimitation: an empirical test case in an endangered Mexican salamander. *Molecular Ecology*, **25**, 5959–5974.
- Ilut, D. C., Nydam, M. L. & Hare, M. P. (2014). Defining loci in restriction-based reduced representation genomic data from nonmodel species: Sources of bias and diagnostics for optimal clustering. *BioMed Research International*, **2014**, 1–9.
- Jermý, T. & Szentesi, Á. (2003). Evolutionary aspects of host plant specialisation—a study on bruchids (Coleoptera: Bruchidae). *Oikos*, **1**, 196–204.
- Johnson, C.D. (1983). Ecosystematics of *Acanthoscelides* (Coleoptera: Bruchidae) of Southern Mexico and Central America. *Miscellaneous Publications of the Entomological Society of America*, **56**, 1-370.
- Johnson, C.D. (1990). Systematics of the seed beetle genus *Acanthoscelides* (Bruchidae) of Northern South America. *Transactions of the American Entomological Society*, **116**, 297-618.
- Johnson, C. D., Kingsolver, J. M. & Teran, A. R. (1989). Sistemática del género *Stator* (Insecta: Coleoptera: Bruchidae) en Sudamérica. *Opera Lilloana*, **37**. Tucumán, República Argentina. 105 p.
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., ... Drummond, A. J. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, **28**, 1647–1649.
- Kingsolver, J. M. & Johnson, C. D. (1986). Systematics of the genus *Mimosestes* (Coleoptera: Bruchidae). *United States Department of Agriculture. Technical Bulletin*, **1590**, 160 p.
- Kingsolver, J. M. (1986). A taxonomic study of the genus *Algarobius* (Coleoptera: Bruchidae). *Entomography*, **4**, 109–136.

- Kingsolver, J. M. (1988). Biosystematics of the genus *Merobruchus* of continental North America and the West Indies (Coleoptera: Bruchidae). *United States Department of Agriculture. Technical Bulletin*, **1744**, 63 p.
- Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B., Bussonnier, M., Frederic, J., ... Willing, C. (2016). Jupyter Notebooks - a publishing format for reproducible computational workflows. *Positioning and Power in Academic Publishing: Players, Agents and Agendas*, 87–90.
- Korenko, S., Spasojevic, T., Pekar, S., Walter, G. H., Korenkova, V., Hamouzova, K., ... Klopstein, S. (2018). One generalist or several specialist species? Wide host range and diverse manipulations of the hosts' web-building behaviour in the true spider parasitoid *Zatypota kauros* (Hymenoptera: Ichneumonidae). *Insect Conservation and Diversity*, **11**, 587–599.
- Kumar, S., Stecher, G. & Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular Biology and Evolution*, **33**, 1870–1874.
- Lanfear, R., Calcott, B., Ho, S. Y. W. & Guindon, S. (2012). PartitionFinder: Combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution*, **29**, 1695–1701.
- Leigh, J. W. & Bryant, D. (2015). POPART: full-feature software for haplotype network construction. *Methods in Ecology and Evolution*, **6**, 1110–1116.
- López-Martínez, V., Figueroa-De La Rosa, J. I. & Romero-Nápoles, J. (2003). Registro de un nuevo huésped para *Stenocorse bruchivora* (Crawford) (Hymenoptera: Braconidae) con notas de su distribución en México. *Acta Zoologica Mexicana*, **89**, 287–289.
- Luo, A., Ling, C., Ho, S. Y. W. & Zhu, C. (2018). Comparison of Methods for Molecular Species Delimitation across a Range of Speciation Scenarios. *Systematic Biology*, **67**, 830–846.
- Maldonado-Sánchez, D., Gutiérrez-Rodríguez, C. & Ornelas, J. F. (2016). Genetic divergence in the common bush-tanager *Chlorospingus ophthalmicus* (Aves: Emberizidae) throughout

- Mexican cloud forests: The role of geography, ecology and Pleistocene climatic fluctuations. *Molecular Phylogenetics and Evolution*, **99**, 76–88.
- Marsh, P. M. (1968). The Nearctic Doryctinae, VI. The genera *Acrophasmus*, *Glyptocolastes*, *Doryctinus*, and a new genus, *Stenocorse* (Hymenoptera: Braconidae). *Proceedings of the Entomological Society of Washington*, **70**, 101–113.
- Marshall, C. J. & Liebherr, J. K. (2000). Cladistic biogeography of the Mexican transition zone. *Journal of Biogeography*, **27**, 203–216.
- Mastretta-Yanes, A., Moreno-Letelier, A., Piñero, D., Jorgensen, T. H. & Emerson, B. C. (2015). Biodiversity in the Mexican highlands and the interaction of geology, geography and climate within the Trans-Mexican Volcanic Belt. *Journal of Biogeography*, **42**, 1586–1600.
- Meyer, C. P. & Paulay, G. (2005). DNA barcoding: error rates based on comprehensive sampling. *PLoS Biology*, **3**, e422.
- Miller, M. A., Pfeiffer, W. & Schwartz, T. (2010). 2010 Gateway Computing Environments Workshop (GCE). *2010 Gateway Computing Environments Workshop (GCE)*, 1–8.
- Montes, C., Bayona, G., Cardona, A., Buchs, D. M., Silva, C. A., Morón, S., ... Valencia, V. (2012). Arc-continent collision and orocline formation: Closing of the Central American seaway. *Journal of Geophysical Research: Solid Earth*, **117**, 1–25.
- Naciri, Y. & Linder, H. P. (2015). Species delimitation and relationships: The dance of the seven veils. *Taxon*, **64**, 3–16.
- Papadopoulou, A., Anastasiou, I. & Vogler, A. P. (2010). Revisiting the insect mitochondrial molecular clock: The mid-aegean trench calibration. *Molecular Biology and Evolution*, **27**, 1659–1672.
- Pedraza-Lara, C., Barrientos-Lozano, L., Rocha-Sánchez, A. Y. & Zaldívar-Riverón, A. (2015). Montane and coastal species diversification in the economically important Mexican grasshopper genus *Sphenarium* (Orthoptera: Pyrgomorphidae). *Molecular Phylogenetics and Evolution*, **84**, 220–231.
- Pons, J., Barraclough, T. G., Gomez-Zurita, J., Cardoso, A., Duran, D. P., Hazell, S., ... Vogler,



- A. P. (2006). Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Systematic Biology*, **55**, 595–609.
- Pritchard, J. K., Stephens, M. & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.
- Quicke, D. L. J. (1997). *Parasitic wasps*. London, England.: Cambridge University Press.
- Quicke, D. L. J. (2015). *The Braconid and Ichneumonid Parasitoid Wasps*.
- Rambaut. (2009). Fig Tree. Retrieved from <http://tree.bio.ed.ac.uk>
- Rambaut, A. & Drummond, A. J. (2009). Tracer.
- Rannala, B. & Yang, Z. (2013). Improved reversible jump algorithms for Bayesian species delimitation. *Genetics*, **194**, 245–253.
- Rannala, B. & Yang, Z. (2017). Efficient Bayesian species tree inference under the multispecies coalescent. *Systematic Biology*, **66**, 823–842.
- Ratnasingham, S. & Hebert, P. D. N. (2013). A DNA-Based Registry for All Animal Species: The Barcode Index Number (BIN) System. *PLoS ONE*, **8**.
- Reaz, R., Bayzid, M. S. & Rahman, M. S. (2014). Accurate phylogenetic tree reconstruction from quartets: A heuristic approach. *PLoS ONE*, **9**.
- Rochette, N. C. & Catchen, J. M. (2017). Deriving genotypes from RAD-seq short-read data using Stacks. *Nature Protocols*, **12**, 2640–2659.
- Rodríguez-Gómez, F., Gutiérrez-Rodríguez, C. & Ornelas, J. F. (2013). Genetic, phenotypic and ecological divergence with gene flow at the isthmus of Tehuantepec: The case of the azure-crowned hummingbird (*Amazilia cyanocephala*). *Journal of Biogeography*, **40**, 1360–1373.
- Rohland, N. & Reich, D. (2012). Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Research*, **22**, 939–946.
- Romero, J. & Johnson, C. D. (2000). Revision of the genus *Zabrotes* Horn of Mexico (Coleoptera: Bruchidae: Amblycerinae). *Transactions of the American Entomological Society*, **126**, 221–274.
- Romero-Nápoles, J., Ayers, T. J. & Johnson, C. D. (2002). Cladistics, Bruchids and Host

- Plants : Evolutionary Interactions in *Amblycerus* ( Coleoptera : Bruchidae ). *Acta Zoologica Mexicana*, **86**, 1–16.
- Ronquist, F. & Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, **19**, 1572–1574.
- Rozas, J., Ferrer-MATA, A., Sánchez-Del Barrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E. & Sánchez-Gracia, A. 2017. DnaSP 6: DNA sequence polymorphism analysis of large datasets. *Molecular Biology and Evolution*, **34**, 3299–3302.
- Salas-Lizana, R. & Oono, R. (2018). Double-digest RADseq loci using standard Illumina indexes improve deep and shallow phylogenetic resolution of *Lophodermium*, a widespread fungal endophyte of pine needles. *Ecology and Evolution*, **8**, 6638–6651.
- Schwarzfeld, M. D. & Sperling, F. A. H. (2015). Comparison of five methods for delimitating species in *Ophion* Fabricius, a diverse genus of parasitoid wasps (Hymenoptera, Ichneumonidae). *Molecular Phylogenetics and Evolution*, **93**, 234–248.
- Shenefelt, R. D. & Marsh, P. M. (1976). Braconidae 9, Doryctinae. In J. van der Vecht & R. D. Shenefelt (Eds.), *Hymenopterum catalogus (nova edito)* (pp. 1263–1424). The Hague: Dr. W. Junk.
- Smith, M. A., Rodriguez, J. J., Whitfield, J. B., Deans, A. R., Janzen, D. H., Hallwachs, W. & Hebert, P. D. N. (2008). Extreme diversity of tropical parasitoid wasps exposed by iterative integration of natural history, DNA barcoding, morphology, and collections. *Proceedings of the National Academy of Sciences*, **105**, 12359–12364.
- Smith, M. A., Woodley, N. E., Janzen, D. H., Hallwachs, W. & Hebert, P. D. N. (2006). DNA barcodes reveal cryptic host-specificity within the presumed polyphagous members of a genus of parasitoid flies (Diptera: Tachinidae). *Proceedings of the National Academy of Sciences*, **103**, 3657–3662.
- Stahlhut, J. K., Fernández-Triana, J., Adamowicz, S. J., Buck, M., Goulet, H., Hebert, P. D. N., ... Smith, M. A. (2013). DNA barcoding reveals diversity of Hymenoptera and the dominance of parasitoids in a sub-arctic environment. *BMC Ecology*, **13**, 1–13.
- Stireman, J. O., Nason, J. D., Heard, S. B. & Seehawer, J. M. (2006). Cascading host-associated

- genetic differentiation in parasitoids of phytophagous insects. *Proceedings of the Royal Society B: Biological Sciences*, **273**, 523–530.
- Suchan, T., Espíndola, A., Rutschmann, S., Emerson, B. C., Gori, K., Dessimoz, C., ... Alvarez, N. (2017). Assessing the potential of RAD-sequencing to resolve phylogenetic relationships within species radiations: The fly genus *Chiastocheta* (Diptera: Anthomyiidae) as a case study. *Molecular Phylogenetics and Evolution*, **114**, 189–198.
- Sukumaran, J. & Knowles, L. L. (2017). Multispecies coalescent delimits structure, not species. *Proceedings of the National Academy of Sciences*, **114**, 1607–1612.
- Swofford, D. L. (2002). *PAUP\*: Phylogenetic Analysis Using Parsimony (⊞ and Other Methods)*. Sunderland, MA: Sinauer Associates.
- Talavera, G., Dincă, V. & Vila, R. (2013). Factors affecting species delimitations with the GMYC model: Insights from a butterfly survey. *Methods in Ecology and Evolution*, **4**, 1101–1110.
- Templeton, A. R., Grandall, K. A. & Sing, F. S. (1992). A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III Cladogram estimation. *Genetics*, **132**, 619–633.
- Toews, D. P. L. & Brelsford, A. (2012). The biogeography of mitochondrial and nuclear discordance in animals. *Molecular Ecology*, **21**, 3907–3930.
- Tripp, E. A., Tsai, Y. H. E., Zhuang, Y. & Dexter, K. G. (2017). RADseq dataset with 90% missing data fully resolves recent radiation of *Petalidium* (Acanthaceae) in the ultra-arid deserts of Namibia. *Ecology and Evolution*, **7**, 7920–7936.
- Tuda, M., Rönn, J., Buranapanichpan, S., Wasano, N. & Arnqvist, G. (2006). Evolutionary diversification of the bean beetle genus *Callosobruchus* (Coleoptera: Bruchidae): Traits associated with stored-product pest status. *Molecular Ecology*, **15**, 3541–3551.
- van Velzen, R., Weitschek, E., Felici, G. & Bakker, F. T. (2012). DNA barcoding of recently diverged species: Relative performance of matching methods. *PLoS ONE*, **7**.
- Veijalainen, A., Wahlberg, N., Broad, G. R., Erwin, T. L., Longino, J. T. & Sääksjärvi, I. E. (2012). Unprecedented ichneumonid parasitoid wasp diversity in tropical forests.

*Proceedings of the Royal Society B: Biological Sciences*, **279**, 4694–4698.

Wiemers, M. & Fiedler, K. (2007). Does the DNA barcoding gap exist? - A case study in blue butterflies (Lepidoptera: Lycaenidae). *Frontiers in Zoology*, **4**, 1–16.

Yang, Z. & Rannala, B. (2010). Bayesian species delimitation using multilocus sequence data. *PNAS*, **107**, 9264–9269.

Yu, H., Liang, D., Tian, E., Zheng, L. & Kjellberg, F. (2018). Plant geographic phenotypic variation drives diversification in its associated community of a phytophagous insect and its parasitoids. *BMC Evolutionary Biology*, **18**, 1–11.

Zhang, C., Zhang, D. X., Zhu, T. & Yang, Z. (2011). Evaluation of a bayesian coalescent method of species delimitation. *Systematic Biology*, **60**, 747–761.

Table 1: Summary statistics of the 3RAD matrices that were examined. Table shows the presence/absence of the out-group in the matrices, minimum sample per locus parameter, number of raw reads in each matrix, coverage range of reads, total loci after filtering, coverage range of final loci, number of variable sites (Var) and number of parsimony informative sites (Pis),

Matrix	Parameters		Reads			Loci		SNPs	
	Out Group	Min. Samp. Locus.	Raw reads	Reads cov. Range	mean num. Reads	Tot. Filt. Loci.	loci. Cov. Range	Var	Pis
M1	no	8	77141303	10605-3579617	952361.7654	7059	57-2786	43740	30301
M2	no	40	77141303	10605-3579617	952361.7654	563	4-487	5319	3804
M3	yes	9	81867492	10605-3579617	951947.5814	6368	57-2675	40759	28507
M4	yes	43	81867492	10605-3579617	951947.5814	434	5-413	4372	3132
M2.5	no	40	77129662	10605-3579617	964120.775	532	25-485	5247	3742

## Figure legends

**Figure 1.** Chronogram derived from the Bayesian relaxed molecular clock analysis conducted with the program BEAST using the COI data set. Black circles near nodes represent posterior probabilities  $\geq 0.95$ . Coloured vertical bars represent the species delineated by the GMYC and BPP methods. Horizontal bars correspond to the 95% HPD interval.

**Figure 2.** A) Map showing the geographic provenance of the specimens of *Stenocorse* examined in this study. Coloured circles correspond to the genetic clusters recovered with 3RAD. Coloured triangles represent species only delimited by COI data. B) Phylogeny recovered with the 3RAD data set using the SVDQuartets method and the M1 matrix. Black circles near nodes represent bootstrap values  $\geq 80\%$ . Coloured boxes on the tips of the tree represent the BPP species delimitation; coloured boxes on the right of the tips correspond to the genetic clusters obtained with STRUCTURE (K=4).

**Figure 3.** Bayesian phylogram of *Stenocorse* derived from the 3RAD matrix M1 using the program Exabayes. Black circles correspond to posterior probability values  $< 95\%$ . Coloured boxes show the species delimited with 3RAD data. Bruchid species that were reared together with the specimens of *Stenocorse* are included and colour-coded on their right along with their host plant species.

## Supplementary Material

**Supplementary Material 1.** List with specimens examined in this study, their localities, physiographic provinces of provenance, DNA voucher and GeneBank accession numbers and bruchid host species and host plant species.

**Supplementary Material 2.** Primer and adapter combinations used for library preparation.

**Supplementary Material 3.** Measures of COI DNA diversity observed for population groups of *Stenocorse*.

**Supplementary Material 4.** COI haplotype network of the examined specimens of *Stenocorse*. Circle size is proportional to the number of specimens with that haplotype. Haplotypes are grouped according to the species delimited in this study.

**Supplementary Material 5.** Ultrametric tree obtained with BEAST showing the species delimited by the GMYC method. Clades in red represent a single specie while black tips represent singletons.

**Supplementary Material 6.** Uncorrected COI distances for the specimens of *Stenocorse* examined in this study. Specimens were grouped according to the clades recovered in the phylogenetic reconstructions (range above diagonal; mean below diagonal).

**Supplementary Material 7.** Bayesian phylogram obtained with MrBayes and the COI data set. Black circles near branches represent posterior probability values  $\geq 0.95$ . Main clades are coloured.

**Supplementary Material 8.** Guide tree showing the eight potential species evaluated with the BPP method. Range of posterior probabilities of each repetition is shown near nodes. PCP-SMS= Pacific Costal Plains subprovince, Sierra Madre del sur province; NCA= Nuclear Central America (Chiapas and Nicaragua); Central America = Costa Rica and Nicaragua; TMVB= Mexican Transvolcanic Belt, Mexico (Puebla, Morelos, Estado de Mexico); CP= Central Plateau, Mexico (Querétaro.); CP1=Central Plateau, Mexico (Querétaro; voucher nos CNIN3461, CNIN3468).

**Supplementary Material 9.** Phylogenetic tree of *Stenocorse* derived from the Exabayes analyses with all data matrices. Black circles near nodes represent posterior probability values < 95%.

**Supplementary Material 10.** Bayesian phylograms of *Stenocorse* derived from the SVDQuartets analyses carried out with all data matrices. Black circles near nodes correspond to posterior probability values  $\geq 80\%$ .



Figure 1.

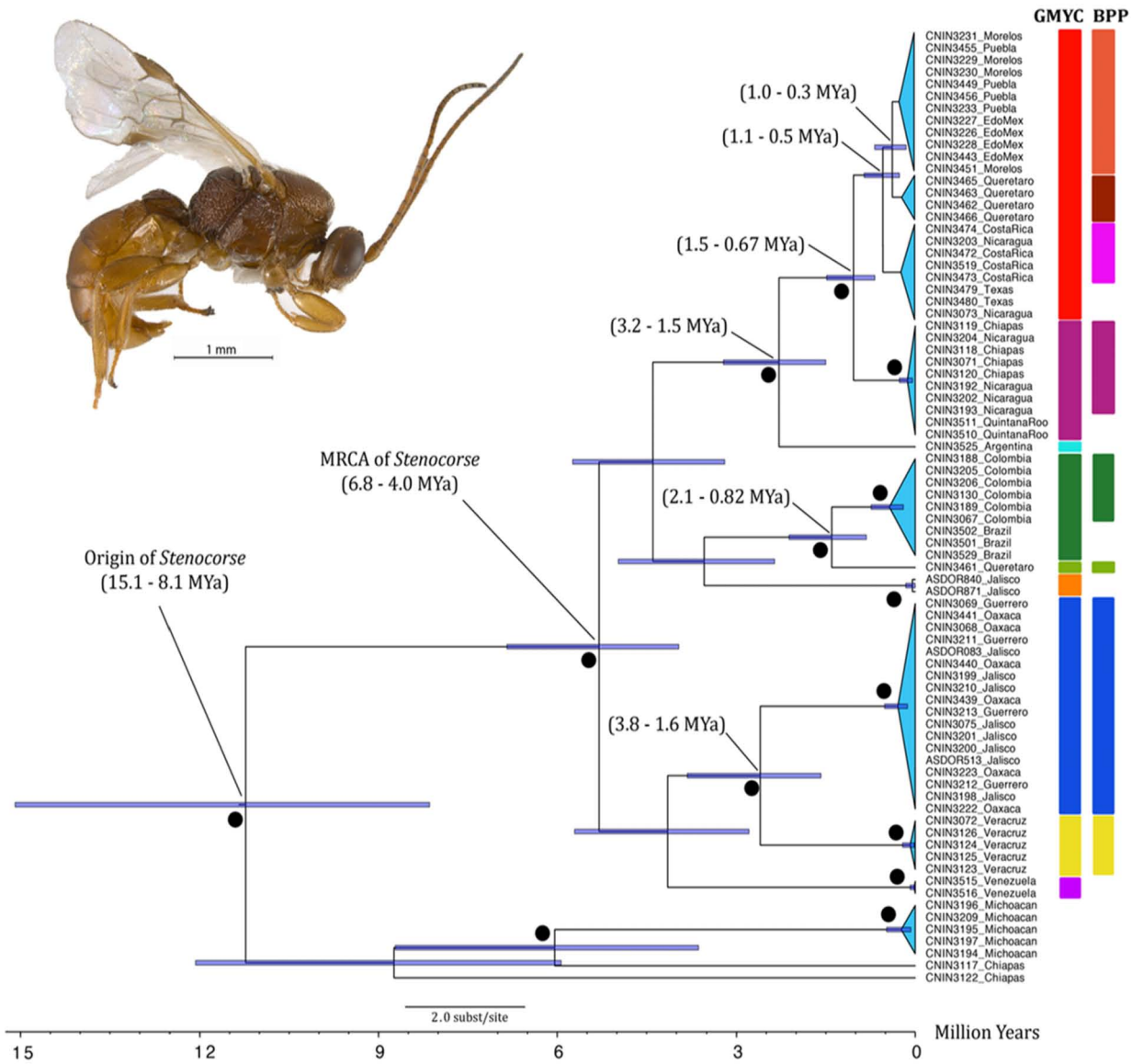


Figure 2.

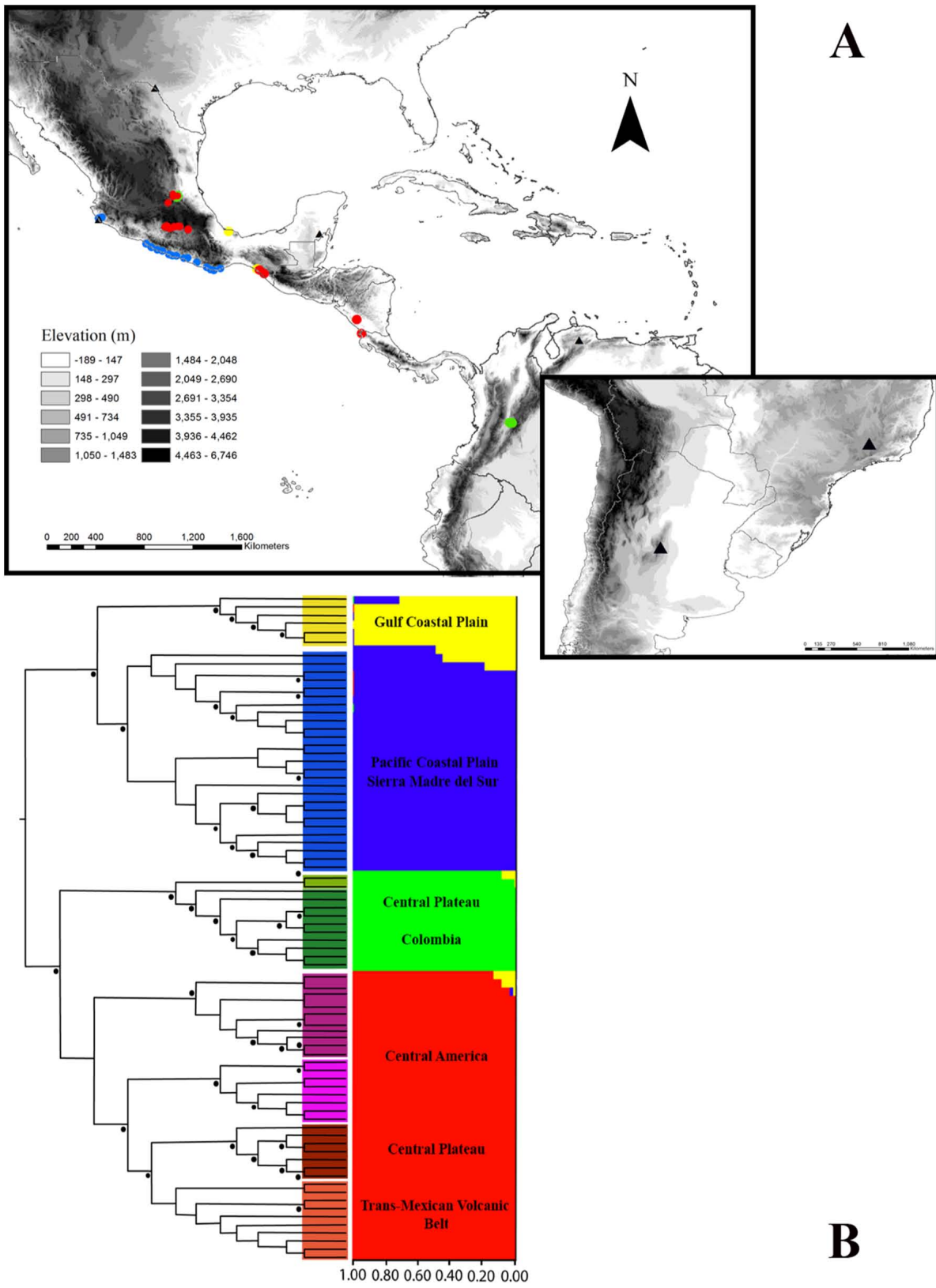
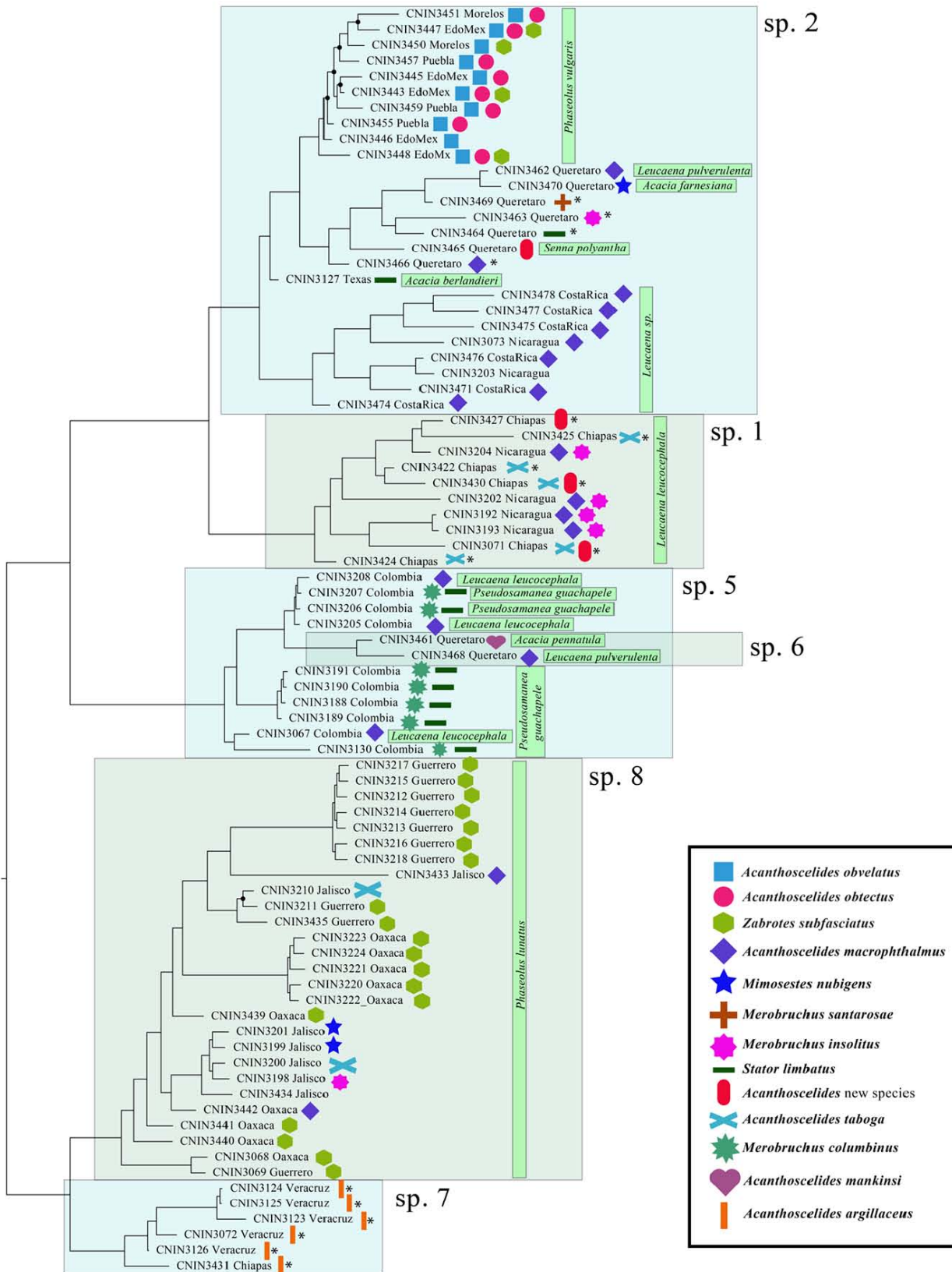


Figure 3.



## DISCUSIÓN

### Tipos de datos y su desempeño

Las avispas parasitoides del género *Stenocorse* son insectos pequeños y poco conspicuos. No obstante presentan interacciones complejas con sus huéspedes herbívoros y afectan indirectamente el desarrollo de las plantas de las que se alimentan sus huéspedes. En el presente trabajo se obtuvieron dos grupos de datos, uno mediante la secuenciación Sanger del gen mitocondrial COI, y otro con la técnica de 3RAD. La finalidad de estos grupos de datos fue investigar la diversidad genética del género y realizar inferencias sobre su evolución y su diversidad taxonómica.

Los genes mitocondriales, y en particular el gen COI, han sido ampliamente utilizados para reconstruir relaciones filogenéticas, realizar inferencias evolutivas y aproximarse a la delimitación de especies (Hajibabaei *et al.*, 2006; Hebert *et al.*, 2003; Stahlhut *et al.*, 2013). Las principales ventajas de utilizar este gen son que, como el resto de la mitocondria, no tiene recombinación genética y su herencia es matrilineal, aunado a que se encuentra relativamente conservado en muchos linajes de seres vivos porque está estrechamente relacionado con el proceso de respiración celular, aunque muestra suficiente variación para poder realizar estudios intraespecífico y su secuenciación es relativamente sencilla y económica (Folmer *et al.*, 1994; Hebert *et al.*, 2004; Papadopoulou *et al.*, 2010).

Las principales desventajas de usar este gen como único marcador son que en algunos casos puede presentar sesgos, debido a su herencia matrilineal y la reconstrucción del árbol del gen puede diferir de la historia de divergencia de las especies (Maddison, 1997; Prugnolle y Meeus, 2002; Singhal *et al.*, 2018). Los problemas que presenta la secuenciación del gen COI, indican que utilizar diversas fuentes de información y análisis de datos para respaldar una hipótesis, por ejemplo de delimitación de especies, puede resultar en una mejor aproximación a la realidad de los fenómenos naturales (Carstens *et al.*, 2013).

En este trabajo se secuenciaron muestras de *S. bruchivora* correspondientes al gen mitocondrial COI como una primera aproximación a investigar la diversidad del grupo. Los análisis realizados con este conjunto de datos fueron contrastados con los resultados del análisis de datos genómicos nucleares, y permitieron establecer hipótesis de delimitación de especies y estimar los tiempos de divergencia de los principales clados del grupo.

Los datos genómicos en este trabajo consisten en la secuenciación de miles de loci representativos del genoma de los organismos utilizando una técnica de secuenciación de nueva generación (SNG). El principal avance de estos métodos, además de ser relativamente más económicos que una secuenciación completa, es la enorme cantidad de información que

recuperan en forma de miles a millones de loci (Goodwin *et al.*, 2016). Esta información permite sobreponerse a dificultades como la incongruencia entre árboles de genes y árboles de especies a través del modelado y análisis estadístico de miles o millones de loci (Luo *et al.*, 2018) y permite la reconstrucción de filogenias a diferentes profundidades de tiempo evolutivo (Eaton *et al.*, 2017; Harvey *et al.*, 2016).

Algunas de las técnicas de SNG más utilizadas en trabajos de filogeografía, filogenética, evolución y genética de poblaciones son la captura de elementos ultra-conservados (UCEs) y la secuenciación asociada a sitios de restricción (RADSeq), en este trabajo utilizamos un método de RADSeq porque que no requiere de un genoma de referencia para el ensamble de loci, por lo que se puede trabajar con organismos no modelo (Andrews *et al.*, 2016a). Además, permite recuperar un mayor número de loci y de sitios nucleotídicos polimórficos (SNPs) a un menor costo para realizar inferencias evolutivas especialmente en grupos de reciente divergencia (Harvey *et al.*, 2016; Leaché *et al.*, 2015). La principal limitante al uso de datos genómicos es computacional, ya que el análisis de grupos de datos tan grandes requiere mucho poder y tiempo, aunado a que aún existe un número limitado de programas adaptados a este tipo de datos.

Algunas de las principales limitaciones del método de RADseq son que requiere de la amplificación de los fragmentos mediante PCR (Polymerase Chain Reaction), lo cual introduce errores de secuenciación en los fragmentos amplificados que no pueden ser identificadas fácilmente. Generalmente se asume que estos errores serán descartados durante el filtrado de datos, ya que la mayoría no estará dentro del rango de similitud establecido por el usuario para considerar secuencias homólogas. Una aproximación a la identificación de estos errores consiste en insertar un primer especial durante la construcción de librerías, este primer está compuesto por una secuencia aleatoria de 8 pares de bases a manera de etiqueta, la cual permitirá identificar los duplicados de PCR durante el procesamiento bioinformático (Andrews *et al.*, 2014; Glenn *et al.*, 2017; Hoffberg *et al.*, 2016). En este estudio se optó por determinar de manera óptima el parámetro correspondiente al umbral de similitud al que se agrupan secuencias homólogas (0.9), calculado a partir de las frecuencias alélicas de cada muestra como proponen Ilut, Nydam y Hare, (2014).

Otro problema que presentan los datos obtenidos mediante métodos de RAD es la pérdida de alelos por polimorfismos en los sitios de restricción, de tal modo que la enzima endonucleasa no reconoce el sitio y el alelo no estará presente en la matriz final (Eaton *et al.*, 2017). Los polimorfismos en estos sitios serán mayores si se trabaja con poblaciones altamente divergentes, ya que la probabilidad de que hayan ocurrido mutaciones en los sitios de restricción es mayor (DaCosta y Sorenson, 2014; Leaché *et al.*, 2015). En el presente trabajo se secuenciaron individuos de un género monotípico del cual se sospecha una mayor riqueza de especies. No obstante, es muy probable que esta diversidad esté asociada a procesos de

divergencia recientes y por lo tanto la pérdida de alelos por polimorfismos en sitios de restricción sea menor.

La obtención de datos genómicos consta de dos etapas principales. La primera es la obtención de los fragmentos, lo cual incluye la elección de la técnica de secuenciación, el trabajo de laboratorio y la secuenciación *per se*. La segunda etapa consiste en el procesamiento bioinformático de los fragmentos, es decir, la asignación de cada uno a una muestra y el ensamble de loci. Existen dos programas que son comúnmente usados para el procesamiento bioinformático de las secuencias obtenidas mediante cualquier variación a la técnica de RADseq: Stacks (Catchen *et al.*, 2013) y iPyrad (Eaton, 2014). Los resultados usando ambos programas son similares (Natalia Bayona, comentario personal), no obstante, iPyrad está asociado a la plataforma Jupyter Notebook (Kluyver *et al.*, 2016), que permite usar el código base de programas como iPyrad para construir un registro del procesamiento bioinformático al que se someten las muestras, facilitando la eficiencia, organización y reproducibilidad de los análisis.

La principal característica de los programas que permiten procesar muestras de SNG es que poseen un gran número de parámetros que pueden ser modificados por el usuario y que permiten ajustar el proceso a las características de los datos, las cuales variarán dependiendo del tipo de enzimas utilizadas para su obtención, el tamaño de los fragmentos, si los organismos pertenecen a linajes cercanamente relacionados o no, la ploidía etc. Los valores asociados a cada parámetro pueden ser variados dentro de un rango definido por el usuario, y afectarán el tamaño y características de la matriz final de datos (Ilut *et al.*, 2014; Paris *et al.*, 2017). La mayoría de los trabajos que usan datos de SNG generan varias matrices usando diferentes valores en los parámetros con el fin de evaluar cómo se ven afectados los resultados en análisis, por ejemplo de reconstrucción filogenética o filogeografía (p. ej Salas-Lizana y Oono, 2018). Generalmente, las matrices finales varían en el número de loci, el número de datos faltantes y el número de SNPs.

En el presente trabajo se utilizó el primer paso de Stacks para asignar los fragmentos por muestra, ya que se decidió utilizar únicamente el primer fragmento producto de la doble digestión enzimática. Los siguientes pasos fueron desarrollados en iPyrad y utilizamos valores diferentes en tres parámetros principales para construir cinco matrices.

a) Clust\_threshold: este parámetro regula el porcentaje de similitud al cual dos secuencias son identificadas como homólogas y asignadas al mismo grupo, permitiendo diferenciar entre secuencias ortólogas y parálogas.

Los cambios en este parámetro implican que si el valor es muy alto, aumenta la probabilidad de separar secuencias en diferentes grupos que en realidad corresponden al mismo, es decir que son alelos del mismo locus. Al contrario, si el valor del parámetro es laxo, se corre

el riesgo de agrupar secuencias que en realidad corresponden a errores de secuenciación o parálogos.

Para definir el valor de similitud apropiado en nuestros datos, utilizamos la aproximación de Ilut *et al.*, (2014) que consiste en graficar la distribución del número de alelos a lo largo de diferentes grados de divergencia entre secuencias, con la finalidad de elegir el valor en donde se minimiza el número de loci homólogos y se estabiliza la acumulación de loci con tres o más alelos. Debido a la alta demanda computacional del proceso de asignación de alelos por loci, se eligieron 40 muestras para calcular la proporción de alelos por cada una de ellas, y se estableció un rango de máxima divergencia permitida de 0 a 20% para tomar en cuenta que la especie *S. bruchivora* es posiblemente un complejo de especies. Con base en el gráfico, se determinó usar el valor de divergencia de 0.1, en donde se observa una tendencia a la disminución de los alelos homocigos y el aumento de los parálogos. Este valor corresponde a un 10% de divergencia permitida y a un valor de 0.9 o 90% de similitud.

- b) Max\_alleles\_consens: este parámetro define cuántos alelos por loci estarán permitidos, lo cual está directamente asociado con la ploidía del organismo.

Dado que las avispas son organismos haplo-diploides, es decir, que los machos son haploides y las hembras diploides, los datos fueron separados dependiendo del sexo y analizados por separado en dos etapas del procesamiento bioinformático que corresponden a la estimación de la heterocigocidad y el filtrado de secuencias con base en la tasa de error. Los valores de este parámetro son 1 y 2 según la ploidía del individuo.

- c) Min\_sample\_locus: este parámetro controla el mínimo número de muestras que deben tener información de un locus para que este sea incluido en la matriz de datos final. El valor de este parámetro determinará el número de loci y de datos faltantes en la matriz resultante.

El resultado de procesar las secuencias obtenidas para las muestras de *Stenocorse* fueron cinco matrices con características diferentes (Tabla 1). Dado el bajo número de loci recuperados para las muestras correspondientes al grupo externo, se construyeron dos grupos de matrices, uno incluyendo las muestras del grupo externo (M3 y M4), y otro sin ellas (M1, M2 y M2.5) con el objetivo de evaluar el efecto de estas muestras en los análisis.

El mayor número de loci y sitios informativos (psi) fue registrado para las matrices M1 y M3, ya que se determinó que los loci retenidos estuvieran compartidos en mínimo 10% de las muestras con el objetivo de realizar el análisis de reconstrucción filogenética mediante inferencia Bayesiana (Eaton *et al.*, 2017; Tripp *et al.*, 2017). Las matrices M2 y M4 fueron construidas para evaluar el efecto de los datos faltantes en los análisis de reconstrucción

filogenética. La matriz M2.5 fue construida con los mismos parámetros que la matriz M2 y con las mismas muestras que todas las otras matrices, únicamente excluyendo la muestra correspondiente a la localidad de Texas, EU dado el bajo número de loci recuperados. Debido a que la matriz M2.5 fue construida reteniendo loci compartidos en mínimo 50% de las muestras representa la matriz con el menos número de datos faltantes, aunque únicamente recuperó 532 locy y 5247 sitios nucleotídicos polimórficos (SNPs). Por sus características, esta matriz fue utilizada para realizar el análisis de estructura genética en el programa STRUCTURE, así como la evaluación de delimitación de especies en el programa BPP. Estos análisis resultan computacionalmente muy demandantes y requieren de matrices de tamaños pequeños y selecciones aleatorias de un número reducido de loci que se encuentren presentes en la mayor cantidad de muestras posibles.

### **Delimitación de especies**

Uno de los principales obstáculos a la delimitación de especies en grupos mega-diversos como los insectos, y en particular las avispas parasitoides, es la ausencia de un sistema de caracteres morfológico estandarizado, ya que generalmente se trata de organismos muy pequeños que no presentan una diferenciación morfológica clara (Cruaud *et al.*, 2018; Peixoto *et al.*, 2018). Las avispas del género *Stenocorse* miden aproximadamente de 3 a 5 mm de longitud y presentan una morfología que las distingue claramente de otros géneros, pero que hasta el momento no ha sido revisada con el fin de investigar si existe una mayor diversidad específica.

Con la finalidad de esclarecer la presencia de linajes independientes dentro del género, primero se realizó un análisis de descubrimiento usando el método de GMYC implementado en R con base en los datos de secuenciación del gen mtCOI (Carstens *et al.*, 2013). Este método busca modelar el punto de transición entre la cladogénesis y la coalescencia de alelos bajo el supuesto de que la primera sucede a una tasa mucho más lenta que la segunda. Una vez que se ha determinado el umbral, los individuos son asignados a posibles especies diferentes si coalescen con otro individuo antes o después de este punto de transición (Carstens *et al.*, 2013; Luo *et al.*, 2018).

La topología recuperada mediante este método presentó las muestras agrupadas en nueve posibles especies separadas en dos clados principales, con cierto grado de estructura geográfica. La composición del clado correspondiente a la especie 2 llama la atención, ya que está formado por cuatro subclados con las muestras de la MC, del ENVT, CA y Texas. Para investigar la existencia de mayor estructura dentro de este clado y para corroborar la asignación de las especies GMYC, se calcularon las distancias genéticas sin corregir.



Nuestros datos muestran que todas las posibles especies delineadas por el método de GMYC presentan una divergencia genética mucho mayor al 2%, excepto en el caso de los grupos de muestras correspondientes a la especie 2. En este caso, el promedio de las diferencias suele ser mayor al 2%, pero menor en casos particulares, y en el caso de las muestras correspondientes al ENVT, el promedio es menor al 2% con respecto a las otras muestras de la especie 2. Este resultado sugiere que todas estas muestras corresponden a un mismo linaje genético que posiblemente presenta una subestructuración a nivel poblacional. No obstante, se requiere más evidencia para corroborar esto.

Se utilizó el grupo de datos de secuenciación genómica para realizar un análisis de delimitación de especies bajo el modelo de coalescencia multiespecies usando el programa BPP (Bayesian Phylogenetics and Phylogeography) (Yang, 2015). Esta aproximación se considera como de validación de especies y procede a las aproximaciones de descubrimiento, como el método de GMYC, ya que arroja valores de probabilidad posterior para un hipótesis de especies previamente asignada por el usuario (Carstens *et al.*, 2013).

En el presente estudio se utilizaron los resultados de GMYC, distancias genéticas, así como la estructuración geográfica de las muestras para establecer las hipótesis de delimitación de especies. Se generó un árbol mostrando estas posibles especies y se utilizó el archivo .loci resultante del procesamiento de las muestras en iPyrad para realizar el análisis de BPP en la plataforma de Jupyter Notebook. Se realizaron diez repeticiones del análisis, de los que no se observó variación, para evaluar la congruencia de los resultados. Por restricciones computacionales se utilizó un muestreo de 50 loci elegidos de manera aleatoria en cada repetición con base en la matriz M2.5. Para explorar el efecto de la asignación de *priors*, el análisis se realizó dos veces usando dos combinaciones diferentes de *priors* difusos correspondientes a tamaños poblacionales pequeños y tiempos de divergencia profundos, así como tamaños poblacionales grandes y tiempos de divergencia recientes (Yang, 2015).

Los resultados de ambos análisis fueron congruentes en las 10 repeticiones; el análisis apoyó con valores de probabilidad posterior mayores al 95% todos los nodos de la hipótesis y solo en algunos casos se recuperó con bajo soporte el colapso del nodo correspondiente a las muestras del ENVT y MC. Este resultado puede ser interpretado como si cada una de nuestras hipótesis de especies correspondieran con linajes con evolución independiente o, como se ha visto en estudios recientes, porque el modelo de coalescencia multi-especies delimita estructura poblacional tanto a nivel de especies como a nivel de poblaciones intraespecíficas, sin un criterio claro para diferenciar una de otra (Sukumaran y Knowles, 2017).

La delimitación de linajes con evolución independiente se realizó con base en la congruencia de los tres métodos de delimitación de especies realizados con genes

mitocondriales y nucleares. De esta manera se propuso un total de seis especies y tres especies adicionales determinadas únicamente con información mitocondrial. El análisis morfológico posterior de estos linajes ayudará a confirmar su estatus.

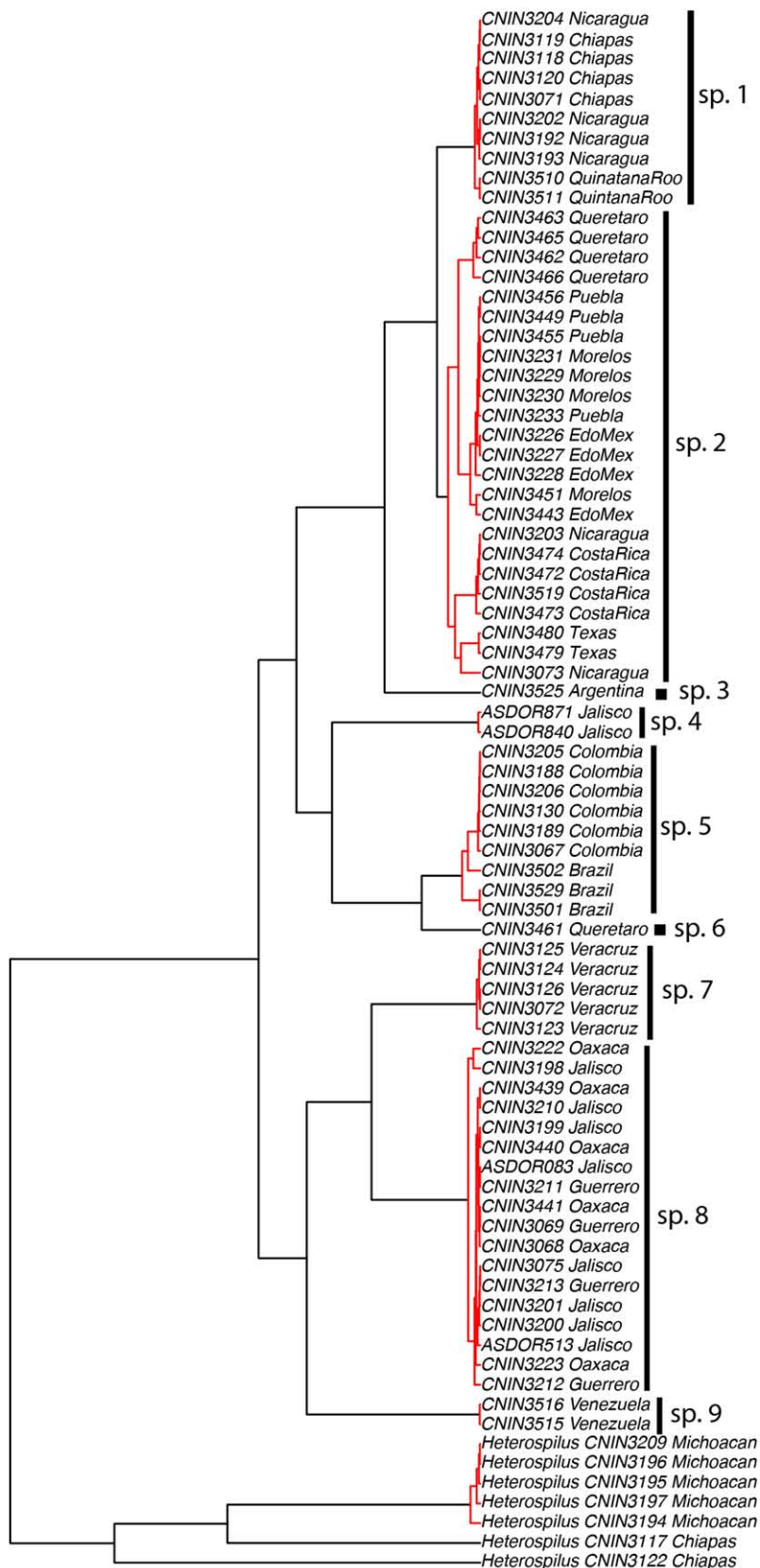
### **Estimación de tiempos de divergencia y relaciones filogenéticas con datos mitocondriales**

Dado que la especiación es un proceso continuo y prolongado, el tiempo de divergencia entre grupos juega un papel muy importante en la formación de especies o de poblaciones estructuradas, de tal manera que a mayor tiempo de divergencia, la probabilidad de que dos poblaciones aisladas hayan divergido será mayor (Carstens *et al.*, 2013; Sukumaran y Knowles, 2017).

La tasa de mutación estimada para el gen mtCOI (Papadopoulou *et al.*, 2010) permitió reconstruir las relaciones de parentesco entre los individuos y estimar los tiempos de divergencia en los principales nodos de la filogenia del género. Las topologías resultantes son altamente congruentes entre ellas y con aquella recuperada por el método de GMYC.

La divergencia del género con respecto al grupo externo data del Mioceno tardío (15-8.1 Ma), mientras que la divergencia de los dos principales clados del grupo data del Mioceno tardío a principios del Plioceno (6.8-3.9 Ma). Se estimó que la diversificación de las posibles especies actuales ocurrió durante el Plioceno tardío hasta el final del Pleistoceno, y en particular que las muestras de la provincia veracruzana divergieron de su grupo hermano compuesto por las muestras de la provincia de las tierras bajas del pacífico hace 3.8 a 1.8 millones de años. Las muestras de la provincia del Eje Neovolcánico Transversal divergieron de aquellas de la meseta central hace 0.8 a 0.3 millones de años (Figura 1).

En conjunto, las estimaciones de tiempos de divergencia hablan de un grupo compuesto de dos clados principales que posiblemente representan linajes con evolución independiente. El tiempo de divergencia estimado para el nodo compuesto por los clados de la provincia de VCZ y del ENVT también indica la posibilidad de que estos dos grupos formen dos linajes independientes. No obstante, el clado compuesto por las muestras de las provincias de la MC, ENVT, CA y Texas presenta tiempos de divergencia muy recientes, razón por la cual el método de GMYC los agrupa en una posible especie, mientras que el análisis BPP probablemente está detectando estructura poblacional intraespecífica. El posterior análisis del grupo de datos correspondiente a estas muestras bajo un enfoque de genética de poblaciones permitirá una mejor comprensión de los procesos que se encuentran influenciando su variabilidad genética y distribución geográfica.



**Figura 4.** Árbol ultramétrico obtenido con el programa BEAST. Con barras negras se muestran las especies delimitadas en el estudio.

## **Estructura y relaciones filogenéticas con datos genómicos**

Para investigar la existencia de estructura genética debido a linajes de especies o linajes de poblaciones realizamos un análisis de Structure usando la matriz M2.5 (Pritchard *et al.*, 2010). El programa evalúa las diferencias en la distribución de la variación genética entre poblaciones usando un algoritmo Bayesiano y asigna las muestras por poblaciones con base en los patrones de variación que comparten (Porrás-Hurtado *et al.*, 2013). Aunado a este análisis, reconstruimos las relaciones de parentesco entre las muestras usando inferencia bayesiana y el método de SVDQuartets con base en las matrices M2, M2, M3 y M4 de los datos genómicos.

La reconstrucción filogenética con base en la matriz M1 recupera topologías altamente congruentes con ambos métodos y los árboles fueron enraizados con base en las relaciones recuperadas en el análisis del grupo de datos mitocondriales. Encontramos incongruencias en el caso de la muestra de Texas, la cual se encuentra cercanamente relacionada al clado ENVT+MC+CA, pero sin resolver su posición dentro del clado. La incongruencia probablemente se debe al bajo número de loci obtenidos para esta muestra. No obstante, se localiza recurrentemente dentro o como grupo hermano del mismo clado, la distancia genética con respecto a las muestras de la provincia CP es muy baja (1.8%) y en la topología bayesiana incluso posee un valor alto de probabilidad posterior, indicando que pertenece a este grupo genético.

Dos muestras de Querétaro (MC) están consistentemente anidadas o como grupo hermano de las muestras de Colombia y Colombia+Brasil en el caso de los datos mitocondriales. Estas muestras fueron revisadas morfológicamente y la similitud indica pertenencia a este clado, sugiriendo que las muestras de Querétaro representan una especie cercanamente relacionada al clado Colombia+Brasil, o que se trata de poblaciones muy estructuradas genéticamente.

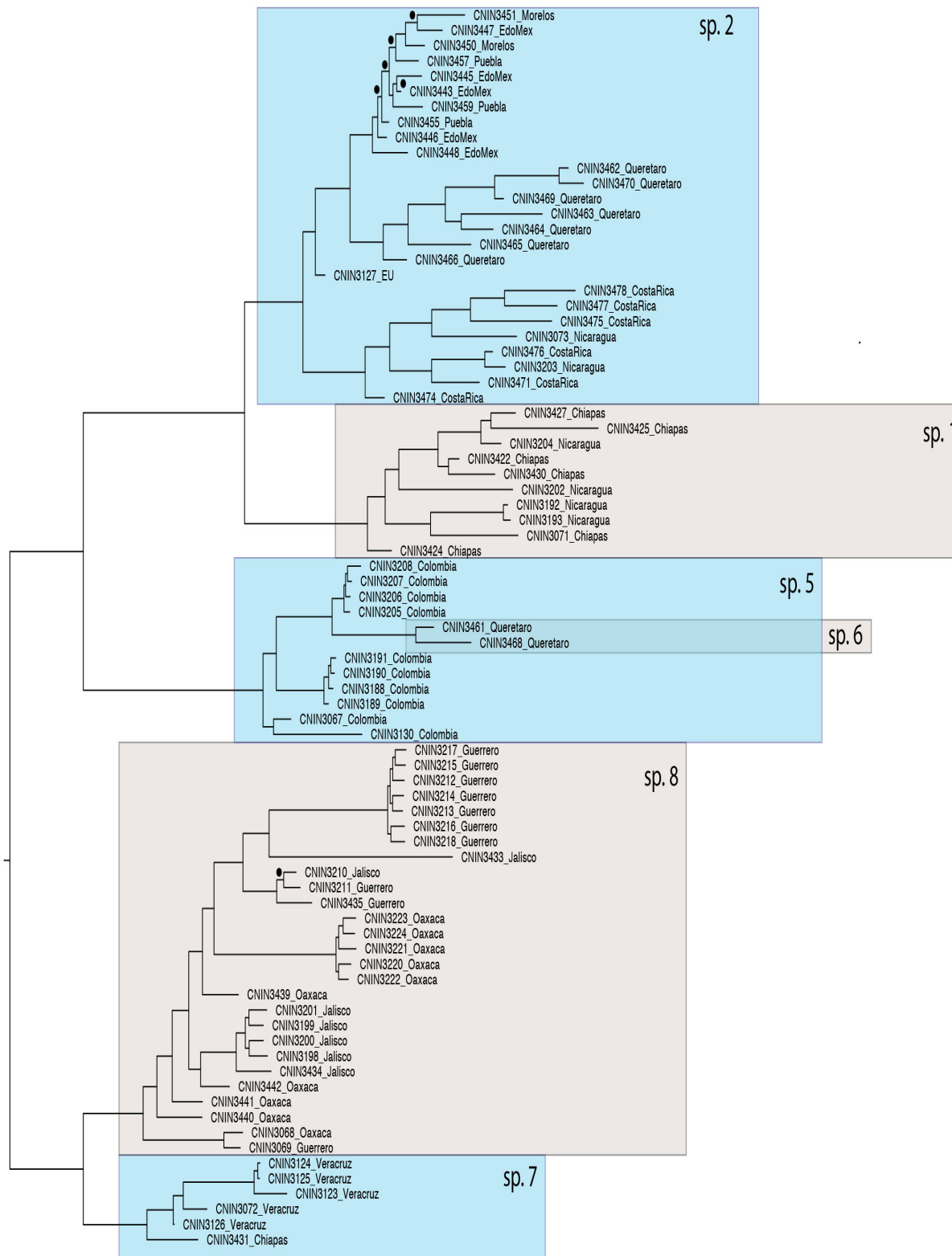
La principal diferencia entre las reconstrucciones filogenéticas con datos genómicos y con datos mitocondriales fue la resolución de los clados de divergencia reciente. La topología del gen mtCOI muestra los principales clados que componen el género y es una buena aproximación al descubrimiento de los principales linajes, así como la estimación del tiempo de divergencia, mientras que los datos genómicos permitieron la resolución de los clados más recientes y una mejor apreciación de su estructura geográfica.

El análisis de Structure recuperó el valor más alto para el promedio de máxima verosimilitud [ $\ln P(D)$ ] en  $K=4$ , indicando la existencia de cuatro grupos genéticos con patrones de variación similares. Los grupos genéticos corresponden con la distribución geográfica de las muestras y con las relaciones recuperadas tanto con los datos genómicos como mitocondriales.

La separación de dos grupos genéticos correspondientes a las muestras de las TBP y VRZ y su congruencia con análisis anteriores es evidencia de que estos dos grupos corresponden a linajes con evolución independiente. La presencia de mezcla entre el grupo genético de VRZ y algunas muestras de las TBP puede deberse a la presencia de polimorfismos ancestrales, introgresión histórica o flujo génico moderado. La introgresión histórica puede ser determinada a través de un análisis de estadístico D (Durand *et al.*, 2011), el cual detecta la presencia de mezcla genética entre linajes divergentes con base en la frecuencia de SNPs que son discordantes con respecto a una hipótesis de especies (Eaton *et al.*, 2015). Esta prueba también permite distinguir entre sorteo incompleto de linajes e introgresión histórica (Alexander *et al.*, 2017). Mientras que una prueba de Mantel, indicadora de aislamiento por distancia, puede arrojar luz sobre los patrones de dispersión del género y la causa de la distribución de su variación genética (Jeffries *et al.*, 2015).

Los grupos genéticos restantes corresponden a dos de los principales clados detectados en la reconstrucción filogenética con ambos marcadores (ENVT y Colombia+Querétaro), de los cuales, cada uno está compuesto de subgrupos que fueron delimitados por el análisis de BPP y que también se encuentran como clados de reciente divergencia en la reconstrucción filogenética (Fig. 2). En el caso de las muestras de Brasil que no pudieron ser incluidas en el grupo de datos de RADseq, estas se agruparon consistentemente junto con las de Colombia usando datos mitocondriales, y su diferenciación genética fue muy baja, sugiriendo que pertenecen a este grupo genético. No obstante, en el caso de las muestras de Venezuela, Argentina y Jalisco, que únicamente poseen información mitocondrial, no fue posible asignarlas un grupo genético definido.

Esta configuración jerárquica en la estructura de las muestras, sugiere que el género *Stenocorse* está compuesto por un mosaico de linajes de especies y linajes de poblaciones estructuradas genéticamente. Los linajes de especies están representados en este trabajo por las muestras de las provincias de las TBP, VRZ y Colombia+Brasil+Querétaro, mientras que los linajes de poblaciones están representados por los clados de CA, MC, ENVT y Texas.



**Figura 5.** Filograma Bayesiano obtenido con base en la matriz M1 usando el programa Exabayes. Los círculos negros corresponden a probabilidades posteriores < 95%. En las cajas de color se indican las especies delimitadas con datos de 3RAD y mitocondriales.

## Biogeografía e interacciones

México es uno de los países con mayor biodiversidad en el mundo, lo cual se correlaciona con su complejidad y variedad fisiográfica que a su vez genera variación en altitud y climas a lo largo del territorio hasta Centro América (Ferrusquía-Villafranca, 1998; Gutiérrez-García y Vázquez-Domínguez, 2013). La formación de esta variedad fisiográfica está estrechamente relacionada con la actividad volcánica causante del levantamiento de cadenas montañosas, aunado a ciclos de glaciaciones y cambios en el nivel del mar que influyeron en la historia evolutiva y patrones de distribución de diversos organismos (Mastretta-Yanes et al., 2015; Pedraza-Lara et al., 2015; Zaldívar-Riverón *et al.*, 2004).

Las cadenas montañosas de las Sierras Madres Oriental y Occidental separan el territorio al este y oeste de la Meseta Central al norte del territorio. En el sur, la formación de la Sierra madre del Sur durante el Mesozoico y Cenozoico temprano (Ferrusquía-Villafranca, 1998), seguido del levantamiento del Eje Neo Volcánico Transversal durante el Mioceno y el Plioceno hace 16 a 1.8 millones de años (Ferrari *et al.*, 2012), forman una barrera entre las costas del Pacífico y el golfo de México.

Los datos obtenidos en este estudio indican que la estructura genética asociada a las muestras de *Stenocorse* posiblemente está influenciada por áreas de elevaciones altas tales como la Sierra Madre del Sur y la Sierra Madre Oriental. Estas cadenas montañosas separan las muestras de las provincias de la MC y el ENVT de aquellas poblaciones asociadas a las provincias costeras, TBP y VRZ. La diversidad genética de las muestras sugiere que se trata de tres linajes (TPB, VRZ y MC+ENVT+Texas) que tal vez se encuentran pasando por un proceso de especiación alopátrica. La presencia de mezcla genética entre las muestras de la provincia de VRZ y TPB puede deberse a la presencia de polimorfismos ancestrales, introgresión histórica o flujo génico. La región del Istmo de Tehuantepec en México es una zona continental con montañas bajas (Gutiérrez-García y Vázquez-Domínguez, 2013), lo cual sugiere que el flujo génico entre las provincias costeras del Pacífico y Atlántico es posible.

Las muestras de las provincias del ENVT, Texas y la MC corresponden a un grupo genético que se extiende hasta Centro América. El tiempo de divergencia estimado para estos clados es relativamente reciente con respecto al de las muestras costeras, indicando la posibilidad de que el grupo se dispersó recientemente de una región a otra. No obstante, es necesario un esfuerzo de colecta mayor para poder entender qué procesos están causando esta distribución que parece no ser interrumpida por barreras geográficas evidentes, y que presentan mezcla genética con el grupo de la provincia de VRZ al sur del Istmo de Tehuantepec. Las muestras colectadas en Sudamérica sugieren que la diversidad del grupo es mucho mayor a la observada en este estudio y que un muestreo exhaustivo podría arrojar un mayor número de linajes con evolución independiente.

Los organismos parasitoides y sus huéspedes están estrechamente relacionados, por lo que estas interacciones son un elemento importante en su historia evolutiva. La selección de caracteres que aumentan la adecuación del parasitoide con respecto a un huésped particular puede causar especiación en la ausencia de barreras geográficas en la forma de especiación simpátrica (Hall *et al.*, 2017; Stireman *et al.*, 2006). Las muestras correspondientes a la región de CA fueron recuperadas en clados diferentes en la reconstrucción filogenética y como dos linajes separados en el análisis de BPP y GMYC, sugiriendo el aislamiento de estas poblaciones por razones que podrían estar más relacionadas con la biología del grupo que con barreras geográficas.

Los organismos parasitoides y en particular las avispas tienen ciclos de vida complejos que están estrechamente asociados con el ciclo de vida de sus huéspedes fitófagos. Se utilizó la topología obtenida mediante inferencia Bayesiana del conjunto de datos genómicos para relacionar las especies de brúquinos y plantas de las que fue obtenida cada muestra, para explorar la existencia de un patrón (Fig. 3). La mayor parte de los escarabajos de la subfamilia Bruchinae poseen un rango de plantas hospederas limitado y conformado por taxones cercanamente relacionados (Kergoat *et al.*, 2004). Esto se ve reflejando en nuestro datos, ya que observamos una tendencia a obtener la misma especie de brúquino en las muestras que fueron colectadas de la misma planta en localidades diferentes. Por otro lado, aquellas muestras correspondientes a un mismo clado y que fueron obtenidas de plantas diferentes, fueron en su mayoría asociadas a especies de brúquinos diferentes. La tendencia de nuestros datos hacia una estructura en la interacción brúquino-parasitoide puede estar causada por la especialización de los grupos de brúquinos a ciertas plantas y no por la especialización del parasitoide en alguna especie de brúquino.

Aunado a esto, observamos que en varias localidades se obtuvieron hasta tres especies de brúquinos emergiendo de las vainas de la misma planta. En estos casos no fue posible determinar con exactitud la especie de brúquino de la que emergieron los parasitoides, e incluso se ha reportado que una semilla puede estar infestada por larvas de hasta cuatro especies diferentes de brúquinos (Jesús Romero Nápoles, comentario personal).

La ausencia de patrones evidentes en la filogenia de *Stenocorse* sugiere que estas avispas poseen hábitos alimenticios generalistas. Esta posibilidad implicaría que las adaptaciones fisiológicas que los brúquinos han desarrollado para contrarrestar las defensas de las plantas y su consecuente especialización, no han causado la diversificación de las avispas del género *Stenocorse*. La falta de especialización en *Stenocorse* puede deberse a que se trata de avispas ectoparasitoides, sobre las cuales las defensas inmunológicas del hospedero no representan una fuerza de selección con la magnitud suficiente para causar una diferencia de adecuación. La falta de especialización en avispas parasitoides ha sido reportada anteriormente (Cronin y Abrahamson, 2001), sugiriendo que la diversificación del huésped, por el mecanismo que sea,



no es la principal causa de diversificación en este tipo de parasitoides. No obstante, se requiere un mayor escrutinio de las características biológicas de estos insectos tales como los mecanismos de localización del huésped, ciclo de vida y formas de apareamiento, así como un muestreo ampliado a todas las familias de plantas en las que se han encontrado brúquinos.

Las plantas de la familia Fabaceae están asociadas a un gran número de insectos, entre ellos, los brúquinos se alimentan en su mayoría de semillas de alguna especie de leguminosa (Araiza *et al.*, 2005). No obstante, se sabe que también utilizan como alimento a las semillas de aproximadamente 30 familias diferentes de plantas, por lo cual sabemos que se trata de un grupo muy diverso con interacciones complejas (Center y Johnson, 1974). Debido a las limitaciones de muestreo del presente trabajo, se obtuvieron muestras de brúquidos y avispas parasitoides de un número limitado de especies de plantas, en su mayoría pertenecientes a la familia Fabaceae. Sin embargo, un mayor muestreo diseñado con base en la diversidad de plantas y su distribución arrojará un mayor número de especies tanto de brúquinos como de avispas parasitoides. Esto también permitirá realizar inferencias sobre la influencia de esta interacción en la evolución y distribución actual de todos los grupos involucrados, sobre lo cual este trabajo representa un precedente.

## CONCLUSIONES

- Se investigó la sistemática y evolución del género de avispas parasitoides *Stenocorse* Marsh, 1968 usando secuencias de ADN nuclear (3RADSeq) y mitocondrial (COI).
- Las filogenias reconstruidas con secuencias nucleares y mitocondriales fueron congruentes mostrando tres clados principales que indican estructura geográfica.
- El análisis de estructura genética recupera las muestras en cuatro grupos que coinciden parcialmente con los clados de la reconstrucción filogenética y con la estructura geográfica que presentan.
- La reconstrucción filogenética y la estructura de la variación genética de las muestras con los dos marcadores permitió obtener una hipótesis de delimitación de especies que fue explorada usando métodos de coalescencia y de divergencia genética.
- Con base en la congruencia de los resultados de la delimitación empleando información nuclear y mitocondrial se sugiere la existencia de seis especies dentro del género *Stenocorse*.
- Se determinaron tres linajes adicionales a partir de la información mitocondrial y que requieren de un mayor muestro para ser validadas.

- Se estimó que la divergencia del género ocurrió durante el Mioceno tardío (15.1–8.14 Ma), y que la divergencia de las especies actuales ocurrió durante el Plioceno temprano y el Pleistoceno tardío (5.71-0.72 Ma).
- Se encontraron tres casos de simpatria en las especies que coexisten en las provincias de la Meseta Central y en las especies de las Planicies Costeras del Golfo, en Querétaro y Veracruz respectivamente, así como en el caso de las especies de Nicaragua, Centro América. Esto confirma la existencia de barreras al flujo génico y sienta las bases para estudios posteriores.
- Los resultados sugieren que las elevaciones altas, tales como el Eje Neovolcánico Transversal y la Sierra Madre del Sur pueden actuar como barreras al flujo génico entre las poblaciones de *Stenocorse*, promoviendo la estructuración de la variación genética. También se propone que la conexión entre América del Sur y América Central, hace 9 Ma, pudo haber causado eventos de dispersión y especiación en el grupo.
- No se encontró evidencia de especialización en los hábitos alimenticios de las especies de *Stenocorse*, lo cual es contrario a lo que se esperaba, debido a la especialización brúquino-planta y el efecto en cascada que esto puede causar en la especiación de sus parasitoides. No obstante, es evidencia que sugiere que el sistema inmunológico en las avispas ectoparasitoides no se encuentra bajo una fuerza de selección tan fuerte como en el caso de las endoparasitoides en donde existe un mayor número de casos de especialización parasitoide-huésped.
- El uso de secuencias mitocondriales como COI permite obtener información importante en el planteamiento de hipótesis de delimitación de especies y representa una adecuada primera aproximación para el estudio de la diversidad de insectos.
- El análisis de secuencias genómicas ha permitido complementar y en muchos casos sobreponerse a las limitaciones de las inferencias realizadas con información de genes puntuales y caracteres morfológicos o ecológicos.
- Conocer la historia natural de los organismos sigue siendo fundamental en la era molecular, y solo la integración del conocimiento permitirá profundizar en el entendimiento de la naturaleza y sus procesos.

## Referencias

- Alexander, A. M., Su, Y. C., Oliveros, C. H., Olson, K. V., Travers, S. L. & Brown, R. M. (2017). Genomic data reveals potential for hybridization, introgression, and incomplete lineage sorting to confound phylogenetic relationships in an adaptive radiation of narrow-mouth frogs. *Evolution*, **71**, 475–488.
- Andrews, K. R., Hohenlohe, P. A., Miller, M. R., Hand, B. K., Seeb, J. E. & Luikart, G. (2014). Trade-offs and utility of alternative RADseq methods: Reply to Puritz et al. *Molecular Ecology*, **23**, 5943–5946.
- Araiza, M. D. S., Nápoles, J. R. & Aguilera, E. G. (2005). Contribución al estudio de los brúquidos (insecta: coleóptera) asociados a fabaceas arbustivas. *Acta Universitaria*, **11**, 26-32.
- Avice, J. C. (2009). Phylogeography: Retrospect and prospect. *Journal of Biogeography*, **36**, 3–15.
- Belokobylskij, S. A., Zaldivar-Riverón, A. & Quicke, D. L. J. (2004). Phylogeny of the genera of the parasitic wasps subfamily Doryctinae (Hymenoptera: Braconidae) based on morphological evidence. *Zoological Journal of the Linnean Society*, **142**, 369–404.
- Bryson, R. W., García-Vázquez, U. O. & Riddle, B. R. (2012). Relative roles of Neogene vicariance and Quaternary climate change on the historical diversification of bunchgrass lizards (*Sceloporus scalaris* group) in Mexico. *Molecular Phylogenetics and Evolution*, **62**, 447–457.
- Center, T. D. & Johnson, C. D. (1974). Coevolution of Some Seed Beetles ( Coleoptera : Bruchidae ) and Their Hosts. *Ecology*, **55**, 1096–1103.
- Davey, J. L. & Blaxter, M. W. (2010). RADseq: Next-generation population genetics. *Briefings in Functional Genomics*, **9**, 416–423.
- Delsuc, F., Brinkmann, H. & Philippe, H. (2005). Phylogenomics and the reconstruction of the tree of life. *Nature Reviews Genetics*, **6**, 361–375.
- Faircloth, B. C., Branstetter, M. G., White, N. D. & Brady, S. G. (2015). Target enrichment of ultraconserved elements from arthropods provides a genomic perspective on relationships among hymenoptera. *Molecular Ecology Resources*, **15**, 489–501.
- Faircloth, B. C., McCormack, J. E., Crawford, N. G., Harvey, M. G., Brumfield, R. T. & Glenn, T. C. (2012). Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. *Systematic Biology*, **61**, 717–726.
- Forbes, A. A., Devine, S. N., Hippee, A. C., Tvedte, E. S., Ward, A. K. G., Widmayer, H. A. & Wilson, C. J. (2017). Revisiting the particular role of host shifts in initiating insect speciation. *Evolution*, **71**, 1126–1137.
- Fujita, M. K., Leaché, A. D., Burbrink, F. T., McGuire, J. A. & Moritz, C. (2012). Coalescent-

- based species delimitation in an integrative taxonomy. *Trends in Ecology and Evolution*, **27**, 480–488.
- Goulet, H. & Huber, J. (1993). Hymenoptera of the World: An Identification Guide to Families. Canada Agriculture. Ottawa, Canada.
- Graham, C. F., Glenn, T. C., McArthur, A. G., Boreham, D. R., Kieran, T., Lance, S., ... Somers, C. M. (2015). Impacts of degraded DNA on restriction enzyme associated DNA sequencing (RADSeq). *Molecular Ecology Resources*, **15**, 1304–1315.
- Gregory, T. R. (2008). Understanding Evolutionary Trees. *Evolution: Education and Outreach*, **1**, 121–137.
- Gutiérrez-García, T. A. & Vázquez-Domínguez, E. (2013). Consensus between genes and stones in the biogeographic and evolutionary history of Central America. *Quaternary Research (United States)*, **79**, 311–324.
- Hamerlinck, G., Hulbert, D., Hood, G. R., Smith, J. J. & Forbes, A. A. (2016). Histories of host shifts and cospeciation among free-living parasitoids of Rhagoletis flies. *Journal of Evolutionary Biology*, **29**, 1766–1779.
- Hanson, P. E. & Gauld, I. D. (2006). *Hymenoptera de la Región Neotropical. Memoirs of the American Entomological Institute* (Vol. 77). Florida, USA.
- Harvey, M. G., Smith, B. T., Glenn, T. C., Faircloth, B. C. & Brumfield, R. T. (2016). Sequence Capture versus Restriction Site Associated DNA Sequencing for Shallow Systematics. *Systematic Biology*, **65**, 910–924.
- Hebert, P. D. N., Penton, E. H., Burns, J. M., Janzen, D. H. & Hallwachs, W. (2004). Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proceedings of the National Academy of Sciences*, **101**, 14812–14817.
- Hennig, W. (1966). Phylogenetic systematics. University of Illinois Press, Urbana.
- Hey, J. & Pinho, C. (2012). Population genetics and objectivity in species diagnosis. *Evolution*, **66**, 1413–1429.
- Hime, P. M., Hotaling, S., Grewelle, R. E., O'Neill, E. M., Voss, S. R., Shaffer, H. B. & Weisrock, D. W. (2016). The influence of locus number and information content on species delimitation: an empirical test case in an endangered Mexican salamander. *Molecular Ecology*, **25**, 5959–5974.
- Hoffberg, S. L., Kieran, T. J., Catchen, J. M., Devault, A., Faircloth, B. C., Mauricio, R. & Glenn, T. C. (2016). RADcap: sequence capture of dual-digest RADseq libraries with identifiable duplicates and reduced missing data. *Molecular Ecology Resources*, **16**, 1264–1278.
- Hull, D. L. (1980). Individuality and selection. *Annual Review of Ecology and Systematics*, **11**, 311–332.

- Jeffries, D. L., Copp, G. H., Lawson Handley, L.-J., Olsén, H., Sayer, C. D. & Hänfling, B. (2015). Comparing RADseq and microsatellites to infer complex phylogeographic patterns, a real data informed perspective in the Crucian carp, *Carassius carassius*, L. *bioRxiv*, 25973.
- Jones, O. R., Purvis, A., Baumgart, E. & Quicke, D. L. J. (2009). Using taxonomic revision data to estimate the geographic and taxonomic distribution of undescribed species richness in the Braconidae (Hymenoptera: Ichneumonoidea). *Insect Conservation and Diversity*, **2**, 204–212.
- Kergoat, G. J., Delobel, A. & Silvain, J. F. (2004). Phylogeny and host-specificity of European seed beetles (Coleoptera, Bruchidae), new insights from molecular and ecological data. *Molecular Phylogenetics and Evolution*, **32**, 855–865.
- Leaché, A. D., Chavez, A. S., Jones, L. N., Grummer, J. A., Gottscho, A. D. & Linkem, C. W. (2015). Phylogenomics of phrynosomatid lizards: Conflicting signals from sequence capture versus restriction site associated DNA sequencing. *Genome Biology and Evolution*, **7**, 706–719.
- Leaché, A. D. & Rannala, B. (2011). The accuracy of species tree estimation under simulation: A comparison of methods. *Systematic Biology*, **60**, 126–137.
- Lemmon, A. R., Emme, S. A. & Lemmon, E. M. (2012). Anchored hybrid enrichment for massively high-throughput phylogenomics. *Systematic Biology*, **61**, 727–744.
- Lozier, J. D., Roderick, G. K. & Mills, N. J. (2009). Molecular markers reveal strong geographic, but not host associated, genetic differentiation in *Aphidius transcaspicus*, a parasitoid of the aphid genus *Hyalopterus*. *Bulletin of Entomological Research*, **99**, 83–96.
- Maddison, W. P. (1997). Gene Trees in Species Trees. *Systematic Biology*, **46**, 523–536.
- Mayden, R. L. (1997). A hierarchy of species concepts: The denouement in the saga of the species problem. Pages 381–424 in *Species: The units of biodiversity* (M. F. Claridge, H. A. Dawah, and M. R. Wilson, eds.). Chapman and Hall, London.
- Mayr, E. (1942). *Systematics and the origin of species*. Columbia University Press, New York.
- Mayr, E. (1963). *Animal species and evolution*. Belknap Press of Harvard University Press, Cambridge, Massachusetts.
- Mccormack, J. E., Faircloth, B. C., Crawford, N. G., Gowaty, P. A., Brumfield, R. T. & Glenn, T. C. (2012). Ultraconserved Elements Are Novel Phylogenomic Markers that Resolve Placental Mammal Phylogeny when Combined with Species Tree Analysis. *Systematic Biology*, **61**, 746–754.
- Mendoza-Ramírez, M., Gutiérrez-Rodríguez, J., Poteaux, C., Ornelas-García, P. & Zaldívar-Riverón, A. (2018). Late Pleistocene genetic diversification and demographic expansion in the widely distributed neotropical ant *Neoponera villosa* (Ponerinae). *Mitochondrial DNA Part A*, **0**, 1–11.

- Morrone, J. J. (2015). Halffter's Mexican transition zone (1962-2014), cenocrons and evolutionary biogeography. *Journal of Zoological Systematics and Evolutionary Research*, **53**, 249–257.
- Nieves-Aldrey, J. L., Fontal-Cazalla, F. . & Fernández, F. (2006). Filogenia y evolución de Hymenoptera. In F. Fernández & M. J. Sharkey (Eds.), *Introducción a los Hymenoptera de la Región Neotropical* (p. 891pp). Bogotá: Sociedad Colombiana de Entomología y Universidad Nacional de Colombia.
- Oliver, K. M., Russell, J. A., Moran, N. A. & Hunter, M. S. (2003). Facultative bacterial symbionts in aphids confer resistance to parasitic wasps. *Proceedings of the National Academy of Sciences*, **100**, 1803–1807.
- Ornelas, J. F., Sosa, V., Soltis, D. E., Daza, J. M., González, C., Soltis, P. S., ... Ruiz-Sanchez, E. (2013). Comparative Phylogeographic Analyses Illustrate the Complex Evolutionary History of Threatened Cloud Forests of Northern Mesoamerica. *PLoS ONE*, **8**.
- Paris, J. R., Stevens, J. R. & Catchen, J. M. (2017). Lost in parameter space: a road map for stacks. *Methods in Ecology and Evolution*, **8**, 1360–1373.
- Peixoto, L., Allen, G. R., Ridenbaugh, R. D., Quarrell, S. R., Withers, T. M. & Sharanowski, B. J. (2018). When taxonomy and biological control researchers unite: Species delimitation of *Eadya* parasitoids (Braconidae) and consequences for classical biological control of invasive paropsine pests of Eucalyptus. *PLoS ONE*, **13**, 1–28.
- Pérez, G. & Bonet, A. (1984). Algunas características biológicas de *Stenocorse bruchivora* (Crawford) (Hymenoptera: Braconidae), ectoparasitoide de *Acanthoscelides obtectus* (Say) (Coleoptera: Bruchidae). *Folia Entomológica Mexicana*, **62**, 59–74.
- Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S. & Hoekstra, H. E. (2012). Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS ONE*, **7**.
- Porrás-Hurtado, L., Ruiz, Y., Santos, C., Phillips, C., Carracedo, J. & Lareu, M. V. (2013). An overview of STRUCTURE: Applications, parameter settings, and supporting software. *Frontiers in Genetics*, **4**, 1–13.
- Pritchard, J. K., Wen, X. & Falush, D. (2010). Documentation for structure software: Documentation for structure software: Version 2 . 3. *University of Chicago & University of Oxford*.
- Prugnolle, F. & Meeus, T. de. (2002). Inferring sex-biased dispersal from population genetic tools: a review. *Heredity*, **88**, 161–165.
- Rodríguez, S. A. (2018). Semiochemistry of the Bruchidae species. *Journal of Stored Products Research*, **77**, 148–155.
- Romero Gómez, G., Romero Nápoles, J., Yus Ramos, R., Burgos Solorio, A., Valdez Carrasco, J. & Flores Morales, A. (2009). Gorgojos De La Familia Bruchidae (Coleoptera)

- Asociados a Semillas De Plantas Silvestres Destinadas Para Germoplasma. *Boletín Sociedad Entomológica Aragonesa*, **44**, 333–342.
- Samacá-Sáenz, E., Belokobylskij, S. A., Quicke, D. L. J. & Zaldívar-Riverón, A. (2016). Systematics of the Neotropical braconid wasps of the *Pedinotus* genus group (Doryctinae). *Systematic Entomology*, **41**, 481–491.
- Sharkey, M. . & Wahl, D. B. (1992). Cladistics of the Ichneumonoidea (Hymenoptera). *Journal of the Hymenoptera Research*, **1**.
- Sharkey, M. J., Carpenter, J. M., Vilhelmsen, L., Herarty, J., John, H., Liljebblad, J., ... Wheeler, W. C. (2012). Phylogenetic relationships among superfamilies of Hymenoptera. *Cladistics*, **28**, 80–112.
- Simpson, G. G. (1961). Principles of animal taxonomy. Columbia University Press, New York.
- Singhal, S., Hoskin, C. J., Couper, P., Potter, S. & Moritz, C. (2018). A Framework for Resolving Cryptic Species: A Case Study from the Lizards of the Australian Wet Tropics. *Systematic Biology*, **67**, 1061–1075.
- VanValen, L. (1976). Ecological species, multispecies, and oaks. *Taxon*, **25**, 233–239.
- West-Eberhard, M. J. (2005). Developmental plasticity and the origin of species differences. *PNAS*, **102**, 6543–6549.
- Yang, Z. (2015). The BPP program for species tree estimation and species delimitation. *Current Zoology*, **61**, 854–865.
- Yang, Z. & Rannala, B. (2010). Bayesian species delimitation using multilocus sequence data. *PNAS*, **107**, 9264–9269.
- Zaldívar-Riverón, A., Belokobylskij, S. A., León-Regagnon, V., Briceño-G., R. & Quicke, D. L. J. (2008). Molecular phylogeny and historical biogeography of the cosmopolitan parasitic wasp subfamily Doryctinae (Hymenoptera: Braconidae). *Invertebrate Systematics*, **22**, 345–363.
- Zaldívar-Riverón, A., Belokobylskij, S. A., Meza-Lázaro, R., Pedraza-Lara, C., García-París, M. & Meseguer, A. S. (2018). Species delimitation, global phylogeny and historical biogeography of the parasitoid wasp genus *spathius* (braconidae: Doryctinae) reveal multiple oligocene?miocene intercontinental dispersal events. *Zoological Journal of the Linnean Society*, **182**, 723–734.
- Zaldívar-Riverón, A., León-Regagnon, V. & Nieto-Montes De Oca, A. (2004). Phylogeny of the Mexican coastal leopard frogs of the *Rana berlandieri* group based on mtDNA sequences. *Molecular Phylogenetics and Evolution*, **30**, 38–49.