UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

POSGRADO EN CIENCIAS BIOLÓGICAS

FACULTAD DE CIENCIAS

SISTEMÁTICA

BIOGEOGRAFÍA EVOLUTIVA Y ECOLÓGICA DE LOS VECTORES

ASOCIADOS CON LA ENFERMEDAD DE CHAGAS

TESIS

QUE PARA OPTAR POR EL GRADO DE:

DOCTORA EN CIENCIAS

PRESENTA:

M. en C. LAURA ALEXANDRA RENGIFO CORREA

TUTOR PRINCIPAL DE TESIS:	**DR. JUAN JOSÉ MORRONE LUPI**
	FACULTAD DE CIENCIAS, UNAM
COMITÉ TUTOR:	**DRA. PAZ MARÍA SILVIA SALAZAR SCHETTINO**
	FACULTAD DE MEDICINA, UNAM
	DR. CHRISTOPHER RHODES STEPHENS STEVENS
	INSTITUTO DE CIENCIAS NUCLEARES, UNAM
TUTOR INVITADO:	**DR. FERNANDO ABAD FRANCH**
	INSTITUTO RENÉ RACHOU, FIOCRUZ

CD. MX.   **JUNIO, 2018**

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

POSGRADO EN CIENCIAS BIOLÓGICAS
FACULTAD DE CIENCIAS
SISTEMÁTICA

BIOGEOGRAFÍA EVOLUTIVA Y ECOLÓGICA DE LOS VECTORES

ASOCIADOS CON LA ENFERMEDAD DE CHAGAS

TESIS

QUE PARA OPTAR POR EL GRADO DE:

DOCTORA EN CIENCIAS

PRESENTA:

M. en C. LAURA ALEXANDRA RENGIFO CORREA

TUTOR PRINCIPAL DE TESIS: **DR. JUAN JOSÉ MORRONE LUPI**
                                FACULTAD DE CIENCIAS, UNAM
COMITÉ TUTOR: **DRA. PAZ MARÍA SILVIA SALAZAR SCHETTINO**
                                FACULTAD DE MEDICINA, UNAM
                                DR. CHRISTOPHER RHODES STEPHENS STEVENS
                                INSTITUTO DE CIENCIAS NUCLEARES, UNAM
TUTOR INVITADO: **DR. FERNANDO ABAD FRANCH**
                                INSTITUTO RENÉ RACHOU, FIOCRUZ

MÉXICO, CD. MX.   **JUNIO, 2018**

UNAM
POSGRADO
Ciencias Biológicas

POSGRADO EN CIENCIAS BIOLÓGICAS
FACULTAD DE CIENCIAS
DIVISIÓN ACADÉMICA DE INVESTIGACIÓN Y POSGRADO

OFICIO FCIE/DAIP/493/2018

ASUNTO: Oficio de Jurado

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

Me permito informar a usted que en la reunión ordinaria del Comité Académico del Posgrado en Ciencias Biológicas, celebrada el día **16 de abril de 2018**, se aprobó el siguiente jurado para el examen de grado de **DOCTORA EN CIENCIAS** del (la) alumno (a) **RENGIFO CORREA LAURA ALEXANDRA** con número de cuenta 512451084 con la tesis titulada: "**Biogeografía evolutiva y ecológica de los vectores asociados con la enfermedad de Chagas**", realizada bajo la dirección del (la) **DR. JUAN JOSÉ MORRONE LUPI**:

Presidente:	**DRA. LIVIA SOCORRO LEÓN PANIAGUA**
Vocal:	**DR. GERARDO SUZÁN AZPIRI**
Secretario:	**DRA. PAZ MARÍA SILVIA SALAZAR SCHETTINO**
Suplente:	**DR. ADOLFO GERARDO NAVARRO SIGÜENZA**
Suplente:	**DR. CHRISTOPHER RHODES STEPHENS STEVENS**

Sin otro particular me es grato enviarle un cordial saludo.

A T E N T A M E N T E
"POR MI RAZA HABLARA EL ESPIRITU"
Ciudad Universitaria, Cd. Mx., a 17 de mayo de 2018

DR. ADOLFO GERARDO NAVARRO SIGÜENZA
COORDINADOR DEL PROGRAMA

AGNS/VMVA/ASR/mnm

AGRADECIMIENTOS

AGRADECIMIENTOS A TITULO PERSONAL

A mi tutor, Juan José Morrone. Gracias a ti las puertas de la UNAM y de México se abrieron. Gracias a tu trato respetuoso, ético y humano pude persistir en mis estudios. Gracias a tus consejos pude elegir un área de investigación que me ayudara a explotar diferentes aptitudes. Gracias por heredarme la proporción de escepticismo/ mentalidad abierta adecuada para desarrollar las habilidades de crítica en innovación tan necesarias en el quehacer científico. Gracias por tu sinceridad amable con la que me corregiste y que derivó en el trabajo que ahora se materializa.

A los miembros de mi comité tutor por aceptar acompañarme y dirigirme durante mis estudios de doctorado, por toda la formación, apoyo académico y paciencia brindados, por compartir su experiencia e intereses académicos y por el apoyo logístico. Gracias a este comité, esta tesis tomó su forma. Además, agradezco a Fernando Abad Franch y a Paz María Salazar Schettino, por introducirme al mundo de los triatominos y la enfermedad de Chagas; a Christopher Stephens y a Fernando Abad Franch por introducirme al mundo de los modelos desde una perspectiva crítica, y a Fernando Abad Franch por sus consejos iniciales de reducir la dimensión de la tesis.

Al jurado de mi tesis de doctorado, por sus valiosos comentarios para mejorar la redacción final de la tesis: Adolfo Navarro, Christopher Stephens, Gerardo Suzán, Livia León, Paz María Salazar.

A los sinodales de mi candidatura, por las críticas constructivas para mejorar el proyecto de doctorado: Alejandro Oceguera, Christopher Stephens, Livia León, Tania Escalante.

A Constantino González Salazar por su asesoría para desarrollar las redes de interacción biótica y por todo el apoyo y los consejos brindados.

A Fernando Martínez Hernández y a Guiehdani Villalobos por introducirme al mundo estocástico de las PCRs y por el apoyo logístico brindado.

A Herón Huerta, a Juan Luis Téllez Rendón y al personal de vigilancia del Instituto de Diagnóstico y Referencia Epidemiológicos (InDRE), por proporcionar los datos de ocurrencia de los triatominos del complejo *Phyllosoma* con los cuales desarrollé los modelos de redes y nichos ecológicos.

DEDICATORIA

A mi mamá, Lilia Correa Carvajal.

Toda una vida de amor y entrega jamás será compensada.

Pero que una gota de alivio sea para ti saber que

tan lejos como lleguen tus semillas tú habrás llegado.

A mi cónyuge, Ángel Gabriel Becerra Vázquez.

De cerca acompañas mis triunfos y mis caídas,

sólo tú conoces el esfuerzo detrás de estas páginas.

"Todo cambia con prisa endemoniada;

cada diez años estos reinos tienen un rostro distinto…

Pero uno sólo ve con nitidez lo que dura:

un mundo que no cesa de cambiar

apenas si produce en los ojos el efecto de un viento."

William Ospina.

"All models are wrong,

but some are useful."

George Box.

ÍNDICE

LISTA DE FIGURAS Y CUADROS

CAPÍTULO 2.

RESUMEN

La enfermedad de Chagas es una de las enfermedades transmitidas por vectores más importantes en Latinoamérica. El agente etiológico de la enfermedad de Chagas es el protozoo *Trypanosoma cruzi*. Este endoparásito desarrolla todo su ciclo de vida en sus huéspedes mamíferos y vectores invertebrados, las chinches de campo o triatominos (Hemiptera: Reduviidae: Triatominae). Para la transmisión del parásito es obligatoria la interacción entre sus huéspedes y vectores. Por lo tanto, las consecuencias ecológicas y evolutivas de dichas interacciones deberían ser estudiadas integralmente. En esta tesis se infirieron algunas consecuencias ecológicas y evolutivas de las interacciones bióticas entre huéspedes y vectores de *T. cruzi* desde el enfoque hipotético – deductivo. Para dicha inferencia se tomó ventaja de la biogeografía ecológica y la evolutiva, la información empírica sobre las interacciones entre pares de especies y la información filogenética. Así, en el capítulo 1 de la tesis se evidenciaron patrones de transmisión de *T. cruzi* entre nueve vectores y más de 100 especies de mamíferos silvestres en áreas no desérticas de México. El descubrimiento de dichos patrones posibilitará la generación de hipótesis sobre la ecología epidemiológica de la enfermedad de Chagas que, a su vez, podrían refutarse en futuros estudios. Adicionalmente, en el capítulo 2 de la tesis se infirió el papel de la hibridación antigua en la historia evolutiva del complejo *Phyllosoma* (Hemiptera: Reduviidae: Triatominae), vectores de la enfermedad de Chagas en México. La consideración del papel de la introgresión en la historia evolutiva del complejo *Phyllosoma* coadyuvó a la inferencia de sus relaciones filogenéticas y podría favorecer la comprensión del papel de los híbridos en la transmisión de *T. cruzi*. Este trabajo representa una contribución al conocimiento de la biogeografía aplicada a la epidemiología y evolución de la enfermedad de Chagas.

Palabras clave: Transmisión enfermedad de Chagas, interacciones bióticas inferidas, red huésped-vector, hibridación antigua, ecosistema de *Trypanosoma cruzi*, filogenia del complejo *Phyllosoma*.

ABSTRACT

Chagas disease is one of the most important vector-borne diseases in Latin America. The etiological agent of Chagas disease is the protozoan *Trypanosoma cruzi*. This endoparasite develops its life cycle inside mammal hosts and invertebrate vectors, the kissing bugs (Hemiptera: Reduviidae: Triatominae). Hosts and vectors must interact for a successful parasite transmission. Therefore, ecological and evolutionary consequences of hosts-vector interactions should be studied integrally. Here, some ecological and evolutionary consequences of biotic interactions between host-vector of *T. cruzi* were inferred from a hypothetic-deductive approach. For the inference, the ecological and the evolutionary biogeographies, the empiric pair-species interaction data and the phylogenetic data were considered. Thus, transmissibility patterns of *T. cruzi* between nine vectors and more than 100 wild mammal species in non-desert areas of Mexico were showed in chapter 1. Pattern discovering will enable hypotheses formulation about ecological epidemiology of Chagas disease, which would be refutable in future studies. Also, the role of ancient hybridization in the evolutionary histories of *Phyllosoma* complex (Hemiptera: Reduviidae: Triatominae), vectors of Chagas disease in Mexico, was inferred in chapter 2. Accounting of the introgression role in the evolutionary history of *Phyllosoma* complex helped the inference of its phylogenetic relationships and would help to understanding of the role of hybrids in *T. cruzi* transmissibility. This research represents a contribution for the biogeography applied to the epidemiology and evolution of Chagas disease.

Key words: Transmission of Chagas disease, biotic interactions, host-vector network, ancient hybridization, ecosystem of *Trypanosoma cruzi*, phylogeny of *Phyllosoma* complex.

INTRODUCCIÓN

Entre las enfermedades transmitidas por vectores, la enfermedad de Chagas es una de las más importantes en Latinoamérica con seis a siete millones de personas infectadas y 70 millones en riesgo de contraerla (WHO, 2015). En la etapa crónica de esta enfermedad se manifiestan patologías diversas a nivel del corazón, sistema digestivo y sistema nervioso periférico, causando discapacidad y muerte (WHO, 2002). Dado que para la enfermedad de Chagas los tratamientos farmacológicos tienen una eficacia limitada (Clayton, 2010), es prioritario prevenirla a través de programas de control de los vectores (Tarleton et al., 2007). Sin embargo, es difícil controlar a los vectores en áreas donde son nativos porque, mientras los programas de control se dirigen a los focos de infestación en ambientes modificados por el hombre, las poblaciones en ambientes silvestres permanecen fuera del alcance de dichos programas (Abad-Franch, 2016; Abad-Franch et al., 2013; Coura, 2013). Para contrarrestar la dificultad del controlar la enfermedad de Chagas, las estrategias de control podrían reestructurarse considerando sus patrones de transmisibilidad.

La transmisibilidad de la enfermedad de Chagas abarca diferentes aspectos de las interacciones entre el parásito *Trypanosoma cruzi* (el agente etiológico de la enfermedad), sus vectores y huéspedes. En un nivel básico, es esencial identificar cuáles son las especies que participan en la transmisión de *T. cruzi*. *Trypanosoma cruzi* desarrolla su ciclo de vida en huéspedes mamíferos (Fig. 1: A, C) y en vectores invertebrados, los triatominos (Hemiptera: Reduviidae: Triatominae) (Fig. 1: B) (Lent y Wygodzynsky, 1979). A pesar del potencial de *T. cruzi* para infectar y ser transmitido por un amplio rango de mamíferos y de triatominos, no todas estas especies tienen la misma importancia en la transmisión del parásito (Jansen y Roque, 2010). Identificar las principales especies que participan en la transmisión de *T. cruzi* ayudaría a priorizar la investigación consecutiva. En un nivel más complejo, es necesario identificar qué lleva a que *T. cruzi* sea transmitido eficazmente. Por ejemplo, se ha encontrado que la virulencia de *T. cruzi* varía en relación con el origen geográfico (por altitud) de las poblaciones de triatominos (de Fuentes-Vicente et al., 2016). También, se ha encontrado que triatominos de origen híbrido son más vulnerables a la infección de *T.*

cruzi que sus parentales (Herrera-Aguilar et al., 2009), lo que podría derivar en una transmisión diferencial del parásito por parte de los parentales y los híbridos. Por consiguiente, es urgente discernir qué procesos están involucrados en la transmisión eficaz de *T. cruzi*. Dichas preguntas de investigación han sido objeto de investigación por décadas.

Las preguntas sobre la transmisión de *T. cruzi* tradicionalmente se han abordado a través de estudios de interacciones entre pares de especies. Aunque esta aproximación empírica es indispensable, genera una visión fragmentaria y segregada de la transmisión de *T. cruzi*. El conocimiento segregado sobre la transmisión de *T. cruzi* dificultaría su comprensión, con consecuencias negativas como la ralentización en la toma de decisiones para el control de la enfermedad de Chagas. Por esta razón, urge integrar el conocimiento sobre las interacciones entre el parásito, los mamíferos (huéspedes) y los triatominos. Una alternativa al estudio de interacciones entre pares de especies sería el estudio simultáneo de múltiples especies. Este estudio simultáneo permitiría acelerar el descubrimiento de patrones de transmisibilidad de la enfermedad de Chagas (Flores-Ferrer et al., 2017; Georgieva et al., 2017; Jansen et al., 2015, Nouvellet et al., 2013). Sin embargo, rara vez se ha aplicado una aproximación multiespecífica a las interacciones entre los agentes involucrados en la transmisión de la enfermedad de Chagas (Ibarra-Cerdeña et al., 2017; Jansen et al., 2015; Rengifo-Correa et al., 2017). Además, empezar *de novo* con una aproximación multiespecífica resultaría extremadamente costoso en tiempo y recursos. Una forma viable de contrarrestar esta situación sería a través del aprovechamiento de la información recopilada en estudios previos. Por lo tanto, es imperioso transformar el conocimiento sobre las interacciones bióticas entre pares de especies de *T. cruzi*, huéspedes y vectores hacia una perspectiva multiespecífica.

Una forma de transformar el conocimiento sobre las interacciones bióticas entre pares de especies hacia un carácter multi-específico sería a través de la biogeografía. Ésta posibilitaría una integración de dicho conocimiento en dos formas simples: la inductiva y la hipotético-deductiva. Bajo un primer enfoque inductivo, el conocimiento científico se construye a partir de la acumulación de observaciones (Morrone y

Escalante 2016). En este caso, la información *in vitro* puede ser relacionada con la de la biogeografía para conjeturar la plausibilidad de una interacción biótica bajo un determinado contexto espacial (Fig. 2 A). Por ejemplo, si se obtuvo evidencia *in vitro* sobre la capacidad de una especie de triatomino para aprovechar en su dieta a cualquier mamífero que se le ofrezca, esta información cobraría relevancia si las distribuciones espaciales del mamífero y vector evaluados coinciden. Asimismo, si dos especies de triatominos pueden producir una progenie fértil *in vitro*, esta información cobraría relevancia si las distribuciones espaciales de estas especies coinciden. En ambos casos, la plausibilidad de la información *in vitro* en un contexto espacial podría conjeturarse con datos simples, como lo son las distribuciones espaciales de las especies. Para este primer enfoque, el patrón multi-específico surgiría cuando varias interacciones por pares de especies se acoplen bajo un mismo contexto espacial (Fig. 2 A). Bajo un segundo enfoque hipotético-deductivo, el conocimiento científico parte de la formulación de hipótesis, acompañadas de la expectativa del comportamiento del fenómeno estudiado (deducciones), que son juzgadas (falsadas) por las observaciones (Morrone y Escalante 2016). El enfoque hipotético-deductivo también permite la integración del conocimiento sobre las interacciones bióticas con ayuda de la biogeografía (Fig. 2 B). Muchos factores (etológicos, fisiológicos, evolutivos, etc.) están involucrados para que ocurra una interacción biótica (Agosta et al., 2010). Sin embargo, un requisito mínimo para que una interacción proceda sería que las especies interactuantes coincidan en el espacio (Stephens et al., 2017). Este criterio espacial simple permite la generación de hipótesis de interacciones bióticas previamente desconocidas. Para que estas hipótesis se puedan falsar, se puede discernir entre interacciones bióticas potenciales vs. interacciones poco probables al confrontarlas con el conocimiento previo sobre las interacciones bióticas por pares de especies (Stephens et al. 2009). En este segundo enfoque, el patrón multi-específico surgiría desde el inicio, cuando las múltiples especies bajo consideración tenían como factor común único el espacio (Fig. 2 B). En síntesis, el conocimiento sobre las interacciones bióticas entre *T. cruzi*, sus huéspedes y vectores adquiriría un carácter multi-específico si: (1) las interacciones entre pares de especies, conocidas empíricamente, se acoplan

bajo un mismo contexto espacial y (2) las interacciones hipotéticas entre múltiples especies, que surgen de criterios espaciales, se confrontan con la información empírica.

En esta tesis, se infieren las interacciones entre varias especies de huéspedes y vectores de *T. cruzi* con ayuda de la biogeografía evolutiva y la ecológica. La biogeografía evolutiva permite reconocer patrones de integración de las biotas en el espacio y el tiempo (Morrone, 2009). Es decir, con ayuda de la biogeografía evolutiva no solo se pueden rescatar los patrones de coincidencia espacial de las especies, sino que también se busca rescatar los contextos temporales de dichos patrones. La biogeografía ecológica permite reconocer y cuantificar la importancia relativa de los factores que explican las distribuciones observadas de las especies e integrar estos factores en un modelo que represente espacialmente dichas distribuciones (Stephens et al., 2009). Por lo tanto, con ayuda de la biogeografía ecológica se pueden estimar las distribuciones de las especies en los espacios geográfico y ambiental (Peterson y Soberón, 2012). Aquí, la biogeografía evolutiva y la ecológica posibilitan: (1) la detección de patrones de asociación espacial entre las especies estudiadas bajo un determinado contexto espacio-temporal y (2) la cuantificación de la importancia relativa de dichas asociaciones. De ahí que, con ayuda de la biogeografía, los patrones de asociación espacio-temporal con significancia estadística son la base para inferir interacciones bióticas entre huéspedes y vectores de *T. cruzi* (Fig. 3).

Esta tesis también se enfoca en interacciones bióticas entre huéspedes y vectores de *T. cruzi* de mayor importancia epidemiológica en México. En México habitan 100 millones de personas, de las cuales alrededor del 1% padece la enfermedad de Chagas (WHO, 2006, 2015). La transmisión de la enfermedad de Chagas ocurre activamente en México, considerando que hay reportes de la enfermedad en población menor de 18 años (Salazar-Schettino et al., 2016). Además, se estima que cada año más de 7000 personas adquieren el parásito *T. cruzi* mediante los vectores (WHO, 2006). Los vectores de la enfermedad de Chagas se han registrado en diversas localidades de la República Mexicana, principalmente en el centro y sur del país (Ramsey et al., 2016). El papel destacado de algunos vectores de *T. cruzi* se

adjudica, entre otras cosas, a su relación estrecha con ambientes modificados por el hombre (Salazar-Schettino et al., 2010). Por ejemplo, algunos vectores pueden desarrollar su ciclo de vida completo (*Triatoma barberi* Usinger y *T. dimidiata* (Latreille)) o en parte (*Triatoma mazzottii* Usinger, *T. pallidipennis* (Stål) y *T. rubida* (Uhler)) dentro de domicilios humanos (Salazar-Schettino et al., 2010). Asimismo, otros vectores se registran con frecuencia en la periferia de domicilios (*T.* sp. *aff. dimidiata*, *T. longipennis* Usinger, *T. gerstaeckeri* (Stål)*, T. mexicana* (Herrich-Schaeffer), *T. phyllosoma* (Burmeister)*, T. picturata* Usinger) (Salazar-Schettino et al., 2010). De esta manera, los vectores funcionan para *T. cruzi* como un puente entre ambientes, porque transmiten el parásito en ambientes modificados (cuando defecan en humanos) luego de adquirirlo en ambientes silvestres (cuando se alimentan de sangre de mamíferos infectados; Lent y Wygodzynsky, 1979). Finalmente, más de 100 especies de mamíferos distribuidas en ocho ordenes se han confirmado como huéspedes de *T. cruzi* en el continente americano (Jansen y Roque, 2010). Esta gran diversidad taxonómica de huéspedes constituye un indicativo del potencial de *T. cruzi* para infectar aún más especies de mamíferos que las descritas. Identificar el papel de los mamíferos como huéspedes potenciales de *T. cruzi* constituye un reto en países como México por su biodiversidad. Por ejemplo, más de 300 especies de mamíferos se han registrado en ambientes silvestres de áreas no desérticas de México (Ceballos y Arroyo, 2012; Ramírez-Pulido et al. 2014). Considerando la transmisión activa de *T. cruzi* en las últimas décadas, la repercusión de la transmisión a una escala geográfica y el amplio número de vectores de *T. cruzi* y la diversidad de mamíferos registrados para México, este territorio se considera inicialmente como contexto espacial de este estudio.

En cuanto al contexto temporal de esta tesis, varía para cada uno de los temas tratados. En la primera parte de esta tesis se aborda la transmisibilidad de *T. cruzi* entre mamíferos de ambientes silvestres (huéspedes potenciales de *T. cruzi*) y los triatominos (vectores de *T. cruzi*) en México, bajo un contexto temporal reciente. En este caso, el contexto temporal está definido por las fechas de colecta asociadas a los registros de mamíferos y triatominos, es decir, en los últimos 60 años. Por otro lado, considerando que el origen híbrido de los vectores es un factor que podría favorecer la

transmisión eficaz de *T. cruzi*, la segunda parte de esta tesis investiga si la hibridación ha sido un proceso común en la historia evolutiva de los principales vectores de la enfermedad de Chagas en México. En este caso, se considera como contexto temporal el periodo comprendido entre el Último Máximo Glacial hasta la actualidad. Este contexto temporal fue seleccionado considerando un periodo posterior al de la especiación de las especies de interés y la disponibilidad de variables ambientales. A continuación, se mencionan más detalles de los temas tratados en los dos capítulos de esta tesis.

En el capítulo 1 de la tesis los objetivos son: (1) identificar las principales especies que participan en la transmisión de *T. cruzi* en México y su importancia relativa e (2) identificar patrones de transmisión de *T. cruzi* desde una perspectiva ecosistémica. Para cumplir con estos objetivos, se estudian nueve especies de triatominos reconocidas por su importancia como vectores en México: *Triatoma barberi*, *T. dimidiata* -*T. dimidiata* 2-, *T.* sp. *aff. dimidiata* -*T. dimidiata* 3-, *T. longipennis*, *T. mazzotii*, *T. mexicana*, *T. pallidipennis*, *T. phyllosoma* y *T. picturata* (Ramsey et al,. 2015). Las ecorregiones (Olson et al., 2001) con registros de estas especies se consideran como área de estudio. Las interacciones potenciales entre los vectores de *T. cruzi* y todas las especies de mamíferos registradas en el área de estudio se infieren mediante el análisis de redes complejas (Stephens et al., 2009). Básicamente, este análisis comprende cuatro pasos: (1) la cuantificación de la fortaleza de la asociación espacial entre pares de especies (triatomino - mamífero); (2) la jerarquización de las asociaciones espaciales de acuerdo con su significancia estadística; (3) construcción de una red compleja con todas las especies cuyas asociaciones espaciales son las más significativas en términos estadísticos, siendo la red una hipótesis de interacción biótica entre múltiples especies, y (4) la falsación de la red compleja, mediante el contraste con información empírica de interacciones bióticas por pares de especies. En este caso, la información empírica son los registros de especies de mamíferos infectadas con el parásito *T. cruzi* en México, considerando que la principal vía de transmisión del parásito es a través de sus vectores. Los registros de mamíferos infectados permiten evaluar la hipótesis sobre si la carencia de aleatoriedad en los

patrones de asociación espacial puede ser interpretada como una medida de la importancia relativa de las interacciones bióticas entre las especies de mamíferos y triatominos. Finalmente, se detectan los patrones de transmisibilidad de *T. cruzi* desde una perspectiva ecosistémica gracias a la red de interacciones potenciales entre vectores y huéspedes de *T. cruzi*.

En el capítulo 2 de la tesis se estudian las interacciones potenciales entre algunos vectores de la enfermedad de Chagas en México (complejo *Phyllosoma*), específicamente, la posibilidad de hibridación antigua y sus consecuencias evolutivas. El complejo *Phyllosoma sensu* Lent y Wygodzynsky comprende a *T. bassolsae*, *T. longipennis*, *T. mazzottii*, *T. pallidipennis*, *T. phyllosoma* y *T. picturata*. Así, los objetivos de este capítulo son: (1) detectar huellas de hibridación antigua en las especies del complejo *Phyllosoma*, mediante un algoritmo que integre los contextos espacio-temporal, ecológico y filogenético, y (2) proporcionar una hipótesis filogenética del complejo *Phyllosoma* que evidencie su historia evolutiva reticulada, mediante un análisis de coalescencia entre múltiples especies. La hibridación antigua puede tener consecuencias evolutivas como la especiación y la introgresión (Baack y Rieseberg, 2007). La introgresión es la integración estable de material genético de una especie en otra mediante cruzamientos reiterados (Baack y Rieseberg, 2007). Una forma de detectar la introgresión es a través de la búsqueda de incongruencias entre árboles filogenéticos de genes con diferentes formas de herencia (Funk y Omland, 2003). Sin embargo, este patrón de incongruencia también puede surgir por otros procesos diferentes a la hibridación, por ejemplo la coalescencia profunda. La coalescencia es el punto del pasado en donde dos alelos convergen en un único ancestro común, mientras que la coalescencia profunda ocurre cuando la coalescencia de dos linajes de genes ocurre mucho más atrás en el tiempo que la divergencia de las especies que albergan estos linajes (Baum y Smith, 2013). El papel de la hibridación en la incongruencia de árboles de genes es discernible del de la coalescencia profunda gracias a la incorporación del contexto espacio-temporal (Funk y Omland, 2003). Más aún, la información ecológica ayuda a la refutabilidad del papel de la hibridación como proceso causante de árboles de genes incongruentes.

Por estas razones, en el capítulo 2 de esta tesis se emplea una aproximación que involucra los contextos espacio-temporal, ecológico y filogenético para detectar introgresión en los especímenes estudiados como consecuencia de hibridación antigua. Básicamente, esta aproximación comprende cuatro pasos: (1) postulación de hipótesis de especies que potencialmente hibridizaron, considerando sus distribuciones y la significancia estadística de su asociación espacial en un contexto espacio-temporal dado; (2) falsación de las hipótesis con datos empíricos de cruzas interespecíficas, o contexto ecológico reproductivo; (3) descubrimiento de incongruencias entre árboles de genes, o contexto filogenético; y (4) detección de muestras que potencialmente presentan introgresión antigua, mediante una correlación entre las muestras que ocasionan incongruencias entre árboles de genes con sus contextos espacio-temporal y ecológico reproductivo. Las muestras que carecen de introgresión se reanalizan bajo un análisis de coalescencia entre múltiples especies para inferir el árbol de especies, o hipótesis filogenética. El pasado evolutivo reticulado de los triatominos del complejo *Phyllosoma* se representa en dicha hipótesis filogenética. De esta manera, el capítulo 2 de la tesis presenta las consecuencias evolutivas de la hibridación antigua entre los principales vectores de la enfermedad de Chagas.

En síntesis, en esta tesis se infieren las consecuencias ecológicas y evolutivas de las interacciones bióticas entre huéspedes (mamíferos) y vectores (triatominos) de *T. cruzi* desde el enfoque hipotético – deductivo. Para dicha inferencia se toma ventaja de la biogeografía ecológica y la evolutiva, la información empírica sobre las interacciones entre pares de especies y la información filogenética. Por una parte, se muestran los patrones de transmisibilidad de *T. cruzi* desde una perspectiva ecosistémica gracias a la red de interacciones potenciales entre vectores y huéspedes de *T. cruzi* en México. Por otra parte, se muestran las consecuencias evolutivas de la hibridación antigua entre algunos vectores de la enfermedad de Chagas en México. Se espera que este trabajo contribuya al conocimiento de la biogeografía aplicada a la epidemiología y evolución de la enfermedad de Chagas.
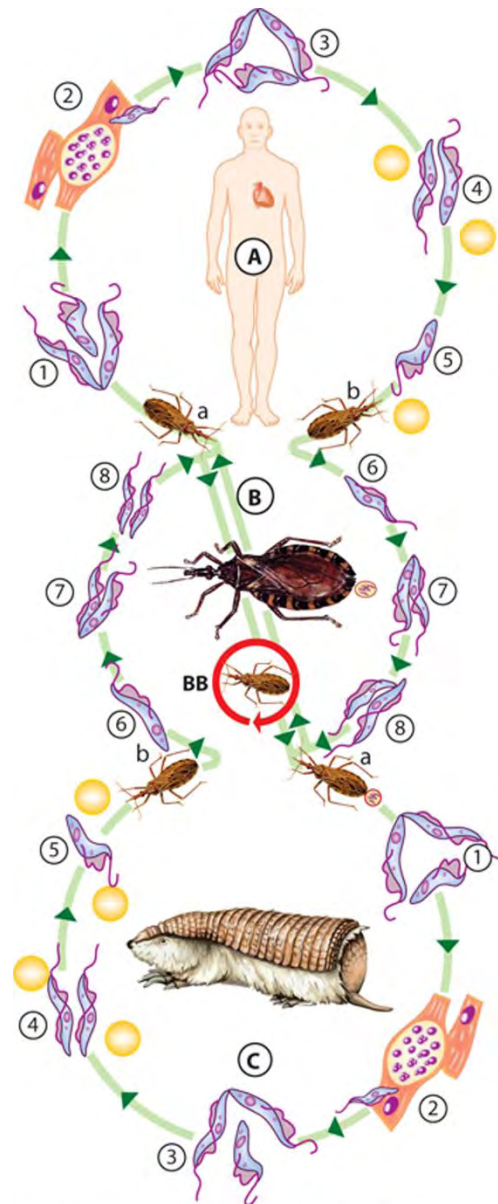
Figura 1. Ciclo de vida del parásito *Trypanosoma cruzi*, en mamíferos (A, C) y en sus vectores invertebrados, los triatominos (B).
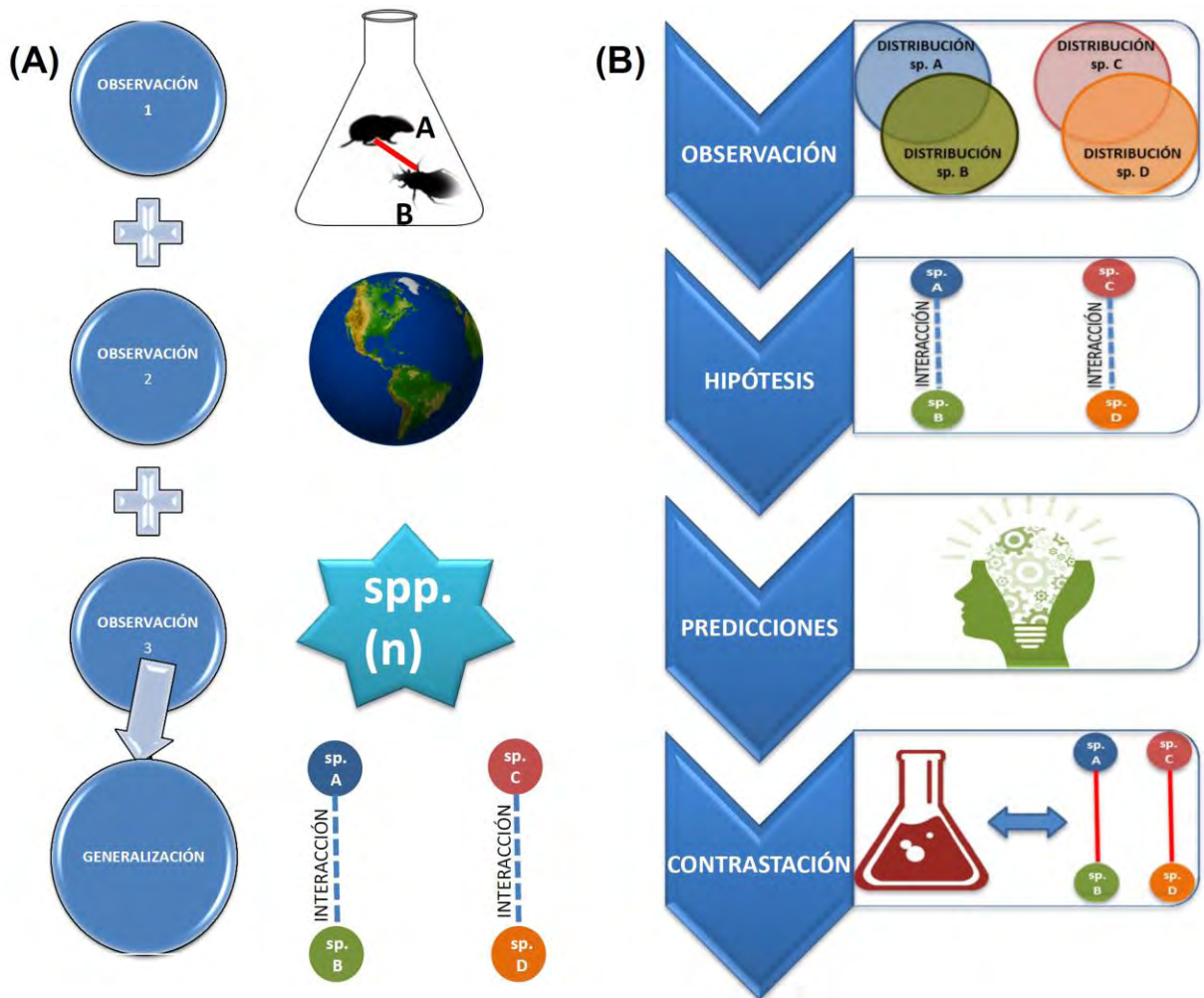
Figura 2. Transformación del conocimiento sobre las interacciones bióticas desde pares de especies hacia múltiples especies a través de la biogeografía. (A) Enfoque inductivo. La información *in vitro* relacionada con la de la biogeografía permite conjeturar la plausibilidad de una interacción biótica bajo un determinado contexto espacial. El patrón multi-específico surge al acoplar varias interacciones por pares de especies bajo un mismo contexto espacial. (B) Enfoque hipotético - deductivo. El patrón multi-específico surge como hipótesis cuando se consideran múltiples especies bajo un mismo contexto espacial. Las hipótesis se puedan falsar con el conocimiento previo sobre las interacciones bióticas por pares de especies
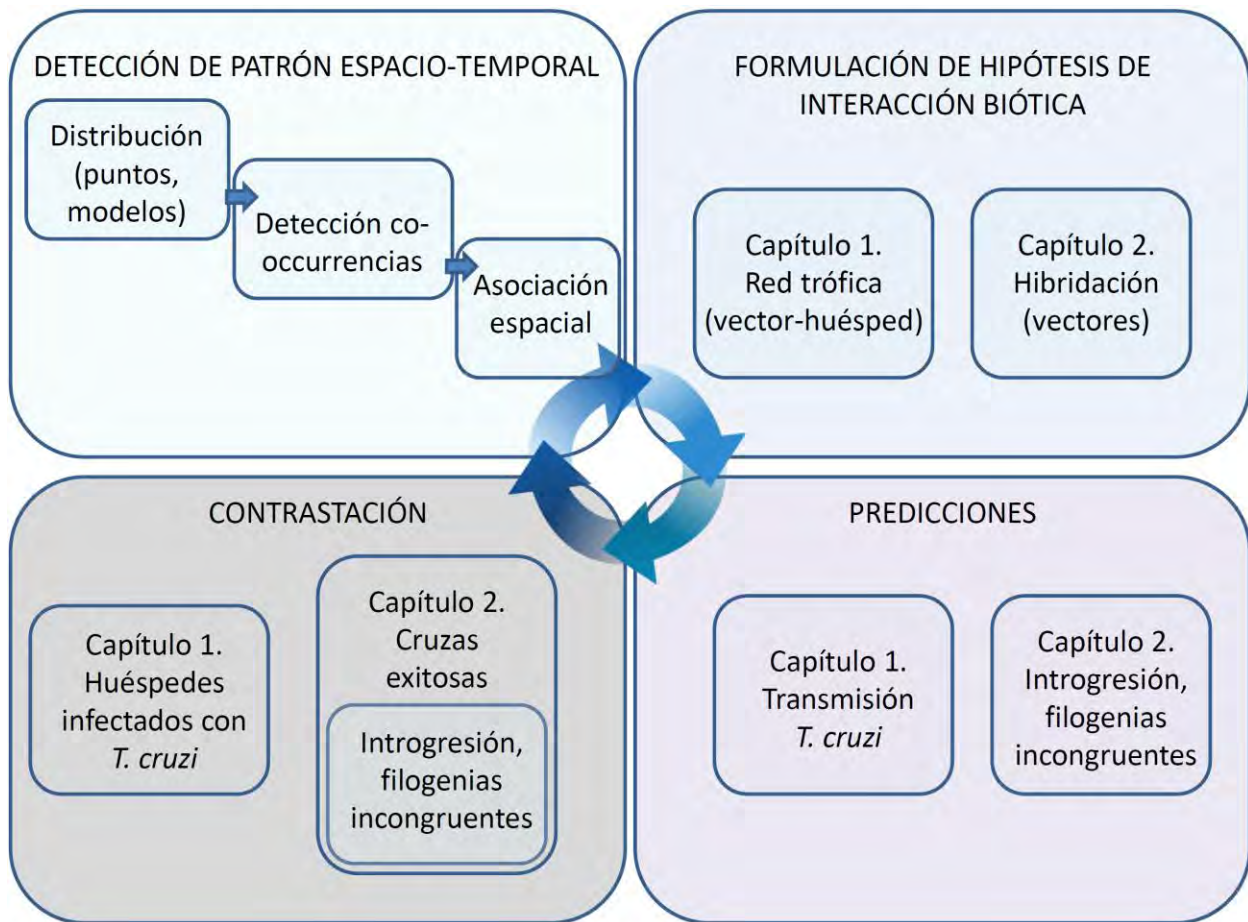
Figura 3. Etapas para inferir las consecuencias ecológicas y evolutivas de las interacciones entre las especies involucradas en la transmisión de *T. cruzi*. Esta aproximación hipotético – deductiva involucra la biogeografía ecológica y la evolutiva, la información empírica sobre las interacciones entre pares de especies y la información filogenética.

CAPÍTULO 1. Understanding transmissibility patterns of Chagas disease through complex vector–host networks

Artículo de requisito publicado en: Parasitology (2017) 144, 760-772.

Understanding transmissibility patterns of Chagas disease through complex vector–host networks

LAURA RENGIFO-CORREA[1], CHRISTOPHER R. STEPHENS[2,3]*, JUAN J. MORRONE[1], JUAN LUIS TÉLLEZ-RENDÓN[4] and CONSTANTINO GONZÁLEZ-SALAZAR[2]

[1] Museo de Zoología 'Alfonso L. Herrera', Departamento de Biología Evolutiva, Facultad de Ciencias, Universidad Nacional Autónoma de México, 04510 Mexico City, Mexico
[2] C3 – Centro de Ciencias de la Complejidad, Universidad Nacional Autónoma de México, 04510 Mexico City, Mexico
[3] Instituto de Ciencias Nucleares, Universidad Nacional Autónoma de México, 04510 Mexico City, Mexico
[4] Instituto de Diagnóstico y Referencia Epidemiológicos (INDRE), 01480 Mexico City, Mexico

SUMMARY

Chagas disease is one of the most important vector-borne zoonotic diseases in Latin America. Control strategies could be improved if transmissibility patterns of its aetiologic agent, *Trypanosoma cruzi*, were better understood. To understand transmissibility patterns of Chagas disease in Mexico, we inferred potential vectors and hosts of *T. cruzi* from geographic distributions of nine species of Triatominae and 396 wild mammal species, respectively. The most probable vectors and hosts of *T. cruzi* were represented in a Complex Inference Network, from which we formulated a predictive model and several associated hypotheses about the ecological epidemiology of Chagas disease. We compiled a list of confirmed mammal hosts to test our hypotheses. Our tests allowed us to predict the most important potential hosts of *T. cruzi* and to validate the model showing that the confirmed hosts were those predicted to be the most important hosts. We were also able to predict differences in the transmissibility of *T. cruzi* among triatomine species from spatial data. We hope our findings help drive efforts for future experimental studies.

Key words: *Trypanosoma cruzi*, potential hosts, spatial data mining, ecological epidemiology.

INTRODUCTION

Chagas disease is one of the most important vector-borne zoonotic diseases in Latin America, with six to seven million people infected and 70 million being at risk of acquiring the disease (WHO, 2015). Infection prevention programs are still the most effective tool for controlling Chagas disease transmission (Rodrigues-Coura, 2013); however, controlling Chagas disease in endemic countries is difficult (Abad-Franch *et al.* 2013; Rodrigues-Coura, 2013). Control strategies could be improved, if the transmissibility patterns of the aetiologic agent, *Trypanosoma cruzi*, were better understood.

To better understand the transmissibility patterns of *T. cruzi*, it is essential to consider the potential interactions among its vectors and hosts at a more integrative level. The parasite has hundreds of potential vectors (Triatominae Reduviidae Hemiptera) and potential hosts (Mammalia) (Jansen and Roque, 2010). Considering this diversity of potential vectors and hosts, it is logistically impossible to understand the transmissibility patterns of *T. cruzi* by an exhaustive experimental examination of all potential triatomine–mammal interactions. In addition, biotic interactions among vectors and hosts

are usually studied at the level of a particular mammal host, rather than considering ecosystemic patterns of the whole host–vector system. Complex Inference Networks (Stephens *et al.* 2009; González-Salazar and Stephens, 2012), applied to zoonoses, provide a useful alternative for understanding the transmissibility patterns of *T. cruzi*. The network represents the most probable ecosystem of *T. cruzi* by showing their potential vectors, hosts and their interactions.

In this paper, we derive a vector–host network for the potential ecological factors – Triatominae and wild mammal species – involved in the transmission cycle of *T. cruzi* in Mexico. Complex Inference Networks in the context of Chagas disease use the statistical significance of co-occurrences of triatomine and wild mammal species as proxies for their potential interactions, where statistically significant co-occurrences are estimated based on the level of overlap of the species' distribution ranges (Stephens *et al.* 2009; González-Salazar and Stephens, 2012). The patterns recovered from the network allow us to formulate and test several hypotheses about the ecological epidemiology of Chagas disease. Primarily, we test the idea that the lack of randomness in the co-occurrence patterns can be interpreted as a measure of the relative level of importance of the biotic interactions between a given mammal and triatomine species. If we identify the most relevant biotic interactions between mammal and triatomine

* Corresponding author: C3 – Centro de Ciencias de la Complejidad, Universidad Nacional Autónoma de México (UNAM), 04510 Mexico City, Mexico. E-mail: stephens@nucleares.unam.mx

species, then, taking the presence of such an interaction as a necessary (but not necessarily sufficient) condition for the transmission of the pathogen, the potential hosts and vectors of Chagas disease can be inferred and transmission patterns of *T. cruzi* can be drawn. In this sense, the vector–host network helps us to detect the transmissibility patterns of *T. cruzi* in an ecosystemic way. We discuss how our findings can be complemented by future experimental studies.

MATERIALS AND METHODS

Study area

To define the study area, records of Triatominae species were projected onto a map of ecoregions (Olson *et al.* 2001). We then considered all ecoregions with at least one species record, which were basically the non-desert ecoregions of the country. Therefore, the study area was all of Mexico, except the Chihuahuan, Sonoran and Baja Californian deserts and the Tamaulipan mezquital (Olson *et al.* 2001).

Species data

We focused on wild mammal species from non-desert ecoregions of Mexico and nine Triatominae species that are known as the main Chagas disease vectors in Mexico (Ramsey *et al.* 2015). We compiled georeferenced localities for Triatominae species according to Lent and Wygodzinsky (1979) and Bargues *et al.* (2008). *Triatoma dimidiata sensu* Lent and Wygodzinsky (1979) is considered a species complex (Bargues *et al.* 2008; Dorn *et al.* 2009; Monteiro *et al.* 2013), so we analysed the two main lineages of the study area separately. The final dataset of 3425 unique point records was obtained from national entomological collections (Instituto de Diagnóstico y Referencia Epidemiológica, InDRE, Mexico City; Colección Nacional de Insectos, CNIN, UNAM, Mexico City) and published records (Ramsey *et al.* 2015). The dataset of mammal species consisted of georeferenced localities for 396 species (Ceballos and Arroyo, 2012). This dataset includes 47 942 unique point records compiled from electronic databases (www.gbif.org, www.conabio.gob.mx).

There is, as has been extensively discussed in the literature, an important question of potential sample bias in such point records; for instance, via the *ad hoc* nature of the collections and the large collecting gaps in space and time (Ponder *et al.* 2001; Graham *et al.* 2004) Such sampling bias constitutes a significant challenge for the success and veracity of analyses of data point records (Yañez-Arenas *et al.* 2014). However, in spite of their potential biases, such databases provide large and important

information resources accumulated over long periods (Ponder *et al.* 2001; Graham *et al.* 2004) for trying to determine the distribution of species as a function of space and time. Thus, it is important to leverage these data, while bearing in mind the impact of such biases. This is even more important for urgent problems of great social impact such as that of emerging diseases.

Inferred interaction network of triatomines and mammals

We adopted a nonparametric spatial data mining approach to infer potential biotic interactions between mammals and triatomine species using the available point collection data. The general modelling methodology of Stephens *et al.* (2009) is based on the idea that biotic interactions can be inferred from the locations of taxa as a function of space and time. Clearly, biotic and ecological interactions in general are very complex, giving rise to spatio-temporal distributions that depend on an enormous number of variables, both biotic and abiotic. However, it is reasonable to suppose that the spatio-temporal distributions of taxa, or other ecological variables, reflect all of the factors and their causal interactions that determine them. The question is: To what extent can the existence of ecological interactions be deduced by an analysis of the positions of taxa? To give a simple example, one would expect competitive interactions to lead to different spatial distributions than mutualistic interactions.

In Stephens *et al.* (2009), the degree of co-occurrence between taxa was taken as an observable measure with which potential interactions could be inferred. Although co-occurrence is not equal to biological interaction, a significantly non-random co-occurrence distribution is a *necessary* condition for a biotic interaction between taxa, and as such it can be used to formulate hypotheses that can be checked experimentally. However, it is clearly not a *sufficient* condition.

Applying the methodology to the present case we observed those co-occurrences between mammals and triatominae in geographical space that are more common than would be expected by chance (Fig. 1). We focused our attention on the spatial dependence of the distributions and ignored the temporal aspect, as the data used are not capable of describing reliably temporal changes. As a measure of statistical association, we consider the probability to find a triatomine given the occurrence of a mammal, $P(T_i|M_j)$, where T_i and M_j represent the presence of the ith triatomine and jth mammal, respectively.

To determine this probability we divide the geographic region of interest into a uniform grid and then count grid cells according to presence of a given triatomine, presence of a given mammal and/
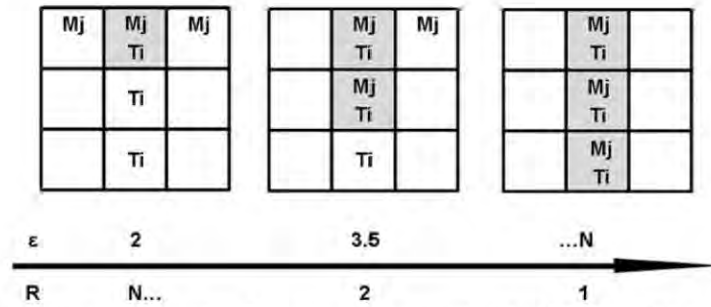
Fig. 1. Co-occurrence pattern of a triatomine (*Ti*) species and a mammal species (*Mj*). Epsilon (*ε*) values increase when the overlap between mammal and triatomine distributions increases. Relative rank (*R*) values are high for species with high *ε* values.

or co-occurrences of both. The choice of grid-cell size has no biological motivation. It is statistically motivated, being associated with maximizing the effective sample size for counting co-occurrences, a co-occurrence essentially being our fundamental statistical unit. The choice of cell size is known in geography as the 'modifiable areal unit problem'. In terms of forming a spatial grid, there are at least two important considerations: the sizes of the statistical samples of the variables and their degree of correlation. Too fine a grid and there will be no co-occurrences, too rough and there will be little to no discrimination. It was checked explicitly in Stephens *et al.* (2009) that the relative ranking of mammals by the model was quite insensitive to the cell size over the range 5–100 km (see also Sierra and Stephens, 2012). However, even though this previous research has shown that predictions of potential feeding resources are robust to large changes in the grid-cell size, we have independently assayed three different grid-cell sizes to check how robust our predictions were (Table S1, Supporting information). Based on these results, for our analysis we used 3535 square grid cells of linear size 20 km, as this resolution has been found to give good overall results when considering a large number of distributions simultaneously (Stephens *et al.* 2009; Sierra and Stephens, 2012).

To evaluate the non-random nature of the co-occurrence distribution we considered the following exact binomial statistical test:

$$\varepsilon(T_i|M_j) = \frac{N_{M_j}(P(T_i|M_j) - P(T_i))}{(N_{M_j}P(T_i)(1 - P(T_i)))^{1/2}} \quad (1)$$

where $P(T_i|M_j) = N_{Ti \text{ and } Mj}/N_{Mj}$ with $N_{Ti \text{ and } Mj}$ being the number of cells where there is a co-occurrence of T_i and M_j, N_{Mj} is the number of cells with presence of M_j and $P(T_i) = N_{Ti}/N$, where N_{Ti} is the number of grid cells with point collections of species T_i and N is the total number of grid cells. This binomial test measures the degree of confidence of the statistical association between T_i and M_j, relative to the null hypothesis that the spatial distribution

of T_i is independent of M_j and distributed randomly over the grid, i.e. $P(T_i)$. The sampling distribution of the null hypothesis is a binomial distribution, where every cell is given a probability $P(T_i)$ of having a point collection of T_i. The numerator of equation (1) is then the difference between the actual number of co-occurrences of T_i and M_j relative to the expected number if the spatial distribution of point collections was obtained from a binomial with sampling probability, $P(T_i)$. The denominator of equation (1) is the standard deviation of the binomial distribution (Stephens *et al.* 2009; González-Salazar *et al.* 2013).

The quantitative values of $\varepsilon(T_i|M_j)$ can be interpreted in the standard sense of hypothesis testing. We consider the *P*-value as the probability that $\varepsilon(T_i|M_j)|$ is at least as large as the observed one and we compare this *P*-value with the required significance level. In the case where $N_{Ti} \geqslant 5$–10, and $P(T_i)$ and $P(T_i|M_j)$ are not close to zero or one, then a normal approximation for the binomial distribution should be adequate, in which case $\varepsilon(T_i|M_j) = 1.96$ would represent the standard 95% confidence interval. Note that such a statistical association does not necessarily prove that there is a direct 'causal' interaction between mammals and vectors. Rather, it allows for a statistical inference to be made or a hypothesis to be formulated that may be validated subsequently (González-Salazar and Stephens, 2012).

We are interested in hypotheses about the transmissibility of the parasite by a specific vector–host interaction. We estimated epsilon, ε, values for a particular mammal species according to its degree of co-occurrence with a given triatomine species. With the values of $\varepsilon(T_i|M_j)$ in hand for every possible triatomine–mammal pair we can compute and visualize a network by considering the nodes of the network to be the mammal and triatomine species and a link between a mammal, M_j, and a triatomine, T_i, to be associated with $\varepsilon(T_i|M_j)$. If all values of $\varepsilon(T_i|M_j)$ are considered, then the network is fully connected. However, if we only draw those links that have a certain degree of statistical significance,

then the network has a different topology, that now represents the principal inferred biotic interactions between mammals and triatominae.

Testing hypotheses of ecological epidemiology of Chagas disease

All else being equal we posit that the higher the value of ε for a host the more epidemiologically important it is in ecological terms. The rational for this is that the greater the degree of spatial overlaps between the distributions, the greater the proportion of potentially infected host individuals due to the higher proportion of individuals that can have a biotic interaction with the vector. Here we do not consider the relative epidemiological importance in terms of human infection. Of course, epidemiological importance, both at the ecological and public health levels is highly complex and multi-factorial involving a host of factors, such as host competence, host/vector abundance, host/vector domiciliation, etc. However, spatial coincidence of vector and host is an absolutely necessary condition on which multiple other factors can and should be included. In the absence of comprehensive, systematic data on these multiple other factors however it is useful to build first-order models based only on occurrence data and use empirical data to test associated hypotheses and models.

To test hypotheses about the ecological epidemiology of Chagas disease, we compiled from the literature a list of confirmed mammal hosts. We searched for mammals species with at least one individual reported as being infected with *T. cruzi* (i.e. confirmed) in non-desert ecoregions of Mexico. Records were found in the Web of Knowledge of the Institute for Scientific Information (ISI – Thomson Scientific, Philadelphia, PA, USA), BibTri (bibtri. com.ar) databases and bibliographic collections of the Laboratorio de Biología de Parásitos, Facultad de Medicina, Universidad Nacional Autónoma de México (Appendix 1, Supporting Information).

We considered all records of confirmed mammal species in our analyses. In our results, we also mention the diagnostic methods by which infection with *T. cruzi* was determined for each mammal species. As the method used here assumes that the larger the geographical overlap the more likely it is that there is a biotic interaction between the species, it may be argued that the mammal species with the largest overlap are simply those with the greatest geographic range, i.e. that a ranking by ε is equivalent to a ranking by the distribution range. To check this we searched for statistical differences, using a *t*-test, between the relative ranking of both ε values and the distribution sizes of mammals, i.e. the number of grid cells with records of a given mammal species. Two sample groups were tested: *T. dimidiata* 2, a relatively widespread species (245 grid cells), and *T. picturata*, a species with a narrow-distribution (20 grid cells).

RESULTS

Inferred interaction network of triatomines and mammals

Our data mining approach allowed us to identify the statistically significant ($\varepsilon > 1 \cdot 96$) potential vector–host pair associations. From these 643 pairs, for each triatomine species we chose the 25% [top quartile (Q4)] of highest ε values. Thus, our network represents the most significant (Q4) positive co-occurrence associations between Triatominae and mammal species (Fig. 2). These potential relationships are the most likely (but not surely) to yield an important biotic interaction between a triatomine and a mammal, yielding the most statistically significant geographic overlaps between triatomines and mammals and therefore the highest potential for encounters given our 'all else being equal' assumption. For a given threshold on ε it is 'maximal' in that it encapsulates the idea of capturing those triatomine–mammal pairs that most satisfy the necessary condition of spatial overlap, but without any assumption of other potential conditions that would cause the interaction to either not be present at all – e.g., the triatomine does not feed on that mammal, or that the interaction does not lead to infection, such as if the mammal has very low competence. It is, of course, of great theoretical and practical interest to know what fraction of that 'maximal' network corresponds to real biotic interactions that also correspond to components of the transmission cycle.

Reviewing the network we note that 116 out of the 396 (29%) total wild mammal species considered could interact with at least one species of Triatominae (Table 1); while 86 of them are potentially associated to only one triatomine species. Ten mammal species probably interact with three or more vectors, among which *Baiomys musculus* and *Liomys pictus* seem the most important as they are potentially associated with five triatomine species (Fig. 2, Table 1). Once again, this network shows species association patterns that are based on the overlap of their geographic ranges, co-occurrence being a necessary condition for a biotic interaction. It does not, however, necessarily prove that there is a direct 'causal' interaction. It can, though, provide testable hypotheses for vector–host interactions.

To check the sensitivity of the network, and therefore our conclusions, to the model parameters and assumptions, such as the grid size and our threshold on ε, we considered how as a base measure the figure of 29% of true positives in Q4, seen in Fig. 3A, changed for three different grid sizes and three different ε thresholds. Checks were made with three grid sizes: the true positive percentages in Q4 were 28% (5 km), 25% (10 km) and 25% (50 km), respectively. We noted that the 29% of true positives in the top quartile does not change significantly ($\chi^2 = 1 \cdot 3$, $P = 0 \cdot 52$) as a function of grid size. On the other

18

Reithrodontomys fulvescens
Sylvilagus floridanus
Dermanura tolteca
Lasiurus blossevillii
Chiroderma salvini
Pappogeomys bulleri
Peromyscus spicilegus
Mephitis macroura
Triatoma longipennis
Cratogeomys fumosus
Artibeus hirsutus
Sigmodon mascotensis
Artibeus jamaicensis
Glossophaga leachii
Sylvilagus cunicularius
Peromyscus hylocetes
Osgoodomys banderanus
Macrotus waterhousii
Desmodus rotundus
Leptonycteris nivalis
Hodomys alleni
Molossus sinaloae
Oryzomys melanotis
Notocitellus annulatus
Triatoma picturata
Peromyscus perfulvus
Sciurus colliaei
Myotis carteri
Pteronotus parnellii
Glossophaga commissarisi
Megasorex gigas
Reithrodontomys hirsutus
Artibeus lituratus

Baiomys taylori
Peromyscus maniculatus
Ictidomys mexicanus
Otospermophilus variegatus
Myotis velifer
Liomys irroratus
Lepus callotis
Baiomys musculus
Balantiopteryx plicata
Liomys pictus
Triatoma pallidipennis

Choeronycteris mexicana
Peromyscus melanophrys
Triatoma barberi
Habromys lepturus
Peromyscus gratus
Tadarida brasiliensis
Peromyscus levipes
Leptonycteris yerbabuenae
Molossus aztecus
Bassariscus sumichrasti
Nyctomys sumichrasti
Lepus flavigularis
Spilogale gracilis
Triatoma phyllosoma
Carollia subrufa
Orthogeomys grandis
Glossophaga morenoi
Peromyscus melanurus
Peromyscus megalops
Sigmodon alleni
Peromyscus mexicanus
Potos flavus
Triatoma mazzottii
Uroderma magnirostrum
Cryptotis phillipsii
Cryptotis goldmani
Tlacuatzin canescens
Lontra longicaudis

Peromyscus pectoralis
Sciurus oculatus
Dermanura azteca
Triatoma mexicana
Microtus quasiater
Sturnira hondurensis
Idionycteris phyllotis
Corynorhinus mexicanus
Peromyscus difficilis
Heteromys desmarestianus
Diphylla ecaudata
Reithrodontomys mexicanus
Sigmodon hispidus
Sphiggurus mexicanus
Platyrrhinus helleri
Glossophaga soricina
Peromyscus furvus
Carollia perspicillata
Triatoma dimidiata 2
Didelphis marsupialis
Rhogeessa tumida
Didelphis virginiana
Oryzomys alfaroi
Philander opossum
Eptesicus furinalis
Oryzomys couesi
Cryptotis mexicana
Sciurus deppei
Orthogeomys hispidus
Molossus rufus
Dasyprocta mexicana
Carollia sowelli
Oligoryzomys fulvescens
Alouatta palliata
Marmosa mexicana
Myotis nigricans
Conepatus semistriatus
Vampyrodes major
Sciurus aureogaster
Mazama americana
Peromyscus leucopus
Triatoma dimidiata 3
Heteromys gaumeri
Otonyctomys hatti
Peromyscus yucatanicus
Reithrodontomys gracilis
Cryptotis mayensis
Peropteryx macrotis
Cuniculus paca
Dasyprocta punctata
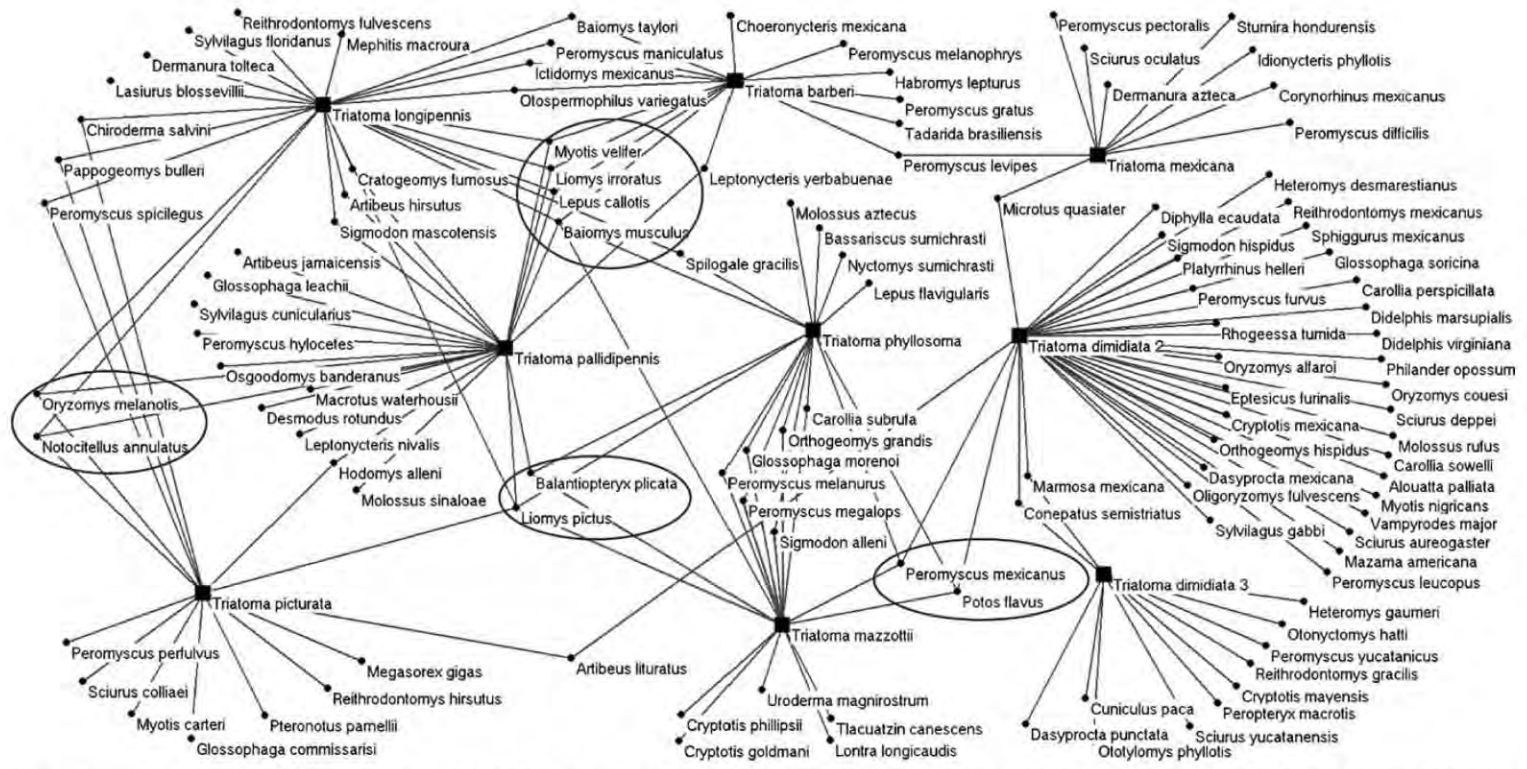Sciurus yucatanensis
Ototylomys phyllotis

Fig. 2. Inferred vector–host network of *Trypanosoma cruzi* from Triatominae (squares) and wild mammals (points) species in non-desert areas of Mexico. Circled mammals potentially interact with three or more triatominae species.

Table 1. Ranked list of potential mammal hosts for *Trypanosoma cruzi* in non-desert areas of Mexico

R	Mammals	ε	Tri	R	Mammals	ε	Tri
1	*Peromyscus yucatanicus*	17·96	1	59	*Rhogeessa tumida*	9·41	1
2	*Orthogeomys hispidus*	16·87	1	60	*Orthogeomys grandis*	9·36[a]	2
3	*Peromyscus mexicanus*	16·65[a]	3	61	*Macrotus waterhousii*	9·3	1
4	*Pappogeomys bulleri*	16·63[a]	2	62	*Lontra longicaudis*	9·24	1
5	*Oligoryzomys fulvescens*	16·33	1	63	*Otospermophilus variegatus*	9·04[a]	2
6	*Baiomys taylori*	15·75[a]	2	64	*Peromyscus pectoralis*	9·04	1
7	*Carollia perspicillata*	15·01	1	65	*Heteromys desmarestianus*	9·01	1
8	*Carollia sowelli*	14·79	1	66	*Didelphis virginiana*	8·91	1
9	*Didelphis marsupialis*	14·59	1	67	*Cryptotis mexicana*	8·82	1
10	*Peromyscus leucopus*	14·27	1	68	*Artibeus lituratus*	8·74[a]	2
11	*Philander oposum*	14·08	1	69	*Reithrodontomys fulvescens*	8·71	1
12	*Osgoodomys banderanus*	13·96	1	70	*Sigmodon alleni*	8·65[a]	2
13	*Peromyscus melanurus*	13·95[a]	2	71	*Myotis nigricans*	8·6	1
14	*Heteromys gaumeri*	13·27	1	72	*Ictidomys mexicanus*	8·59[a]	2
15	*Microtus quasiater*	13·22[a]	2	73	*Molossus rufus*	8·53	1
16	*Reithrodontomys gracilis*	13·18	1	74	*Platyrrhinus helleri*	8·53	1
17	*Oryzomys couesi*	13·14	1	75	*Sigmodon mascotensis*	8·50[a]	2
18	*Lepus callotis*	13·07[a]	3	76	*Mazama americana*	8·46	1
19	*Ototylomys phyllotis*	12·84	1	77	*Vampyrodes major*	8·45	1
20	*Peromyscus furvus*	12·83	1	78	*Hodomys alleni*	8·43[a]	2
21	*Dasyprocta mexicana*	12·5	1	79	*Peropteryx macrotis*	8·36	1
22	*Sigmodon hispidus*	12·47	1	80	*Molossus sinaloae*	8·34	1
23	*Sciurus deppei*	12·28	1	81	*Oryzomys melanotis*	8·19[a]	3
24	*Cryptotis mayensis*	12·19	1	82	*Glossophaga commissarisi*	7·98	1
25	*Otonyctomys hatti*	12·15	1	83	*Leptonycteris nivalis*	7·84	1
26	*Reithrodontomys mexicanus*	12	1	84	*Liomys pictus*	7·65	5
27	*Sylvilagus gabbi*	11·97	1	85	*Cryptotis goldmani*	7·49	1
28	*Sciurus colliaei*	11·93	1	86	*Leptonycteris yerbabuenae*	7·41	1
29	*Molossus aztecus*	11·86	1	87	*Nyctomys sumichrasti*	7·33	1
30	*Reithrodontomys hirsutus*	11·76	1	88	*Bassariscus sumichrasti*	7·23	1
31	*Baiomys musculus*	11·63[a]	5	89	*Lepus flavigularis*	7·2	1
32	*Oryzomys alfaroi*	11·6	1	90	*Peromyscus melanophrys*	7·18	1
33	*Balantiopteryx plicata*	11·60[a]	3	91	*Uroderma magnirostrum*	7·16	1
34	*Dasyprocta punctata*	11·56	1	92	*Sylvilagus cunicularius*	7·09	1
35	*Carollia subrufa*	11·48[a]	2	93	*Chiroderma salvini*	7·04[a]	2
36	*Artibeus hirsutus*	11·31[a]	2	94	*Idionycteris phyllotis*	6·89	1
37	*Sciurus aureogaster*	11·04	1	95	*Artibeus jamaicensis*	6·86	1
38	*Liomys irroratus*	10·88[a]	3	96	*Cryptotis phillipsii*	6·84	1
39	*Eptesicus furinalis*	10·75	1	97	*Dermanura tolteca*	6·77	1
40	*Peromyscus spicilegus*	10·72[a]	2	98	*Sciurus oculatus*	6·67	1
41	*Cuniculus paca*	10·59	1	99	*Glossophaga leachii*	6·61	1
42	*Glossophaga soricina*	10·58	1	100	*Cratogeomys fumosus*	6·54	1
43	*Myotis carteri*	10·41	1	101	*Mephitis macroura*	6·43	1
44	*Megasorex gigas*	10·29	1	102	*Myotis velifer*	6·37[a]	3
45	*Sciurus yucatanensis*	10·2	1	103	*Peromyscus perfulvus*	6·36	1
46	*Marmosa mexicana*	10·12[a]	2	104	*Sylvilagus floridanus*	6·29	1
47	*Sphiggurus mexicanus*	10·11	1	105	*Lasiurus blossevillii*	6·17	1
48	*Dermanura azteca*	10·07	1	106	*Sturnira hondurensis*	6·04	1
49	*Alouatta palliata*	9·95	1	107	*Peromyscus hylocetes*	6·03	1
50	*Peromyscus levipes*	9·90[a]	2	108	*Desmodus rotundus*	6·02	1
51	*Notocitellus annulatus*	9·87[a]	3	109	*Peromyscus difficilis*	6	1
52	*Glossophaga morenoi*	9·84[a]	2	110	*Tadarida brasiliensis*	5·86	1
53	*Spilogale gracilis*	9·84[a]	2	111	*Tlacuatzin canescens*	5·82	1
54	*Peromyscus maniculatus*	9·74[a]	2	112	*Corynorhinus mexicanus*	5·64	1
55	*Conepatus semistriatus*	9·66	1	113	*Choeronycteris mexicana*	5·44	1
56	*Diphylla ecaudata*	9·63	1	114	*Pteronotus parnellii*	5·17	1
57	*Potos flavus*	9·60[a]	3	115	*Peromyscus gratus*	5·15	1
58	*Peromyscus megalops*	9·55	1	116	*Habromys lepturus*	4·74	1

R, relative rank of a mammal species, the lowest values of R being the most important; ε, epsilon values for mammal species estimated according to the level of co-occurrence with a given triatominae species; Tri., number of species of Triatominae potentially interacting with a given mammal species.

[a] Only the highest ε value is reported for a mammal species associated to two or more triatominae species.

20
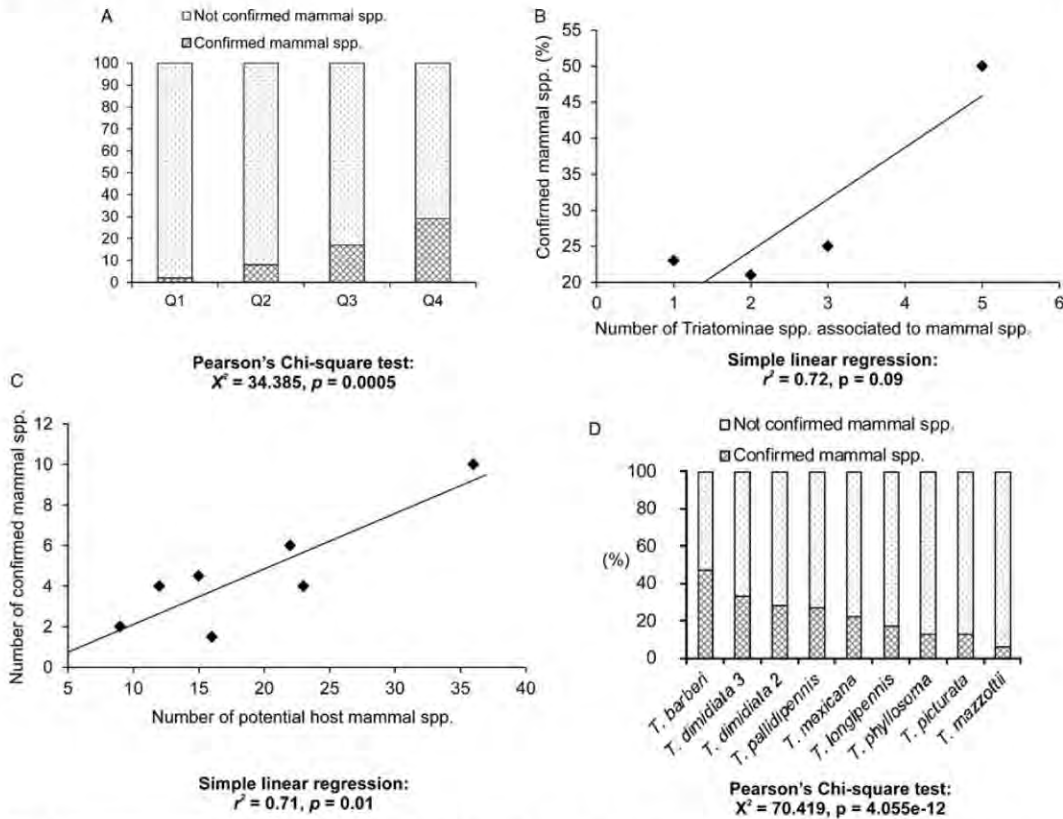
Fig. 3. Hypotheses of ecological epidemiology of Chagas disease in Mexico. Hypotheses were examined considering the wild mammal species confirmed as *Trypanosoma cruzi* hosts in Mexico. (A) H1. The probability for a mammal species to be confirmed for *T. cruzi* is different among quartiles. (B) H2. The probability for a mammal species to be confirmed is not associated to the number of Triatominae species associated with a mammal species. (C) H3. The number of confirmed mammal species is correlated to the number of mammal species associated to a triatomine species. (D) H4. Percentages of confirmed mammals are different among triatomine species.

hand, checking three different ε thresholds ($\varepsilon > 1\cdot96$, $\varepsilon > 4$ and $\varepsilon > 6$), the percentage of true positives increased significantly in high quartiles as a function of the ε threshold, for example, with threshold $\varepsilon > 4$, the true positives percentage was higher in Q3 (27%), than with threshold $\varepsilon > 1\cdot96$ Q3 (17%) ($\chi^2 = 6\cdot7$, $P = 0\cdot03$), and with threshold $\varepsilon > 4$, the true positives percentage was higher in Q4 (33%), than with threshold $\varepsilon > 1\cdot96$ Q4 (29%) ($\chi^2 = 24\cdot2$, $P < 0\cdot0001$). These results are to be expected given that a higher threshold on ε means that we are restricting attention to a smaller subset of relations which are more statistically significant and therefore more likely to be associated with a true positive.

Ecological epidemiology of Chagas disease

Our literature survey uncovered 37 wild mammal species confirmed as hosts of *T. cruzi* in non-desert areas of Mexico (Table 2). Of these species 43% are of the order Rodentia, 38% are Chiroptera and the

other records are Carnivora, Didelphimorphia and Xenarthra (19%). Of these 37 species, 32 have been identified as positive using multiple diagnostic tests. We were not able to determine the used diagnostic test for *T. cruzi* for the remaining five species: *Carollia perspicillata, Dasypus novemcinctus, Hodomys alleni, Ototylomys phyllotis* and *Tylomys nudicaudus*.

With our list of mammals ranked by ε, and the list of confirmed hosts, we can test the network as a predictive model and also construct some simple hypotheses based on the overall structure of the network. More sophisticated hypotheses will potentially need additional data beyond just point collection data. Overall, all the confirmed mammal species were predicted by our analysis as being potentially associated in a statistically significant way to at least one triatomine species ($\varepsilon > 1\cdot96$). This allows us to formulate a first prediction: for a given mammal species, to be a host it must co-occur with the vector, we posit then that it is more likely to be

21

Table 2. Wild mammal species confirmed as *Trypanosoma cruzi* hosts in Mexico

	Confirmed mammal	OR	Method	Q
1	*Artibeus lituratus*	Ch	P	4
2	*Baiomys musculus*	Rd	B,C,P	4
3	*Carollia perspicillata*	Ch	ND	4
4	*Carollia sowelli (brevicauda)*	Ch	P	4
5	*Dasyprocta punctata*	Rd	A,B	4
6	*Didelphis marsupialis*	Dp	A,B,C,X	4
7	*Didelphis virginiana*	Dp	B,C,P,X	4
8	*Glossophaga soricina*	Ch	B,C	4
9	*Heteromys desmarestianus*	Rd	B,C,X	4
10	*Heteromys gaumeri*	Rd	PCR	4
11	*Liomys irroratus*	Rd	B,C,P	4
12	*Otospermophilus (Spermophilus) variegatus*	Rd	B,C	4
13	*Ototylomys phyllotis*	Ch	ND	4
14	*Peromyscus leucopus*	Rd	A,B	4
15	*Peromyscus levipes*	Rd	P	4
16	*Peromyscus mexicanus*	Rd	B,C,X	4
17	*Peromyscus yucatanicus*	Rd	A,B,P	4
18	*Philander oposum*	Dp	R	4
19	*Reithrodontomys fulvescens*	Rd	P	4
20	*Sigmodon hispidus*	Rd	B,C,P,X	4
21	*Artibeus jamaicensis*	Ch	B,C,P,X	3
22	*Choeronycteris mexicana*	Ch	B,C	3
23	*Desmodus rotundus*	Ch	B,C,X	3
24	*Hodomys alleni*	Rd	ND	3
25	*Leptonycteris yerbabuenae (curasoae)*	Ch	B,C	3
26	*Myotis keaysi*	Ch	P	3
27	*Nasua narica*	Cr	P	3
28	*Peromyscus melanophrys*	Rd	B,C	3
29	*Sturnira hondurensis (ludovici)*	Ch	P	3
30	*Sturnira lilium*	Ch	B,C,P	3
31	*Tylomys nudicaudus*	Rd	ND	3
32	*Dasypus novemcinctus*	Xn	ND	2
33	*Dermanura phaeotis*	Ch	P	2
34	*Neotoma mexicana*	Rd	B,C,P	2
35	*Procyon lotor*	Cr	P	2
36	*Pteronotus parnellii*	Ch	B,C	2
37	*Urocyon cinereoargenteus*	Cr	A,B	1

OR, order of mammal species; Q, quartil of ε values, quartile 1 (Q1) being the lowest ε values and Q4 the highest; A, antibodies; B, blood smear; C, culture; Ch, Chiroptera; Cr, Carnivora; Dp, Didelphimorphia; P, polymerase chain reaction; R, random amplified polymorphic DNA; Rd, Rodentia; X, Xenodiagnosic; Xn, Xenarthra; ND, no data.

confirmed as a host of *T. cruzi* if it has a statistically significant overlap with a triatomine species (Table 3, H1). To test this hypothesis, we ranked all mammal species with significant co-occurrence associations ($\varepsilon > 1.96$) with triatomine species according to their ε values and classified them in quartiles. So, quartile 1 (Q1) represents the mammals with the lowest ε values and Q4 the highest, ranking from low to high statistically significant associations between triatomine and mammal species. The corresponding quartiles of confirmed hosts were then assigned (Table 2). We found that the level of association between a mammal and a triatomine species correlated very well with the probability to be a confirmed host for *T. cruzi* ($\chi^2 = 34.385$, $P = 0.0005$, Fig. 3A). Thus, we can see that our inferred interaction network (Fig. 2) serves as a good prediction model for the vector–host system. Note that although only 29% of mammal species in Q4 have been

confirmed as hosts this serves only as a lower bound as many of the species in Q4 that have not been confirmed have either not been collected and tested for presence of *T. cruzi* or in such small numbers that a statistically significant rejection of them as hosts given a null hypothesis about the expected infection rate is not possible. The data in Fig. 3 were split into quartiles to facilitate the visual inspection of the relation between the true positive rate and the average value of ε in the quartiles in a way that presenting the regression coefficients and R^2 value for the logistic regression does not. The coarse graining we use is not *ad hoc*. In the case of deciles rather than quartiles it is the standard grouping into by risk score used in the Hosmer–Lemeshow test often used with logistic regressions. We have also carried out a logistic regression at the species level. The associated relation is: Logit $P = -3.648 + 0.235 \times \varepsilon$, with a P-value < 0.001 on the regression coefficient. This

22

Table 3. Hypotheses of ecological epidemiology of Chagas disease in Mexico

Hyp.	Data	Ho.	Hi.
H1	Mammal and triatominae species with significant values of ε (Q1 to Q4)	The probability for a mammal species to be confirmed is independent of the level of co-occurrence with a triatomine species (Quartil)	The probability for a mammal species to be confirmed depends on the level of co-occurrence with a triatomine species (Quartil)
H2	Only mammal and triatominae species with the highest significant values of ε (Q4)	The probability for a mammal species to be confirmed is not correlated to the number of triatomine species co-occurring with the mammal species	The probability for a mammal species to be confirmed increases when the number of triatomine species co-occurring with the mammal species increases
H3		The number of confirmed mammal species is not correlated with the number of mammal species co-occurring with a triatomine species	The number of confirmed mammal species increases when the number of mammal species co-occurring with a triatomine species increases
H4		The ability of triatomine species to transmit the parasite after a feeding interaction of its individuals is independent of the probability that a mammalian species is confirmed	The ability of triatomine species to transmit the parasite after a feeding interaction of its individuals explains the probability that a mammalian species is confirmed

Hypotheses were formulated considering patterns of interaction among potential vectors and hosts of *Trypanosoma cruzi* (ε data).

confirms the statistically significant relation between ε as a statistical measure of geographical overlap and the probability to be a host of *T. cruzi*.

A second hypothesis is that mammal species that co-occur significantly with several triatomine species have a higher chance of being hosts of *T. cruzi* than mammal species that co-occur with few triatomine species (Table 3, H2). If this hypothesis is confirmed, then we expect an increment in the proportion of confirmed mammal host species as the number of associations increase. To test this hypothesis we assumed that all triatomine species have the same competence to transmit *T. cruzi* and the same population density. In this case, we cannot reject the null hypothesis at any level of statistical confidence and so we conclude that the probability to be a confirmed mammal host does not increase proportionally to the number of triatomine species for which these mammal species co-occur ($r^2 = 0.72$, $P = 0.09$, Fig. 3B).

A third hypothesis is that the transmission of *T. cruzi* is a more common process for triatomine species associated with many mammal species than for triatomine species associated with only a few (Table 3, H3). If this hypothesis is valid, then the number of confirmed mammal host species will increase as the number of mammal associated to a triatomine species increases. Again, to test this hypothesis we assumed that all triatomine species have the same competence to transmit *T. cruzi* and the same population density. Our statistical test indicated that when the number of mammal associated with a triatomine species increases the number of confirmed mammal host species also increases ($r^2 = 0.72$, $P = 0.01$, Fig. 3C).

The previous hypotheses explain the role of mammal species in the transmission of *T. cruzi*, but ignore the explicit role of each vector. Also, we are implicitly considering the importance of a given mammal species for transmission of *T. cruzi*, without taking into account specific DTUs. In contrast to the above analyses, to explain the role of a specific vector we should ignore the role of the interaction in the transmissibility of the parasite. In other words, we assume that every triatomine species and their potential feeding resources are isolated from the other triatomine species or triatomine species that do not share any mammal species. In accordance with this setting, we can evaluate whether the transmission characteristics of distinct triatomine species are different (Table 3, H4). From the total set of potential feeding resources of a triatomine species, we compared the confirmed and non-confirmed mammal species percentages. If the transmissibility of triatomine species were the same, we would expect the percentages to be conserved among triatomine species. We found that the percentages of confirmed mammal hosts were different among triatomine species ($\chi^2 = 70.419$, $P = 4.055 \times 10^{-12}$), being the highest for *T. barberi*, *T. dimidiata* 2 and 3, and *T. pallidipennis* (Fig. 3D). Our results could be interpreted as showing that every triatomine species has a different competence to transmit *T. cruzi* and/or a different population density. It is interesting that in this way one can potentially infer indirectly vector competence in terms of the proportion of species it may infect.

Finally, to test the hypothesis that the statistical associations between vector and potential host do not simply reflect the relative range size of the different mammal (i.e., that a ranking by ε is different to a ranking by the distribution range), we show in Table S2 of the supporting information

23

that these distributions are quite distinct both *T. dimidiata* 2 ($t = 2 \cdot 53$, $P = 0 \cdot 01$) and *T. picturata* ($t = -2 \cdot 57$, $P = 0 \cdot 01$). Therefore, mammals' range sizes do not explain the observed co-occurrence patterns.

DISCUSSION

We inferred the potential vectors and hosts involved in the transmission of *T. cruzi* in non-desert ecoregions of Mexico and deduced the possible epidemiological consequences of triatomine–mammal interactions based on their geographic co-occurrence patterns. Certainly, transmission of *T. cruzi* could potentially occur in ways that do not directly involve a vector, e. g. maternal infection, feeding on infected mammals, etc. (Jansen *et al.* 2015). However, for a mammal species sharing most of its distribution with a triatomine species, we would expect that an important transmission route should be through vector interactions. An advantage of the type of analysis carried out here is that mammal occurrence data are much more complete and widely available than abundance data. Therefore, we posit that interaction networks inferred from co-occurrence patterns are efficient proxies with which to recognize potential hosts of *T. cruzi* and to understand their macro-level transmission dynamics in megadiverse countries.

As Mexico is a megadiverse country, there is a huge number of possible components of the vector-host system: 550 wild mammals and more than 30 Triatominae species (Ceballos and Arroyo, 2012; Ramsey *et al.* 2015). Complex Inference Networks allow us to recognize the most likely and most important wild hosts of *T. cruzi*, considering only those species with a significant co-occurrence. We predicted which were the most important mammal (116 species) and triatomine species involved in the ecological epidemiology of Chagas disease in Mexico. The high level of coincidence found between the predicted and confirmed hosts (Table 2), implies that many mammal species in our vector–host system can be considered as potential hosts or *T. cruzi*. Our results can help drive efforts for future experimental studies to confirm if the most probable predicted hosts are actually reservoirs of *T. cruzi*.

Interaction networks allow us to recognize patterns of transmissibility of *T. cruzi*. Testing our hypothesis 1 we observed the most of mammals confirmed positives to *T. cruzi* in the top quartile of our ranked list (Fig. 3A). This top quartile includes the most important spatial associations of triatomine and mammal species, relative to quartiles 1 to 3. Therefore, we rejected the null hypothesis of a vector-independent transmission of *T. cruzi*, which could exhibit a same number of confirmed mammals for all mammal species with a statistically significant geographical overlap with triatomines

(Q1 to Q4). We concluded that hosting a *T. cruzi* can be correlated to mammal and vector co-occurrence and we interpreted this as an evidence of biotic interaction between mammals and triatomines.

Of course, the nature of this biotic interaction can itself be quite complex and multi-faceted. The most natural interaction, given that Triatominae are hematophagous, should be a feeding interaction, whereby a triatomine takes a bloodmeal from the mammal and the consequent triatomine defecation leads to an infection. This type of interaction can be confirmed with studies of blood meal origin at mammal species level. For example, our prediction of *Mephitis macroura* (Mephitidae: Carnivora), *Reithrodontomys fulvescens* and *Sigmodon mascotensis* (Cricetidae: Rodentia) as feeding resources of *T. longipennis* has been confirmed by a blood meal origin study (Bosseno *et al.* 2009). Similarly, *B. musculus* has been confirmed as a feeding resource of *T. barberi*, *T. pallidipennis* and *T. phyllosoma* (Mota *et al.* 2007). Given the scarcity of research about blood meal origin for triatomine species in Mexico (Mota *et al.* 2007; Bosseno *et al.* 2009; Ramsey *et al.* 2012), the interactions inferred by our model remain mostly as potential species. Another plausible type of interaction is that triatomine species are feeding resources for a given mammal species. However, this scenario seems less common in our network as most of the mammal species inferred are not insectivores (75%; González-Salazar *et al.* 2014) (Table S3, Supporting Information).

It is important to note that, although the detailed nature of the biotic interaction between an individual vector and an individual host may be important at some level, it does not affect our results, which are at a macro, ecosystemic level and therefore independent of the precise details of the interaction. That is not to say such details are unimportant. Moreover, this complexity extends further, considering the inclusion of the parasite itself, in that different mammals could have quite different competencies. Each species could deal in a different way with an infection by distinct *T. cruzi* DTUs. Similarly, the detailed behavioural traits of different mammal species can affect transmission probabilities. For instance, the confirmed host *D. novemcinctus* constructs burrows that upon abandonment are often used as shelter or breeding sites by other mammal species and triatominae thereby allowing for a Triatomine to feed on multiple mammal species. In principle, our methodology could take into account much more complexity if there were data to support it. For instance, there is no comprehensive database that lists the competencies of all mammal species with respect to all DTUs for instance.

The data that do exist for all species is where they are, at least as proxied by point collection data. Our

24

research in that sense provides a first, crude but effective approximation to a very complex system: the vector-host ecosystem of Chagas disease in Mexico. In this respect, hypotheses 2 and 3 attempt to explain the role of potential interactions between mammal and triatomine species in the transmission of *T. cruzi*. We are acutely aware that transmissibility of *T. cruzi* involves many factors, not only mammal and triatomine co-occurrence. Here, the explicit role of every vector and mammal species, i.e. its competence and population density, was assumed similar in the absence of standard data of competence of triatomine species and available data on population densities. These are model assumptions. The approximate validity of those assumptions is tested by the results of the model. From the available data, we predict that transmissibility stays relatively constant for mammal species associated to a few or a lot of triatomine species (H2) with about 20% of mammals with significant ε values being confirmed hosts independent of the associated vector. Also, our data allow us to predict that the probability of transmission of *T. cruzi* increases for triatomine species associated from a few to a lot of mammal species (H3). This is not just a question of expecting that the more mammal species that are sampled the more positives one would expect. If mammals with low ε values were sampled then one would expect only a small number of confirmed mammals or none, no matter how many mammal species were sampled.

Testing differences in the transmissibility of *T. cruzi* among triatomine species from spatial data (Hypothesis 4), we observed that *Triatoma barberi*, *T. dimidiata* and *T. pallidipennis* have a particularly important role in *T. cruzi* transmission. This result might be expected as these species have the widest distribution in Mexico (Ramsey *et al.* 2015), but for the first time we were able to predict the epidemiological importance of some triatominae species by considering the number of potential feeding resources and confirmed mammal hosts. The transmission of *T. cruzi* is a process that usually occurs by contact between a mammal and a vector, such as via contaminated triatomine feces, a process that is assumed to be repeated in proportion to the populations and distribution sizes of the corresponding mammal and triatomine species. Due to the nature of the point collection data used, the population density of the species is unknown. However, we do know that all the interactions (links in our network) come from a large spatial overlap between mammal and triatomine distributions. Hence, a link in the network between mammal and triatomine species means a high potential for *T. cruzi* transmission. Therefore, we were able to test hypotheses of transmissibility with our vector–host system because potential interactions between species are quantifiable.

We note that differences in the transmissibility of *T. cruzi* between triatomine species are not only a result of mammal and vector densities and distributions, but also a result of different vector competences. For example, *Triatoma barberi* exhibits the best competence to transmit *T. cruzi* having the highest natural infection index, the highest frequency of trypomastigotes and the shortest time for defecation among the main vectors of Chagas disease in Mexico (Salazar-Schettino *et al.* 2005). Likewise, *T. dimidiata* and *T. pallidipennis* are recognized by their high degree of competence among the main vectors of Chagas disease in Mexico (Martínez-Ibarra and Novelo-López, 2004; Salazar-Schettino *et al.* 2005; Dorn *et al.* 2007). Even though there are no differences in the competence of distinct lineages of *T. dimidiata*, there are differences in their spatial dynamics (Herrera-Aguilar *et al.* 2009). *Triatoma dimidiata* 3 participates in the flow between sylvatic and domestic environments whereas *T. dimidiata* 2 does not, being restricted to only domestic habitats exclusively (Herrera-Aguilar *et al.* 2009).

Finally, we were able to recognize the epidemiological consequences of interactions between mammal and triatomine species, in spite of the limitations of our data and assumptions. Certainly, distribution datasets accumulate taxonomic and geographic sampling biases, for instance, accounting for the fact that some areas have been subject to intense field surveys while others have not. However, it is essential that we take advantage of the huge quantity of accumulated data (Varela *et al.* 2014). Although we did not analyse in depth the potential effect of sampling biases in our data, we believe that our coarse graining can reduce to some extent some collection bias by only counting once multiple collections in one grid square. In addition, Complex Inference Networks have been shown to be predictive in spite of sampling collection biases, as has been shown in the case of Leishmaniasis (Berzunza-Cruz *et al.* 2015; Stephens *et al.* 2016).

We are aware that the list of confirmed mammal species with *T. cruzi* has some limitations. In the first place, the list is not definitive and does not come from systematic samples of Mexico, but still represents the most complete knowledge available for the country today. Our list also reflects the diversity of mammal species by order in Mexico (Ceballos and Arroyo, 2012) and seems to not be biased for widely distributed mammal species. Secondly, some level of uncertainty is to be expected in *T. cruzi* determination depending on the diagnostic method used. Finally, we did not include information of DTU's of *T. cruzi* because this information is scarce in Mexico. However, one would expect that any pathogen that has a transmission cycle that involves triatomines and mammals will show a similar ecosystemic network. We would argue then that the list of confirmed mammal species with *T. cruzi* was sufficient to test our hypotheses.

With respect to our assumptions, we recognize that not only direct biotic interactions, such as feeding, but also other ecological interactions and evolutionary biogeographic processes can cause co-occurrence patterns between mammal and triatomine species (Morrone, 2009). In other words there may exist confounding factors such that a perceived pair correlation is really intermediated by another latent variable, such as climate. This can only be checked thoroughly by exhaustively including every potential confounding variable and checking if it is more predictive than its proxy. However, even if co-occurrence patterns had not been a direct consequence of feeding interactions, the fact that triatomines would feed more on those mammals that have a higher fraction of co-occurrences seems plausible because of their generalist habits. Only a few triatomine species have definite host preferences (Lent and Wygodzinsky, 1979) and they are not within the set of species studied herein (Dorn *et al.* 2007; Bosseno *et al.* 2009; Villalobos *et al.* 2011; Ramsey *et al.* 2012). We hope to examine some of the model assumptions within our vector-host system when more data are obtained. Finally, we hope that our findings will stimulate other researchers in the direction of making corresponding epidemiological hypotheses that can be verified by further experimental and ecological research.

Finally, we have emphasized throughout that co-occurrence, although an important necessary condition for a mammal to participate in the transmission cycle of Chagas disease, is not sufficient as the transmission cycle is highly multifactorial, with factors such as species abundance, phylogeny, sampling frequency, phenotypic characteristics of the potential host, to name but a few could play a significant role. Our formalism lends itself to the incorporation of such factors in a democratic fashion. What is lacking is to have databases that contain such information to integrate with the purely spatial data used here. This is an ongoing effort and will be reported in a future publication.

SUPPLEMENTARY MATERIAL

The supplementary material for this article can be found at https://doi.org/10.1017/S0031182016002468.

REFERENCES

Abad-Franch, F., Diotaiuti, L., Gurgel-Gonçalves, R. and Gürtler, R. E. (2013). Certifying the interruption of Chagas disease transmission by native vectors: cui bono? *Memórias do Instituto Oswaldo Cruz* **108**, 251–254.

Bargues, M. D., Klisiowicz, D. R., Gonzalez-Candelas, F., Ramsey, J. M., Monroy, C., Ponce, C., Salazar-Schettino, P. M., Panzera, F., Abad-Franch, F., Sousa, O. E., Schofield, C. J., Dujardin, J. P., Guhl, F. and Mas-Coma, S. (2008). Phylogeography and genetic variation of *Triatoma dimidiata*, the main Chagas disease vector in Central America, and its position within the genus *Triatoma*. *PLoS Neglected Tropical Diseases* **2**, e233.

Berzunza-Cruz, M., Rodriguez-Moreno, Á., Gutiérrez-Granados, G., González-Salazar, C., Stephens, C. R., Hidalgo-Mihart, M., Marina, C. F., Rebollar-Téllez, E., Bailón-Martínez, D., Balcells, C., Ibarra-Cerdeña, C. N., Sánchez-Cordero, V. and Becker, I. (2015). *Leishmania* (L.) *mexicana* infected bats in Mexico: novel potential reservoirs. *PLoS Neglected Tropical Diseases* **9**, 1–15.

Bosseno, M. F., Barnabé, C., Ramirez, M. J., Kengne, P., Guerrero, S., Lozano, F., Ezequiel, K., Gastélum, M. and Brenière, S. F. (2009). Wild ecotopes and food habits of *Triatoma longipennis* infected by *Trypanosoma cruzi* lineages I and II in Mexico. *The American Journal of Tropical Medicine and Hygiene* **80**, 988–991.

Ceballos, G. and Arroyo-Cabrales, J. (2012). Lista actualizada de los mamíferos de México 2012. *Revista Mexicana de Mastozoología, Nueva época* **2**, 28–80.

Dorn, P., Monroy, C. and Curtis, A. (2007). *Triatoma dimidiata* (Latreille, 1811): a review of its diversity across its geographic range and the relationship among populations. *Infection, Genetics and Evolution* **7**, 343–352.

Dorn, P. L., Calderón, C., Melgar, S., Moguel, B., Solorzano, E., Dumonteil, E., Rodas, A., Rua, N., Garnica, R. and Monroy, C. (2009). Two distinct *Triatoma dimidiata* (Latreille, 1811) taxa are found in sympatry in Guatemala and Mexico. *PLoS Neglected Tropical Diseases* **3**, e393.

González-Salazar, C. and Stephens, C. (2012). Constructing ecological networks: a tool to infer risk of transmission and dispersal of Leishmaniasis. *Zoonoses and Public Health* **59**, 179–193.

González-Salazar, C., Stephens, C. and Marquet, P. A. (2013). Comparing the relative contributions of biotic and abiotic factors as mediators of species' distributions. *Ecological Modelling* **248**, 57–70.

González-Salazar, C., Martinez-Meyer, E. and López-Santiago, G. (2014). A hierarchical classification of trophic guilds for North American birds and Mammals. *Revista Mexicana de Biodiversidad* **85**, 931–941.

Graham, C. H., Ferrier, S., Huettman, F., Moritz, C. and Peterson, A. T. (2004). New developments in museum-based informatics and applications in biodiversity analysis. *Trends in Ecology and Evolution* **19**, 497–503.

Herrera-Aguilar, M., Be-Barragán, L. A., Ramirez-Sierra, M. J., Tripet, F., Dorn, P. and Dumonteil, E. (2009). Identification of a large hybrid zone between sympatric sibling species of *Triatoma dimidiata* in the Yucatán Peninsula, Mexico, and its epidemiological importance. *Infection, Genetics and Evolution* **9**, 1345–1351.

Jansen, A. M., and Roque, A. R. L. (2010). Domestic and wild mammalian reservoirs. In *American Trypanosomiasis Chagas Disease. One Hundred Years of Research* (ed. J. Telleria and M. Tibayrenc), pp. 249–276. Elsevier, Amsterdam.

Jansen, A. M., Xavier, S. C. C. and Roque, A. R. L. (2015). The multiple and complex and changeable scenarios of the *Trypanosoma cruzi* transmission cycle in the sylvatic environment. *Acta Tropica* **151**, 1–15.

Lent, H. and Wygodzinsky, P. (1979). Revision of the Triatominae (Hemiptera, Reduviidae) and their significance as vectors of Chagas disease. *Bulletin of the American Museum of Natural History* **163**, 123–520.

Martinez-Ibarra, J. A. and Novelo-López, M. (2004). Blood meals to molt, feeding time and postfeeding defecation delay of *Meccus pallidipennis* (Stal, 1872) (Hemiptera: Reduviidae) under laboratory conditions. *Folia Entomologica Mexicana* **43**, 313–319.

Monteiro, F. A., Peretolchina, T., Lazoski, C., Harris, K., Dotson, E. M., Abad-Franch, F., Tamayo, E., Pennington, P. M., Monroy, C., Cordon-Rosales, C., Salazar-Schettino, P. M., Gómez-Palacio, A., Grijalva, M. J., Beard, C. B. and Marcet, P. L. (2013). Phylogeographic pattern and extensive mitochondrial DNA divergence

disclose a species complex within the Chagas disease vector *Triatoma dimidiata*. *PLoS ONE* **8**, e70974.

Morrone, J. J. (2009). *Evolutionary Biogeography: an Integrative Approach with Case Studies*. Columbia University Press, New York.

Mota, J., Chacon, J. C., Gutierrez, A., Sanchez-Cordero, V., Wirtz, R., Ordoñez, R., Panzera, F. and Ramsey, J. M. (2007). Identification of Chagas disease vector blood meal source using a multiplex cytochrome b PCR assay. *Vector Borne Zoonotic Disease* **7**, 617–627.

Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V. N., Underwood, E. C., D'Amico, J. A., Itoua, I., Strand, H. E., Morrison, J. C., Loucks, C. J., Allnutt, T. F., Ricketts, T. H., Kura, Y., Lamoreux, J. F., Wettengel, W. W., Hedao, P. and Kassem, K. R. (2001). Terrestrial ecoregions of the world: a new map of life on Earth. *Bioscience* **51**, 933–938.

Ponder, W. F., Carter, G. A., Flemons, P. and Chapman, R. R. (2001). Evaluation of museum collection data for use in biodiversity assessment. *Conservation Biology* **15**, 648–657.

Ramsey, J. M., Gutierrez-Cabrera, A. E., Salgado-Ramírez, L., Peterson, T. A., Sanchez-Cordero, V. and Ibarra-Cerdeña, C. N. (2012). Ecological connectivity of *Trypanosoma cruzi* reservoirs and *Triatoma pallidipennis* in an anthropogenic landscape with endemic Chagas disease. *PLoS ONE* **7**, e46013.

Ramsey, J. M., Peterson, A. T., Carmona-Castro, O., Moo-Llanes, D. A., Nakazawa, Y., Butrick, M., Tun-Ku, E., Cruz-Félix, K. and Ibarra-Cerdeña, C. N. (2015). Atlas of Mexican Triatominae (Reduviidae: Hemiptera) and vector transmission of Chagas disease. *Memórias do Instituto Oswaldo Cruz* **110**, 339–352.

Rodrigues-Coura, J. (2013). Chagas disease: control, elimination and eradication. Is it possible? *Memórias do Instituto Oswaldo Cruz* **108**, 962–967.

Salazar-Schettino, P. M., Arteaga, I. H. and Cabrera, M. (2005). Tres especies de triatominos y su importancia como vectores de *Trypanosoma cruzi* en México. *Medicina Buenos Aires* **65**, 63–69.

Sierra, R. and Stephens, C. R. (2012). Exploratory analysis of the interrelations between co-located Boolean spatial features using network graphs. *International Journal of Geographical Information Science* **26**, 441–468.

Stephens, C. R., Heau, J. G., González, C., Ibarra-Cerdeña, C., Sánchez-Cordero, V. and González-Salazar, C. (2009). Using biotic interaction networks for prediction in biodiversity and emerging diseases. *PLoS ONE* **4**, e5725.

Stephens, C. R., González-Salazar, C., Sánchez-Cordero, V., Becker, I., Rebollar-Tellez, E., Rodriguez-Moreno, Á., Berzunza-Cruz, M., Domingo Balcells, C., Gutiérrez-Granados, G., Hidalgo-Mihart, M., Ibarra-Cerdeña, C. N., Ibarra López, P., Iñiguez Dávalos, L. I. and Ramírez Martínez, M. M. (2016). Can you judge a disease host by the company it keeps? Predicting disease hosts and their relative importance: a case study for Leishmaniasis. *PLoS Neglected Tropical Diseases* **10**, 1–21.

Varela, S., Anderson, R. P., García-Valdés, R. and Fernández-González, F. (2014). Environmental filters reduce the effects of sampling bias and improve predictions of ecological niche models. *Ecography* **37**, 1084–1091.

Villalobos, G., Martínez-Hernández, F., de la Torre, P., Laclette, J. P. and Espinoza, B. (2011). Entomological indices, feeding sources, and molecular identification of *Triatoma phyllosoma* (Hemiptera: Reduviidae) one of the main vectors of Chagas disease in the Istmo de Tehuantepec, Oaxaca, Mexico. *The American Journal of Tropical Medicine and Hygiene* **85**, 490–497.

World Health Organization (2015). *Chagas Disease in Latin America: an Epidemiological Update Based on 2010 Estimates*. WHO Weekly Epidemiological Record 90. World Health Organization, Geneva, Switzerland.

Yañez-Arenas, C., Guevara, R., Martínez-Meyer, E., Mandujano, S. and Lobo, J. M. (2014). Predicting species' abundances from occurrence data: effects of sample size and bias. *Ecological Modelling* **294**, 36–41.

27

CAPÍTULO 2. Ancient hybridization detection under an ecological-biogeographical approach: footprints in the main Chagas disease vectors in Mexico (*Phyllosoma* complex)

Manuscrito para ser enviado a Molecular Ecology.

ANCIENT HYBRIDIZATION DETECTION UNDER AN ECOLOGICAL-BIOGEOGRAPHICAL APPROACH: FOOTPRINTS IN THE MAIN CHAGAS DISEASE VECTORS IN MEXICO (*PHYLLOSOMA* COMPLEX)

ANCIENT HYBRIDIZATION IN CHAGAS DISEASE VECTORS

Laura Rengifo-Correa[1], Fernando Abad-Franch[2], Fernando Martínez-Hernández[3], Paz M. Salazar-Schettino[4], Juan L. Téllez-Rendón[5], Guiehdani Villalobos[3], and Juan J. Morrone[1]

[1] Museo de Zoología ‗Alfonso L. Herrera', Departamento de Biología Evolutiva, Facultad de Ciencias, Universidad Nacional Autónoma de México, 04510 Mexico City, Mexico

[2] Grupo Triatomíneos, Instituto René Rachou – Fundação Oswaldo Cruz Minas Gerais, Belo Horizonte, Brazil

[3] Departamento de Ecología de Agentes Patógenos, Hospital General Dr. Manuel Gea González, 14080 Mexico City, Mexico

[4] Laboratorio de Biología de Parásitos, Departamento de Microbiología y Parasitología, Facultad de Medicina, Universidad Nacional Autónoma de México, 04510 Mexico City, Mexico

[5] Instituto de Diagnóstico y Referencia Epidemiológicos, 01480 Mexico City, Mexico

Correspondence

Juan J. Morrone. Museo de Zoología ‗Alfonso L. Herrera', Departamento de Biología Evolutiva, Facultad de Ciencias, Universidad Nacional Autónoma de México, 04510 Mexico City, Mexico. Email: juanmorrone2001@yahoo.com.mx

ABSTRACT

Several processes would entangle the phylogenetic relationships of species complexes, as introgression caused by ancient hybridization events. Although the role of the spatial context of hybridization events has been early highlighted, it has been neglected by phylogenetic methods attempting introgression detection. Here, we propose an ecological-biogeographical approach for detection of introgression, by considering not only genetic signals of introgression, but also the spatio-temporal and reproductive ecological contexts of introgression. Our efforts are focused on the *Phyllosoma* complex, a group of kissing bug species (Heteroptera: Reduviidae: Triatominae) exhibiting nearly allopatrid distributions, similar morphology and high reproductive compatibility, and well-known by their medical relevance in Chagas disease transmission in Mexico. Spatial co-occurrences of *Phyllosoma* complex species were modeled on Last Glacial Maximum, using ecological niche modeling (generalized linear models, complex networks). Thus, hypotheses of ancient hybridization between species of the *Phyllosoma* complex were postulated. Plausibility of these hypotheses was asserted through reproductive ecological data. A correlation between gene tree incongruences (ITS-2, *Cyt b*), the spatio-temporal and reproductive ecology contexts made possible the detection of potential ancient introgression in *Phyllosoma* complex samples. Some samples of *Triatoma mazzottii, T. mexicana, T. picturata,* and *T.* sp. aff. *dimidiata* exhibited potential ancient introgression. Samples without potential ancient introgression were reanalyzed with *BEAST, to model incomplete linage sorting in the species tree of *Phyllosoma* complex. Support values of the phylogeny without potential introgression were higher than the values of the phylogeny with potential introgression, exemplifying the relevance of ancient introgression detection.

KEYWORDS

1. INTRODUCTION

The evolutionary consequences of hybridization should not be ignored in phylogenetic analyses. Hybridization in the wild is more common than usually believed, with estimated frequencies around 10% of animal and 25% of plant species (Mallet, 2005). As a result of ancient hybridization events, introgression and hybrid speciation generate reticulate evolutionary histories that may cause gene tree disagreement (Maddison, 1997) and low phylogenetic resolution (Dowling & Secor, 1997). Recently, hybridization is being increasingly recognized as one of the important drivers of low support values or low resolution in phylogenetic inference based on demographic thinking (García et al., 2017; Meyer, Matschiner, & Salzburger, 2016; Wagner, Härtl, Vogt, & Oberprieler, 2017; Wallis et al., 2017). Phylogenetic methods based on demographic thinking might be severely affected by just a single sample of hybrid origin, as population parameters are usually inferred from a small sample size (Jones, Aydin, & Oxelman, 2015). However, phylogenetic methods have hardly ever considered hybridization among studied processes (Chaudhary, Burleigh, & Fernández-Baca, 2013; Heled & Drummond, 2010; Mirarab, Reaz, Bayzid, Zimmermann, Swenson, & Warnow, 2014). Therefore, in order to avoid misleading phylogenetic inferences, phylogeneticists should be concerned about the appropriate detection of hybridization.

Detection of hybridization is a difficult task by several reasons. First, as hybridization and incomplete lineage sorting (another source of gene tree disagreement) generate similar patterns in gene trees, distinguishing between them may be difficult (Funk & Omland, 2003; Holder, Anderson, & Holloway, 2001). Some previously proposed methods have inferred hybridization and deep coalescence together in a phylogenetic context, for instance, under the assumption that all disagreement between gene trees that could not be expected under a coalescent model alone is hybridization (Meng & Kubatko, 2009; Yu, Dong, Liu, & Nakhleh, 2014). In other words, distinguishing between hybridization and deep coalescence depends on how close are real and estimated species tree parameters. To avoid a circular reasoning, the detection of hybridization should be achieved by an independent process from the phylogenetic inference. Second, the distinction between introgression and hybrid

speciation seems untractable for some methods. For instance, some methods have detected hybridization exclusively based on the existence of intermediate genotypes (Gauthier & Lapointe, 2007; Gompert & Buerkle, 2009). As this approach lacks a temporal context, the origin of admixture genotypes (introgression or hybrid speciation) could be assumed only. Therefore, spurious patterns of a phylogenetic analysis would remain, because a differential treatment of samples according to their hybrid origin cannot be considered in the analysis design. A differential treatment of samples means that, on one hand samples with introgressed DNA would be removed from the analyses (Meyer et al., 2016; Wagner et al., 2017) and, on the other hand, samples from species that arose by hybrid speciation should be analyzed. Finally, the minimum requirement of a hybridization process is the sympatry or spatial co-occurrence of potentially interbreeding species (Stephens, Sánchez-Cordero, & González-Salazar, 2017). Even though sympatrid patterns between potentially interbreeding species would clarify an entangled interspecific gene flow, the spatial context of hybridization events has been excluded from phylogenetic context mainly because species´s distribution changes over time (Funk & Omland, 2003). Considering all the previous concerns, it would be important to develop a protocol for detecting hybridization before a phylogenetic analysis.

We propose herein an integrative approach to detect introgression from past events of hybridization between extant species, combining ecological niche modeling (ENM) with the reproductive ecology and phylogenetic contexts. This approach follows four steps to infer species that potentially hybridized in a given spatio-temporal context (steps 1 and 2), and to detect samples of species that potentially suffered ancient introgression (steps 3 and 4) (Figure 1). These steps are: (1) postulating hypotheses of species that potentially hybridized, by considering potential distributions and significant co-occurring species in a given spatio-temporal context; (2) an empirical falsification of hypotheses, by means of interspecific crosses background, i.e. the reproductive ecology context; (3) a discovering of conflictive relationships between gene trees, i.e. the phylogenetic context; and (4) a detection of samples with potential ancient introgression, by a correlation between samples causing gene tree incongruences with

their spatio-temporal and reproductive ecology contexts. Once potential ancient introgression is detected on the samples, the species tree is inferred by reanalyzing samples lacking ancient introgression under a multispecies coalescent context.

We apply this approach to the kissing bugs of the *Phyllosoma* complex (Heteroptera: Reduviidae: Triatominae). This complex includes the main vectors of the parasite causing Chagas's disease, *Trypanosoma cruzi*, in Mexico (Ramsey et al., 2015). The monophyly of the *Phyllosoma* complex is controversial, being conformed for six (Carcavallo, Jurberg, Lent, Noireau, & Galvão, 2000; Lent & Wygodzinsky, 1979) to twelve species (de la Rúa et al., 2014). Also, the phylogenetic relationships between *Phyllosoma* complex species remain unresolved. Previous attempts to infer a molecular phylogenetic hypothesis for this complex have evidenced disagreements between nuclear (ITS-2) and mitochondrial (*Cyt b*) gene trees (Espinoza et al., 2013; Martínez et al., 2006). It has been suggested that hybrids play a role to explain this disagreement (Espinoza et al., 2013; Martínez et al., 2006), due to successful interspecific crosses between *Phyllosoma* complex species in laboratory (Martínez-Ibarra et al., 2008a, 2009, 2011a, b, 2015) and in the field (Martínez-Hernandez et al., 2010). Specimens of hybrid origin usually have one parent phenotype, rather than an intermediate phenotype (Martínez-Hernandez et al., 2010; Martínez-Ibarra et al., 2009). Therefore, detection of samples with ancient introgression is relevant for phylogenetic studies of *Phyllosoma* complex.

Our main goal is to detect ancient introgression between extant species of *Phyllosoma* complex, by an algorithm in which spatio-temporal, ecological and phylogenetic contexts being integrate it. Likewise, we provide a multispecies coalescent inference of *Phyllosoma* complex phylogeny, evidencing a past reticulate evolutionary history. Hopefully, our study will promote a discussion about an integrative insight in phylogenetic methods, allowing a better understanding of the complex evolutionary history behind a phylogeny.

2. MATERIALS AND METHODS

2.1 Species-level taxonomy

We followed the morphological classification of species of Triatominae of Lent & Wygodzinsky (1979). Even though the classification of the *Phyllosoma* complex species has been criticized by authors using the biological species concept, we agree with Lent & Wygodzinsky (1979) that interfertility between populations under laboratory conditions may be not a valid indicator of species rank, but intersterility would strongly suggest specific rank of the putative parents. In addition, a cryptic species of *Triatoma dimidiata*, here referred as *T.* sp. aff. *dimidiata* (Bargues et al., 2008; García et al., 2013), is considered in this study. Under the cohesion species concept (Templeton, 1989), it is expected that demographic exchangeability mechanisms lead to stronger cohesion within each one of studied species than between studied species. In support of this expectation, differences on developmental life cycle parameters (Martínez-Ibarra et al., 2005, 2012, 2013), and habitat selection (Tamay-Segovia et al., 2008), have been recorded for *Phyllosoma* complex species and relatives.

2.2 The spatio-temporal context

In order to detect species that potentially hybridized, we discover their spatio-temporal co-occurrence in a given spatio-temporal context (Figure 1). The procedure followed four steps: 1) to look for a time-frame when the studied taxa might have already speciated, 2) to infer the environmental requirements of the *Phyllosoma* complex species from recent occurrence data, to identify potential distribution of studied species in the Last Glacial Maximum, 3) to look for significant co-occurrences between *Phyllosoma* complex species in the Last Glacial Maximum, and 4) to represent significant co-occurrences in a Complex Inference Network. Details of this procedure are given below.

2.2.1 Occurrence data

We compiled georeferenced localities for five species of the *Phyllosoma* complex (*sensu* Lent & Wygodzinsky, 1979) and three other species, *T. dimidiata*, *T.* sp. aff.

dimidiata, and *T. mexicana*, which have been considered close to the complex (de la Rúa et al., 2014; Lent & Wygodzinsky, 1979). *Triatoma bassolsae*, from the *Phyllosoma* complex (*sensu* Carcavallo et al., 2000), was not considered because of its scarce records. Occurrence data were obtained from the Instituto de Diagnóstico y Referencia Epidemiológica, InDRE (Mexico City). Specimens were collected on villages and surrounding areas of Mexico, from 1999 to 2013, by trained surveillance staff. Samples were determined by trained surveillance staff, and a centralized quality control of taxonomic performance was conducted by InDRE staff every year. All records were georeferenced to locality level, following the catalogs of Instituto Nacional de Estadística y Geografía, INEGI (Mexico City). This dataset includes 5924 point records. An independent dataset was obtained from the atlas of Triatominae of medical relevance in Mexico (Ramsey et al., 2015). This dataset includes 1536 point records.

2.2.2 Study area and temporal considerations

To define the study area, records of the eight species mentioned above were projected onto a map of biomes (Olson et al., 2001). Biomes with at least one species record were considered. Therefore, the study area included from desert and xeric shrublands to tropical and subtropical coniferous forests of America (Olson et al., 2001).

Choosing an appropriate temporal context is necessary to avoid overestimating ancient introgression events by accounting them together with events of speciation by hybridization, and to find different species potential distributions from which known on the present. Speciation events in *Phyllosoma* complex species seems to have occurred during the Pliocene – Early Pleistocene, i.e. 5.33 to 0.78 Ma (Ibarra-Cerdeña, Zaldívar-Riverón, Peterson, Sánchez-Cordero, & Ramsey, 2014), or more recently, at Early Pleistocene, i.e. 2.28 to 0.74 Ma (Bargues et al., 2000). Therefore, a post-speciation period of time for *Phyllosoma* complex species could be Last Glacial Maximum (21 to 18 ka). Also, the Last Glacial Maximum is a period of time known by its extreme climatic conditions around the world (Ruddiman, 2008). As a consequence of climatic conditions differences between Last Glacial Maximum and the present, changes have been

expected between species potential distributions under Last Glacial Maximum climatic conditions and species potential distributions under recent climatic conditions.

2.2.3 Potential distributions of *Phyllosoma* complex species in the Last Glacial Maximum

To identify potential distribution of species of the *Phyllosoma* complex in the Last Glacial Maximum, several steps of the ecological niche modeling's routines (Peterson, 2011) were followed. First, recent and past environmental variables were selected. Second, the environmental requirements of the studied species were inferred from occurrence data and recent environmental variables, by an ecological niche modeling algorithm. Third, model performance was evaluated. Finally, assuming that environmental requirements of studied species have been retained since the past, areas that fulfill the inferred environmental requirements were represented in a map with Last Glacial Maximum environmental conditions. Such areas represent potential distributions of the studied species. The below specifications were followed.

Variables describing climate, microhabitat availability, and the historically accessible area have been considered as covariates of this study, whereas anthropogenic landscape variable has been considered as a cofounder variable. The climatic variables were: mean diurnal temperature range, annual temperature range, annual precipitation, precipitation seasonality, relative humidity. These variables are known for their general ecological relevance to triatomines. Recent climatic layers and corresponding palaeoclimatic layers from the Last Glacial Maximum were obtained with a resolution of 30 arcsec (Collins et al., 2006; Fick & Hijmans, 2017). Variable used as proxy of microhabitat availability was slope. The historically accessible area (Barve et al., 2011) of the studied species was included as biogeographic provinces (Morrone, Escalante, & Rodríguez-Tapia, 2017). Anthropogenic landscape variable was considered as a cofounder of the remaining variables, to control the bias of sample collection strategies and human-made environmental changes. Alternative hypotheses about the effect of each variable on populations of triatomines successful were

postulated, considering the knowledge of Triatominae' biology, to approach to the environmental requirements of *Phyllosoma* complex species.

The environmental requirements of the studied species were inferred under a multi-model inference approach. Performance of several generalized linear models, combining alternative hypotheses about the effect of a given set of current variables, was explored. As generalized linear models required absent data, and only occurrence data are available, records of all species were considered for an estimation of non-occupancy odds of the species of interest. The best model set was used for spatial projection of potential species distribution projection. Finally, areas that fulfill the inferred requirements of studied species were represented on Last Glacial Maximum output maps. Assumptions about conservative environmental requirements seem plausible for studied species, because niche conservatism is a common pattern among Triatominae species from North and Central America (Ibarra-Cerdeña et al., 2014).

2.1.4 Co-occurrence patterns between *Phyllosoma* complex species

The spatial co-occurrence patterns between the potential distribution of studied species were identified by executing the Complex Network Inference method (Stephens et al., 2009). This method allows discriminating species whose co-occurrence patterns are statistically significant from species whose co-occurrence patterns are only expected by random (Stephens et al., 2009). Here, significant co-occurrence patterns were interpreted as a chance for spatio-temporal co-occurrence of studied species, whereas random co-occurrence patterns were discarded, by assuming them as an artifact of the modeling process. Methods are explained in detail by Stephens et. al (2009). Only a modification was explored: instead of point record data, polygon areas (potential distributions) were compared.

We are aware that co-occurrence patterns from potential distributions differ from co-occurrence patterns from actual distributions, by an over-prediction in the former. For instance, negative biotic interactions like competence between close species are not considered under potential distributions, which entail over-predicted areas of distributions. To control the overprediction of co-occurrence patterns from potential

distributions on Last Glacial Maximum, we compared them with co-occurrence patterns from current potential distributions of *Phyllosoma* complex species. Co-occurrence patterns of the Last Glacial Maximum as significant of more significant than current co-occurrence patterns were considered to build the network. A Complex Inference Network was drawn, being only represented significant co-occurring species, and their links, that represented a statistical significance of their spatial association. If interbreeding is possible among all species in the network, the network would represent a hypothesis of introgression events occurring in the LGM.

2.3 The reproductive ecology context

Under this context, our main goal was to falsify empirically hypotheses of hybridization events occurring among compared species in the Last Glacial Maximum (Figure 1). We infer the species that potentially hybridized in a Last Glacial Maximum spatio-temporal context, considering their significant spatial co-occurrence on the Last Glacial Maximum and current hybrids successful. Furthermore, the non-hybridizing species in a Last Glacial Maximum spatio-temporal context are inferred, considering their lacking of co-occurrence on the Last Glacial Maximum or current hybrids unviability. An argument in favor of the preceding comparative process is that the isolation mechanisms are developed progressively under spatial overlap of species distributions (Mallet, 2005). From this argument, we assumed that species without current hybrids had an older co-occurrence than species with current hybrids. In order to fulfill this assumption, the compared species should be meet two prerequisites: 1) to have similar biotic mechanisms to develop reproductive isolation, because are close species, and 2) to have similar ages, because they belong to a fast evolving species complex. As a consequence, we expect old co-occurring species develop reproductive isolation before recent co-occurring species. Therefore, species lacking current hybrids could have been reproductively isolated in the Last Glacial Maximum, if fulfill above mentioned prerequisites.

Interspecific crosses background of *Phyllosoma* complex species, *T. dimidiata, T. sp. aff. dimidiata*, and *T. mexicana*, was recovered from the literature. This information

was categorized in two groups to falsify each hypothesis: 1) all crosses resulting in a viable and fertile F1 (first hybrid), even when experimental percentages were low, and 2) all crosses resulting in non-viable or non-fertile F1 or without F1. On the one hand, hypotheses of hybridization events occurring in the LGM would not be rejected if crosses resulting in a viable and fertile F1 have been reported among compared species (group 1). On the other hand, hypotheses on hybridization events occurring in the Last Glacial Maximum would be rejected if crosses resulting in non-viable or non-fertile F1 or without F1 were reported among compared species (group 2).

2.4 The phylogenetic context

The goals of this context were: 1) to identify conflictive relationships between nuclear (ITS-2) and mitochondrial (*Cyt b*) gene trees; 2) to detect samples with ancient introgression, by a correlating of conflictive phylogenetic relationships with the spatio-temporal and reproductive ecology contexts (Figure 1); and 3) to infer the species tree under a multispecies coalescent context, leaving explicit ancient gene flow. The following specifications were considered.

2.4.1 Taxon sampling, molecular biology techniques and alignments

Six species of the *Phyllosoma* complex *sensu* Lent & Wygodzinsky (1979) and Carcavallo et al. (2000), five species of the *Phyllosoma* complex *sensu* de la Rúa et al. (2014), and *Triatoma infestans*, *T. protracta*, and *Rhodnius neglectus* were considered for the phylogenetic analysis (Table 1). Specimens from Genbank (Table 1) were selected, considering information availability of two studied locus. Other specimens were obtained from colonies (Laboratorio de Biología de Parásitos, Facultad de Medicina, Universidad Nacional Autónoma de México). Only two specimens of *T. pallidipennis* and two specimens of *T. dimidiata* were obtained from the field. *Triatoma dimidiata* holds the largest number of specimens, to properly reflect its genomic variation given its wide spatial distribution. Morphology of all specimens fit with taxonomic keys for studied species (Lent & Wygodzinsky, 1979).

Total genomic DNA was extracted from the six legs of each insect. The legs of each specimen were frozen at -18°C and suspended in 1 mL of lysis solution (50 mM Tris–HCl, 50 mM EDTA pH 8, 50 mM NaCl, 1% SDS, 20 ng/μL proteinase K) and incubated at 37 °C overnight. DNA was extracted using the phenol chloroform technique (Sambrook et al., 1989). Approximately 200 ng of genomic DNA were amplified by the polymerase chain reaction (PCR). Oligonucleotides considered in this study have been described by Marcilla et al. (2001) and Monteiro et al. (2013), being ITS-2 and *Cyt b*, respectively. Amplification and DNA purification protocols followed Martínez et al. (2006). All PCR products were purified and sent to Laboratorio de Secuenciación Genómica de la Biodiversidad y de la Salud (Instituto de Biología, Universidad Nacional Autónoma de México), for sequencing service.

Alignments were performed for gene sequences using the online version of MAFFT v. 7. The alignment of ITS-2 was examined visually and minor corrections were made manually to get a better fit of nucleotide correspondences through indel regions. Orthology of ITS-2 studied sequences was verified according to recommendations of Bargues, Zuriaga, & Mas-Coma (2014). Orthology of *Cyt b* was verified by inspecting for an open reading frame of the sequences of the *Phyllosoma* complex.

2.4.2 Conflictive relationships between gene trees

Gene trees were recovered using a multispecies coalescent analysis (*BEAST, Heled & Drummond, 2010) of all samples. This method coestimates the species tree and gene trees in one Bayesian MCMC analysis, by modeling incomplete lineage sorting and assuming that samples lack of introgression and gene duplication (Heled & Drummond, 2010). Modelling incomplete lineage sorting would be beneficial for phylogenetic inference of *Phyllosoma* complex, because this process is an expected source of gene tree disagreement in fast evolving species (Heled & Drummond, 2010). *BEAST analysis consisted of 500 million generations saving every 50000. A Birth-Death model of speciation was implemented in the analysis, considering its consistent performance under several kinds of data sets (Ritchie, Lo, & Ho, 2016). The ITS-2 gen tree and the *Cyt b* gen tree used a HKY + G nucleotide substitution model, the best-fit models based

on the Bayes score inferred from PartitionFinder v. 1.1.1 (Lanfear, Calcott, Ho, & Guindon, 2012). A suboptimal partitioning scheme of *Cyt b* (one partition) was performed instead of the suggested scheme (three partitions) by PartitionFinder, to avoid an overparametrization considering the small sample size of this sequence. Stationarity and ESS scores (ESS>200) for estimated model parameters were assessed in TRACER v1.6. Tree samples were processed with TREEANNOTATOR v 2.3.2 to calculate maximum clade credibility, without the first 25% of iterations ＿buning in". Conflictive relationships between gene trees were identified visually.

2.4.3 Detection of samples with ancient introgression

Conflictive relationships between ITS-2 (nDNA) and *Cyt b* (mtDNA) gene trees offer clues of ancient introgression. As mitochondrial genes lack recombination, complete mtDNA could reflect their heterospecific origin and be fixed faster than heterospecific nDNA (Funk & Omland, 2003). Therefore, *Cyt b* gene tree could be more susceptible to introgression than ITS-2 gen tree. Furthermore, some samples of a particular nominal species could be closely related to a different and non-related species than to the nominal species, i.e. the nominal species is polyphyletic. Polyphyletic species only described by a mitochondrial gene tree, like *Cyt b*, could be a signal of introgressed mtDNA (Funk & Omland, 2003).

The introgressed origin of samples is clarified by correlating polyphyletic species according *Cyt b* gene tree with their spatio-temporal and reproductive ecology contexts. Introgression is inferred for sympatric samples sharing a *Cyt b* sequence, from genetically (ITS-2) divergent species (Funk & Omland, 2003). In addition to sympatrid patterns of samples, plausibility of interbreeding among nominal species is considered.

2.4.4 Species tree

A species tree was recovered using a multispecies coalescent analysis (*BEAST, Heled & Drummond, 2010), with same specifications of ＿2.4.2 Conflictive relationships between gene trees". Samples with ancient introgression were eluded from this analysis to properly fulfill *BEAST requirements (Heled & Drummond, 2010). The inferred

maximum clade credibility species tree is complemented with a schematic representation of ancient introgression in *Phyllosoma* species and relatives.

3. RESULTS

3.1 The spatio-temporal context

From recent areas of distribution of *Phyllosoma* complex species and relatives, four groups of species co-occurrence can be described (Figure 2). *Triatoma* sp. aff. *dimidiata* and *T. dimidiata* co-occur on the Northern and Central Maya Lowlands. *Triatoma dimidiata* and *T. mexicana* co-occur on the costal lowlands of the Gulf of Mexico, on the eastern part of the Transmexican Volcanic Belt. *Triatoma mazzottii*, *T. pallidipennis*, and *T. phyllosoma* co-occur on the southwest of the Transmexican Volcanic Belt, surrounding the Sierra Madre del Sur. Finally, *T. longipennis*, *T. pallidipennis*, and *T. picturata* co-occur on the west portion of the Transmexican Volcanic Belt.

Seven hypotheses of potential interbreeding arose from the co-occurrence patterns among recent areas of distribution of *Phyllosoma* complex species (Figure 3). Hypotheses here enumerated are five times less than expected hypotheses of hybrid crosses among all species without any prior knowledge, as species distribution or reproductive ecology. Hypotheses from the co-occurrence patterns will be falsified empirically on the next section.

3.2 The reproductive ecology context

Results of crosses between *Phyllosoma* complex species (*sensu* Carcavallo et al. (2000) and Lent & Wygodzinsky (1979)) and relatives, available from the literature, are summarized on table 2. Information was recovered for 65% of 45 possible crosses, being scarce information about crosses between *T.* sp. aff. *dimidiata* and *T. dimidiata* with *Phyllosoma* complex species. Most of crosses between *Phyllosoma* complex species yield successful F2, proving viability and fertility of F1 parents (Table 2). Some exceptions are crosses between *T. bassolsae* with *T. mazzottii* and *T. phyllosoma*, *T. pallidipennis* with *T. mazzottii* and *T. phyllosoma*, and *T. picturata* with *T. mazzottii.* Although F1 adults can be obtained from crosses between *T. recurva* and most of

Phyllosoma complex species (Martínez-Ibarra et al., 2015), only crosses between *T. recurva* with *T. pallidipennis* and *T. phyllosoma* yield on successful F2 specimens (Table 2). Viable and fertile hybrids of sympatrid *T.* sp. aff. *dimidiata* and *T. dimidiata* are reported (Table 2), in contrast to non-successful crosses of allopatrid *T.* sp. aff. *dimidiata* and *T. dimidiata* (García et al., 2013). Finally, crosses between *T. mexicana* and other compared species (Table 2) cannot yield viable F1 specimens (Martínez-Ibarra et al., 2011a).

Most hypotheses of potential interbreeding among *Phyllosoma* complex species and relatives cannot be rejected (Figure 3). Only hypotheses of potential interbreeding among *T. pallidipennis* with *T. mazzottii* and *T. phyllosoma* are rejected on the spatio-temporal context considered. Therefore, the network represents validated hypotheses of interbreeding events occurring among *Phyllosoma* complex species and relatives in the considered spatio-temporal context. As a consequence, samples with introgressed DNA derived from interbreeding events from the spatio-temporal context considered are predicted: 1) *Triatoma* sp. aff. *dimidiata* and *T. dimidiata*, 2) *Triatoma mazzottii* and *T. phyllosoma*, 3) *Triatoma pallidipennis* and *T. longipennis*, 5) *Triatoma pallidipennis* and *T. picturata*, and 6) *Triatoma longipennis* and *T. picturata*. These predictions will be correlated with data from the phylogenetic context.

3.3 The phylogenetic context
3.3.1 Conflictive relationships between gene trees
Gene trees of ITS-2 and *Cyt b* of *Phyllosoma* complex specimens (*sensu* Carcavallo et al. (2000), and Lent & Wygodzinsky (1979)) and relatives differ in several ways (Figure 4). First, the ITS-2 gen tree recovers a monophyletic *Phyllosoma* complex, with the six species: *T. bassolsae, T. longipennis, T. mazzottii, T. pallidipennis, T. phyllosoma,* and *T. picturata*. The *Phyllosoma* complex is paraphyletic with respect to *T.* sp. aff. *dimidiata*, *T. dimidiata*, *T. gerstaeckeri*, *T. mexicana*, and *T. recurva* in the *Cyt b* gen tree. Second, most of the species are recovered as monophyletic in the ITS-2 gen tree, excluding *T. mazzottii*, which is paraphyletic with respect to *T. phyllosoma*, and *T. mexicana*, which is paraphyletic with respect to *T. gerstaeckeri*. Whereas *T.* sp. aff.

dimidiata, *T. mazzottii*, *T. mexicana*, and *T. picturata* are polyphyletic species according to the *Cyt b* gen tree. Finally, phylogenetic relationships of ITS-2 and *Cyt b* differ considerably. According to the ITS-2 gen tree, *T. recurva* is the sister group of the *Phyllosoma* complex; *Triatoma* sp. aff. *dimidiata* + *T. dimidiata*, and *T. gerstaeckeri* + *T. mexicana*, are in different clades; and *T. mazzottii* + *T. phyllosoma* clade is the sister group of the clade comprised of *T. bassolsae, T. longipennis, T. pallidipennis,* and *T. picturata*. Whereas, according to *Cyt b* gene tree, *T. recurva* is paraphyletic to *T. longipennis*, *T.* sp. aff. *dimidiata, T. dimidiata*, *T. gerstaeckeri,* and *T. mexicana*, are a clade, and *T. phyllosoma* is the sister group of the clade comprised of *T. bassolsae, T. mazzottii*, *T. mexicana, T. pallidipennis,* and *T. picturata*.

3.3.2 Detection of potential ancient introgression on samples

By an eco-biogeographic approach, potential ancient introgression was detected from samples of mtDNA of two species: *Triatoma picturata* and *T.* sp. aff. *dimidiata*. *Triatoma* sp. aff. *dimidiata* is a polyphyletic species according to the *Cyt b* gen tree, related to both *T. mexicana* + *T. gerstaeckaeri* clade and *T. dimidiata* (Figure 4). Interbreeding events occurring between *T.* sp. aff. *dimidiata* and *T. dimidiata* are expected, according to the spatio-temporal and reproductive ecology contexts. There is not another competing hypothesis, like interbreeding events occurring among *T.* sp. aff. *dimidiata* and *T. mexicana*. Therefore, it is inferred potential ancient introgression of *T. dimidiata*´s mtDNA on Tspaffdim_BZt1 sample. *Triatoma picturata* is a polyphyletic species according to the *Cyt b* gen tree, related to both the clade comprised of *T. mexicana, T. mazzottii,* and *T. bassolsae,* and the clade of *T. longipennis* + *T. recurva*. Interbreeding events occurring among *T. picturata* and *T. longipennis* are expected, according spatio-temporal and reproductive ecology contexts, and there is not another competing hypothesis. Therefore, potential ancient introgression on Tpicturata_MXny6 sample is inferred. A multispecies coalescent analysis was performed without Tspaffdim_BZt1 and Tpicturata_MXny6 samples (Supplementary material X). When specimens with potential ancient introgression were removed from the multispecies coalescence analysis, any

huge change was observed on the resulting species tree topology (Supplementary material X).

According to the *Cyt b* gen tree, *T. mexicana* is a polyphyletic species, related to *T. gerstaeckaeri* and *T. mazzottii*, as well as *T. mazzottii,* related to both *T. gerstaeckaeri* + *T. mexicana* clade and *T. mexicana* (Figure 4). *Triatoma mazzottii* and *T. mexicana* are not close related species, according ITS-2 gen tree. Due to introgression origin of samples of *T. mazzottii* and *T. mexicana* cannot be clarified by the spatio-temporal context considered here, the reproductive ecology context is considered exclusively. *Triatoma mazzottii* seems close related to *Phyllosoma* complex species, according their viable and fertile hybrids records (Table 2). *Triatoma mexicana* seems not close related to *Phyllosoma* complex species, according their non-viable hybrids records (Table 2). Therefore, Tmazzotti_MXgr38 and Tmexicana_MXgu1cs samples are inferred as introgression consequence. A multispecies coalescent analysis without Tmazzotti_MXgr38, Tmexicana_MXgu1cs, Tspaffdim_BZt1 and Tpicturata_MXny6 samples is explained on the next section.

3.3.3 Species tree

Most of the gene tree incongruence was resolved when samples without potential introgression were reanalyzed (Figure 5). *Phyllosoma* complex is a well-supported monophyletic group, with respect to the clade comprised of *T. dimidiata*, *T.* sp. *aff. dimidiata*, *T. gerstaeckeri* and *T. mexicana* (Figure 6). *Triatoma recurva* belongs to *Phyllosoma* complex and its sister-relationship with *T. longipennis* is relatively well supported. *Triatoma picturata* is the sister-group of *T. recurva* + *T. longipennis* clade. Another relatively well supported clade of *Phyllosoma* complex species tree hold *T. bassolsae, T. mazzottii, T. pallidipennis,* and *T. phyllosoma*. Posterior probability values of the inferred maximum clade credibility species tree are low (Figure 6). However, the posterior probability values improved with respect to the maximum clade credibility species tree considering samples with introgression (supplementary material X).

4. DISCUSSION

4.1 Ancient introgression from an ecological-biogeographical insight

Our eco-biogeographic approach enabled detection of samples of *Phyllosoma* complex with potential ancient introgression. This approach takes advantage of empirical fundamentals, as knowledge of reproductive ecology and species distributions. Relevance of species distribution co-occurrence for constraining the inference of hybridization events has been proposed previously (Dowling et al., 1997; Funk & Omland, 2003). However, spatial context has been not considered yet by available methods for inferring hybridization in a phylogenetic context. Therefore, hybridization in a phylogenetic context could be underestimated, by considering only species with current co-occurrence and hybridization (Gauthier & Lapointe, 2007; Gompert & Buerkle, 2009), or overestimated, by inferring ancient hybridization between all sampled taxa (Meng & Kubatko, 2009; Yu et al., 2014), no matter their spatial distribution. Here, the spatio-temporal context was useful for constraining the number of hypotheses of past hybridization events between *Phyllosoma* complex species. Plausibility of these hypotheses of interbreeding was asserted on the reproductive ecology context. As a result, from a relatively large set of hybridization events (all species vs. all species), a plausible and small set of samples was detected exhibiting potential ancient introgression.

Also, the relevance of a proper detection of introgression on samples of *Phyllosoma* complex was corroborated by our eco-biogeographic approach. This approach emphasizes the detection of introgression and inference of the species tree as independent processes. Following this two steps, we realized about the improvement of support values of the species tree of *Phyllosoma* complex without potential introgression. An improvement of species tree after removing potential introgression has been previously reported for cichlids exhibiting adaptive radiation (Meyer et al., 2016). 2017). Therefore, the phylogenetic inference would be benefited by any attempt of detection of introgression than just assuming lacking of introgression on samples. As introgression seems a common process causing gene tree incongruence in several taxa (García et al., 2017; Mas-Coma & Bargues, 2009; Meyer et al., 2016; Wagner et al.,

2017; Wallis et al., 2017), we reinforce the idea that phylogeneticists should be concerned about introgression.

Detection of introgression under our eco-biogeographic approach could be limited by lacking of environmental data. The only spatio-temporal context studied here was Last Glacial Maximum, but hybridization events between *Phyllosoma* complex species could have been occurred before of this temporal context. For instance, hybridization between *T. mazzottii* and *T. mexicana* was inferred considering incongruences ITS-2 and *Cyt b* gene tree. These species seem unclose related according ITS-2 gen tree, the published knowledge of their morphology (Carcavallo et al., 2000; Lent & Wygodzinsky, 1979; Salazar-Schettino, Rojas-Wastavino, Rosales-Piña, Vences-Blanco, & Cabrera-Bravo, 2013), and their reproductive biology (Martínez-Ibarra et al., 2011a). Given that *T. mazzottii* and *T. mexicana* are not able to produce viable hybrids in a current temporal context, we concluded that: 1) a recent introgression of samples, occasioned by possible unplanned crosses on colony handling, can be discarded; 2) the minimum requirement for reproductive isolation development between *T. mazzottii* and *T. mexicana* is the spatio-temporal co-occurrence (see 2.3 The reproductive ecology context); 3) a hypothesis of hybridization between *T. mazzottii* and *T. mexicana* was not formulated for Last Glacial Maximum spatio-temporal context, because these species did not co-occur under Last Glacial Maximum context; 4) the spatio-temporal co-occurrence of *T. mazzottii* and *T. mexicana* probably occur before of Last Glacial Maximum, given *T. mazzottii* and *T. mexicana* distributions are separated by the Transmexican Volcanic Belt. Considering divergence times of *Phyllosoma* complex species (Bargues et al., 2000; Ibarra-Cerdeña et al., 2014), the spatio-temporal co-occurrence of *T. mazzottii* and *T. mexicana* is presumed during the current morphology acquirement of the Belt (Late Pliocene) (Ferrari, Orozco-Esquivel, Manea, & Manea, 2012). Unfortunately, environmental data for such spatio-temporal context are not available with the required quality. However, the biogeographic data might allow a recovering of the spatio-temporal context of hybridization events. Data from biogeography of areas might supply lacking of environmental data of very ancient contexts, for future attempts to detect ancient introgression.

An interesting topic of ancient introgression detection is the temporal context of hybridization events. Ancient hybridization events inferred at Last Glacial Maximum seems to tangle phylogenetic relationships of close species. Whereas, the species tree of *Phyllosoma* complex would be entangled at deep phylogenetic relationships level, by probably hybridization events before of Last Glacial Maximum. Hypothesis of hybridization at Last Glacial Maximum are *T. picturata* with *T. bassolsae, T. longipennis*, and *T. mazzottii,* and *T.* sp. aff. *dimidiata* with *T. dimidiata, T. mexicana*, and *T. gerstaeckaeri*. All these species seem close related according their reproductive biology, and the published knowledge of their morphology (Carcavallo et al., 2000; Lent & Wygodzinsky, 1979). When specimens of *T. picturata* and *T.* sp. aff. *dimidiata* with potential ancient introgression were removed from the multispecies coalescence analysis, few changes were observed on the resulting species tree topology. Otherwise, spurious relationship between *T. mazzottii* and *T. mexicana* causes most of the incongruence between ITS-2 and *Cyt b* gene trees, being disentangled when specimens of *T. mazzottii* and *T. mexicana* with potential ancient introgression were removed from the multispecies coalescence analysis. Potential hybridization events before the full formation of the Transmexican Volcanic Belt are presumed between *T. mazzottii* and *T. mexicana*. Therefore, temporal sequence of ancient hybridization events seems to tangle phylogenetic relationships at different levels of complexity.

The limited number of loci and samples of our phylogenetic analysis were not an impediment for detection of introgression under our eco-biogeographic approach. We were able to detect ancient introgression using just two unlinked loci exhibiting different inheritance strategies. Even if a large number of unlinked loci were sampled, most of the irreconcilable differences among gene trees have been detected between biparental inheritance nuclear loci and clonal inheritance loci (García et al., 2017; Wallis et al., 2017). To avoid sampling of redundant information, detection of introgression with few loci would be an advisable initial step of a design of a phylogenetic study. So that, after samples with and without potential introgression were detected, sequencing efforts would be focus on a large number of loci of samples without potential introgression. Also, samples of *Phyllosoma* complex with potential ancient introgression were

detected, even though the small sample size of the studied specimens. We presume that introgression in specimens of *Phyllosoma* complex would be very common, such as other complexes of Triatominae (Mas-Coma & Bargues, 2009). Therefore, for future studies, we suggest considering a larger sample size of specimens than the sample size here studied. An intensive specimen sampling would improve the chance of an adequate inference of the demographic processes describing ancient introgression. Thus, rather than a huge effort on loci sampling, we suggest an improvement on specimen sampling effort for future studies attempting detection of introgression.

4.2 Phylogeny of *Phyllosoma* complex

Our study provides the first thorough analysis of potential factors causing incongruence between gene trees for species of *Phyllosoma* complex. These factors are horizontal gene transfer –or hybridization-, incomplete lineage sorting, extinction, and gene duplication –causing pseudogenes- (Maddison, 1997). Ancient hybridization and incomplete lineage sorting seem causative of most of the incongruence between ITS-2 and *Cyt b* gene trees of *Phyllosoma* complex. Even more, the monophyly of *Phyllosoma* complex was only clarified after samples without potential introgression were reanalyzed. Extinction was also modeled, by considering a birth-death model of speciation in the multispecies coalescent analysis. However, extinction seems an unlikely process in the evolutionary history of the *Phyllosoma* complex, given the data. This discovery seems feasible because species of the complex have evolved recently (Bargues et al., 2000; Ibarra-Cerdeña, Zaldívar-Riverón, Peterson, Sánchez-Cordero, & Ramsey, 2014). Finally, data quality (sequences and alignments) was assessed with special care to minimize the noise from pseudogenes. Given the wide evolutionary insight of our analysis, our study provides the most comprehensive phylogeny of the *Phyllosoma* complex until now.

The *Phyllosoma* complex is a well-supported monophyletic group. This complex comprises the six species considered *a priori* (*Triatoma bassolsae, T. longipennis, T. mazzottii, T. pallidipennis, T. phyllosoma,* and *T. picturata*) (Carcavallo et al., 2000; Lent & Wygodzinsky, 1979), plus *T. recurva*. The inclusion of *T. recurva* in the *Phyllosoma*

complex, and its sister relationship with *T. longipennis*, were initially proposed considering *Cyt b* information (Pfeiler, Bitler, Ramsey, Palacios-Cardiel, & Markow, 2006). Incorporation of *T. recurva* to the *Phyllosoma* complex is also supported by its reproductive compatibility with *T. phyllosoma* and *T. pallidipennis* (Martínez-Ibarra et al., 2015). Nevertheless, our ITS-2 gen tree recovers to *T. recurva* as sister species of the clade comprised of the remaining six species of the *Phyllosoma* complex, being in agreement with Espinoza et al.'s (2013) finding. Here, *T. recurva* is confirmed as sister species of *T. longipennis* in the phylogeny of *Phyllosoma* complex.

Our proposal of the *Phyllosoma* complex differs from other proposals, mainly because disagreement about species complex treatment. We treat a species complex as a group of nearly allopatrid species, with reproductive compatibility and similar morphology (Carcavallo et al. 2000; Usinger, 1944; Usinger, Wygodzinsky, Ryckman, 1966). These features are fulfilled by the seven species of the *Phyllosoma* complex. Other species have been proposed as part of the *Phyllosoma* complex, considering the most inclusive monophyletic group (de la Rúa et al., 2014; Espinoza et al., 2013; Martínez et al., 2006). However, not all monophyletic groups are species complexes, even though monophyly is an expectable pattern in a species complex. Therefore, we exclude *T. dimidiata*, *T.* sp. *aff. dimidiata*, *T. gerstaeckeri*, and *T. mexicana* of the *Phyllosoma* complex (*sensu* de la Rúa et al., 2014). These species are easily distinguished from species from the *Phyllosoma* complex proposed here, by morphological characters, and do not produce viable hybrids (Carcavallo et al., 2000; Martínez-Ibarra et al. 2011a). Females of *T. mexicana* resemble to species of *Phyllosoma* complex, by its abdomen strongly widened (Lent & Wygodzinsky, 1979), but this similarity seems superficial. There are several differences between *T. mexicana* and *Phyllosoma* complex species, as pronotum shape, measures of female body, male genitalia, and biological parameters (Martínez-Ibarra, et al., 2008; Salazar-Schettino et al., 2013). *Triatoma hegneri* is another species mentioned as part of the *Phyllosoma* complex by de la Rúa et al. (2014). This insular species has been suggested as a melanic polymorphism of *T. dimidiata*, because *T. hegneri* is embedded in *T. dimidiata* clade (Bargues et al., 2008; Justi et al., 2018). We could not include *T. hegneri* in our

analysis, but we agree with de la Rúa et al. (2014) in its treatment as full species. The paraphyly of *T. dimidiata*, occasioned by *T. hegneri* treatment as full species, has not any detriment in *T. dimidiata* status as full species, according the species concept followed here. Also, in addition to morphological differences, *Triatoma hegneri* exhibits reproductive incompatibility with *T. dimidiata* (Mazzottii, 1943). We were no able to find information about crosses between *T. hegneri* and species of *Phyllosoma* complex, but *T. hegneri* is distinguished from species of the *Phyllosoma* complex proposed here by morphological characters. Therefore, we also exclude *T. hegneri* from the *Phyllosoma* complex (*sensu* de la Rúa et al., 2014).

The phylogenetic relationships between the species of the *Phyllosoma* complex are more resolved in our study than in other previous studies, although these relationships are weakly supported by Bayesian posterior probabilities of our multispecies coalescence analysis. Multispecies coalescence analyses usually have lower support values than concatenated analyses, because depicting incongruences between genealogical histories of genes (Kubatko & Degnan, 2007). Our ecological-biogreoaphical approach keeps the impact of the most of the potential factors causing gene tree incongruences (gene flow, gene duplication, extinction) as low as possible in our data. Thus, low support values would indicate a requirement of a better model of incomplete lineage sorting processes of the *Phyllosoma* complex, by a sampling of a large number of unlinked loci. Even thought our model of incomplete lineage sorting processes of the *Phyllosoma* complex is still incipient, our model gets an appropriate consideration of variability of previously neglected mitochondrial data. Mitochondrial (*Cyt b*) and nuclear (ITS-2) DNA data have been analyzed together under a concatenated approach, obtaining entangled phylogenies (Espinoza et al., 2013; Martínez et al., 2006). As ITS-2 gen tree have seem congruent with the biological knowledge of *Phyllosoma* complex, ITS-2 phylogeny have been the preferred data for phylogenetic hypothesis in spite of its low resolution (de la Rúa et al., 2014; Marcilla et al., 2001; Mas-Coma & Bargues, 2009). Here, we provide a different and fully resolved phylogenetic hypothesis of *Phyllosoma* complex, according the two gene trees (ITS-2, *Cyt b*), and incomplete lineage sorting and introgression processes.

4.2 Conclusions

Our study advocates for an integrative consideration of processes causing gene tree incongruences, to better understanding of the complex evolutionary history of taxa, like *Phyllosoma* complex species. An appropriate detection of introgression on samples seems as important as incomplete lineage sorting modeling to inferring a phylogeny. Introgression seems common not only in *Phyllosoma* complex, but also in other triatomines, and complexes of several animal and plants. But, to difference of incomplete lineage sorting, detection of introgression is beneficed by an explicit ecological-biogegraphical background. For instance, our ecological-biogegraphical approach helped us to detecting potential introgression on a plausible set of samples of *Phyllosoma* complex species and relatives. We suggest keeping in mind the association between hybridization events and spatio-temporal contexts, and empirical knowledge like reproductive biology, to improve reality of evolutionary processes modeling.

AUTHOR CONTRIBUTIONS

L.R-C., F.A-F., P.M.S-S., and J.J.M. conceived and designed this study. L.R-C. and J.L.T-R. compiled and curated data for the environmental niche models. L.R-C. and F.A-F. performed ecological niche modeling. L.R-C., F.M-H., P.M.S-S., and G.V. reviewed reproductive ecology literature. L.R-C., F.M-H., and G.V. performed molecular experiments and analyses. L.R-C. wrote the manuscript, with contributions of J.J.M. All authors read and approved the final manuscript.

REFERENCES

Bargues, M. D., Klisiowicz, D. R., Gonzalez-Candelas, F., Ramsey, J. M., Monroy, C., Ponce, C., … Mas-Coma, S. (2008). Phylogeography and genetic variation of *Triatoma dimidiata*, the main Chagas disease vector in Central America, and its position within the genus *Triatoma*. *PLoS Neglected Tropical Diseases*, *2*, e233.

Bargues, M. D., Marcilla, A., Ramsey, J. M., Dujardin, J.-P., Schofield, C. J., & Mas-Coma, S. (2000). Nuclear rDNA-based molecular clock of the evolution of Triatominae (Hemiptera: Reduviidae), vectors of Chagas disease. *Memórias do Instituto Oswaldo Cruz*, *95*, 567–573.

Bargues, M. D., Zuriaga, M. A., & Mas-Coma, S. (2014). Nuclear rDNA pseudogenes in Chagas disease vectors: evolutionary implications of a new 5.8S+ITS-2 paralogous sequence marker in triatomines of North, Central and northern South America. *Infection, Genetics and Evolution*, *21*, 134–156.

Barve, N., Barve, V., Jiménez-Valverde, A., Lira-Noriega, A., Maher, S. P., Peterson, A. T., … Villalobos, F. (2011). The crucial role of the accessible area in ecological niche modeling and species distribution modeling. *Ecological Modelling*, *222*, 1810–1819.

Carcavallo, R. U., Jurberg, J., Lent, H., Noireau, F. & Galvão, C. (2000). Phylogeny of the Triatominae (Hemiptera, Reduviidae). Proposals for taxonomic arrangements. *Entomología y Vectores, 7*, 1-99.

Chaudhary, R., Burleigh, J. G., & Fernández-Baca, D. (2013). Inferring species trees from incongruent multi-copy gene trees using the Robinson-Foulds distance. *Algorithms for Molecular Biology, 8,* 28.

Collins, W. D., Bitz, C. M., Blackmon, M. L., Bonan, G. B., Bretherton, C. S., Carton, J. A., … Smith, R. D. (2006). The community climate system model version 3 (CCSM3). *Journal of Climate, 19,* 2122–2143.

de la Rúa, N. M., Bustamante, D. M., Menes, M., Stevens, L., Monroy, C., Kilpatrick, C. W., … Dorn, P. L. (2014). Towards a phylogenetic approach to the composition of species complexes in the North and Central American *Triatoma*, vectors of Chagas disease. *Infection, Genetics and Evolution*, *24*, 157–166.

Dorn, P. L., Calderon, C., Melgar, S., Moguel, B., Solorzano, E., Dumonteil, E., … Monroy, C. (2009). Two distinct *Triatoma dimidiata* (Latreille, 1811) taxa are found in sympatry in Guatemala and Mexico. *PLoS Neglected Tropical Diseases*, *3*(3), e393.

Dorn, P. L., de la Rúa, N. M., Axen, H., Smith, N., Richards, B. R., Charabati, J., … Stevens, L. (2016). Hypothesis testing clarifies the systematics of the main Central American Chagas disease vector, *Triatoma dimidiata* (Latreille, 1811), across its geographic range. *Infection, Genetics and Evolution*, *44*, 431–443.

Dowling, T. E., & Secor, L, C. (1997). The role of hybridization and introgression in the diversification of animals. *Annual Review of Ecology and Systematics*, *28*, 593–619.

Espinoza, B., Martínez-Ibarra, J. A., Villalobos, G., de La Torre, P., Laclette, J. P., & Martínez-Hernández, F. (2013). Genetic variation of north american triatomines (Insecta: Hemiptera: Reduviidae): initial divergence between species and

populations of Chagas disease vector. *The American Journal of Tropical Medicine and Hygiene*, *88*, 275–284.

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.

Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, *37*, 4302–4315.

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.

García, M., Menes, M., Dorn, P. L., Monroy, C., Richards, B., Panzera, F., & Bustamante, D. M. (2013). Reproductive isolation revealed in preliminary crossbreeding experiments using field collected *Triatoma dimidiata* (Hemiptera: Reduviidae) from three ITS-2 defined groups. *Acta Tropica*, *128*, 714–718.

García, N., Folk, R. A., Meerow, A. W., Chamala, S., Gitzendanner, M. A., de Oliveira, R. S., … Soltis, P. S. (2017). Deep reticulation and incomplete lineage sorting obscure the diploid phylogeny of rain-lilies and allies (Amaryllidaceae tribe Hippeastreae). *Molecular Phylogenetics and Evolution*, *111*, 231–247.

Gauthier, O., & Lapointe, F.-J. (2007). Hybrids and phylogenetics revisited: a statistical test of hybridization using quartets. *Systematic Botany*, *32*, 8–15.

Gompert, Z., & Buerkle, C. A. (2009). A powerful regression-based method for admixture mapping of isolation across the genome of hybrids. *Molecular Ecology*, *18*, 1207–1224.

Heled, J., & Drummond, A. J. (2010). Bayesian inference of species trees from multilocus data. *Molecular Biology and Evolution*, *27*, 570–580.

Herrera-Aguilar, M., Be-Barragán, L. A., Ramirez-Sierra, M. J., Tripet, F., Dorn, P., & Dumonteil, E. (2009). Identification of a large hybrid zone between sympatric sibling species of *Triatoma dimidiata* in the Yucatan peninsula, Mexico, and its epidemiological importance. *Infection, Genetics and Evolution*, *9*(6), 1345–1351.

Holder, M. T., Anderson, J. A., & Holloway, A. K. (2001). Difficulties in detecting hybridization. *Systematic Biology, 50,* 978–982.

Ibarra-Cerdeña, C. N., Zaldívar-Riverón, A., Peterson, A. T., Sánchez-Cordero, V., & Ramsey, J. M. (2014). Phylogeny and niche conservatism in North and Central American triatomine bugs (Hemiptera: Reduviidae: Triatominae), vectors of Chagas' disease. *PLoS Neglected Tropical Diseases*, *8*, e3266.

Jones, G., Aydin, Z., & Oxelman, B. (2015). DISSECT: an assignment-free Bayesian discovery method for species delimitation under the multispecies coalescent. *Bioinformatics*, *31*, 991–998.

Justi, S. A., Cahan, S., Stevens, L., Monroy, C., Lima-Cordón, R., & Dorn, P. L. (2018). Vectors of diversity: genome wide diversity across the geographic range of the Chagas disease vector *Triatoma dimidiata sensu lato* (Hemiptera: Reduviidae). *Molecular Phylogenetics and Evolution, 120,* 144–150.

Kubatko, L. S., & Degnan, J. H. (2007). Inconsistency of phylogenetic estimates from concatenated data under coalescence. *Systematic Biology*, *56*, 17–24.

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.

Lent, H., & Wygodzinsky, P. (1979). Revision of the Triatominae (Hemiptera, Reduviidae) and their significance as vector of Chagas' disease. *Bulletin of the American Museum of Natural History, 163,* 123–520.

Maddison, W. P. (1997). Gene trees in species trees. *Systematic Biology*, *46*(3), 523.

Mallet, J. (2005). Hybridization as an invasion of the genome. *Trends in Ecology & Evolution*, *20*, 229–237.

Marcilla, A., Bargues, M. D., Ramsey, J. M., Magallon-Gastelum, E., Salazar-Schettino, P. M., Abad-Franch, F., … Mas-Coma, S. (2001). The ITS-2 of the nuclear rDNA as a molecular marker for populations, species, and phylogenetic relationships in Triatominae (Hemiptera: Reduviidae), vectors of Chagas disease. *Molecular Phylogenetics and Evolution*, *18*, 136–142.

Martínez, F. H., Villalobos, G. C., Cevallos, A. M., de la Torre, P. D., Laclette, J. P., Alejandre-Aguilar, R., & Espinoza, B. (2006). Taxonomic study of the *Phyllosoma* complex and other triatomine (Insecta: Hemiptera: Reduviidae) species of epidemiological importance in the transmission of Chagas disease: using ITS-2 and mtCytB sequences. *Molecular Phylogenetics and Evolution*, *41*, 279–287.

Martínez-Hernández, F., Martínez-Ibarra, J. A., Catalá, S., Villalobos, G., de la Torre, P., Laclette, J. P., … Espinoza, B. (2010). Natural crossbreeding between sympatric species of the *Phyllosoma* Complex (Insecta: Hemiptera: Reduviidae) indicate the existence of only one species with morphologic and genetic variations. *The American Journal of Tropical Medicine and Hygiene*, *82*, 74–82.

Martínez-Ibarra, J. A., Grant-Guillén, Y., Delgadillo-Aceves, I. N., Zumaya-Estrada, F. A., Rocha-Chávez, G., Salazar-Schettino, P. M., … Nogueda-Torres, B. (2011a). Biological and genetic aspects of crosses between phylogenetically close species of Mexican Triatomines (Hemiptera: Reduviidae). *Journal of Medical Entomology*, *48*, 705–707.

Martínez-Ibarra, J. A., Grant-Guillén, Y., Ventura-Rodríguez, L. V., Osorio-Pelayo, P. D., Macías-Amezcua, M. D., Meillón-Isáis, K., … Nogueda-Torres, B. (2011b). Biological and genetic aspects of crosses between species of the genus *Meccus* (Hemiptera: Reduviidae Triatominae). *Memórias do Instituto Oswaldo Cruz*, *106*, 293–300.

Martínez-Ibarra, J. A., Nogueda-Torres, B., García-Benavídez, G., Vargas-Llamas, V., Bustos-Saldaña, R., & Montañez-Valdez, O. D. (2012). Bionomics of populations of *Meccus pallidipennis* (Stål), 1872 (Hemiptera: Reduviidae) from Mexico. *Journal of Vector Ecology*, *37*, 474–477.

Martínez-Ibarra, J. A., Nogueda-Torres, B., González, E. P., Alejandre-Aguilar, R., Solorio-Cibrián, M., Barreto, S. P., … Trujillo-García, J. C. (2005). Development of *Triatoma rubida sonoriana, Triatoma barberi,* and *Meccus mazzottii* (Heteroptera, Reduviidae) under laboratory conditions. *Journal of the American Mosquito Control Association*, *21*, 310.

Martínez-Ibarra, J. A., Nogueda-Torres, B., Licón-Trillo, Á., Alejandre-Aguilar, R., Salazar-Schettino, P. M., & Vences-Blanco, M. O. (2015). Biological aspects of

crosses between *Triatoma recurva* (Stål), 1868 (Hemiptera: Reduviidae: Triatominae) and other members of the *Phyllosoma* complex. *Journal of Vector Ecology*, *40*, 117–122.

Martínez-Ibarra, J. A., Nogueda-Torres, B., Licón-Trillo, Á., Villagrán-Herrera, M. E., de Diego-Cabrera, J. A., Montañez-Valdez, O. D., & Rocha-Chávez, G. (2013). Comparative bionomics of four populations of *Meccus longipennis* (Hemiptera: Reduviidae: Triatominae) under laboratory conditions. *Memorias do Instituto Oswaldo Cruz*, *108*, 239–242.

Martínez-Ibarra, J. A., Salazar-Schettino, P. M., Nogueda-Torres, B., Vences, M. O., Tapia-González, J. M., & Espinoza-Gutiérrez, B. (2009). Occurrence of hybrids and laboratory evidence of fertility among three species of the *Phyllosoma* complex (Hemiptera: Reduviidae) in Mexico. *Memórias Do Instituto Oswaldo Cruz*, *104*, 1125–1131.

Martínez-Ibarra, J. A., Salazar-Schettino, P. M., Solorio-Cibrián, M., Cabrera-Bravo, M., Novelo-López, M., Vences-Blanco, M. O., … Nogueda-Torres, B. (2008). Influence of temperature and humidity on the biology of *Triatoma mexicana* (Hemiptera: Reduviidae: Triatominae) under laboratory conditions. *Memórias do Instituto Oswaldo Cruz, 103,* 719–723.

Martínez-Ibarra, J. A., Ventura-Rodríguez, L. V., Meillon-Isais, K., Barajas-Martínez, H., Alejandre-Aguilar, R., Lupercio-Coronel, P., … Nogueda-Torres, B. (2008). Biological and genetic aspects of experimental hybrids from species of the *Phyllosoma* complex (Hemiptera: Reduviidae: Triatominae). *Memorias do Instituto Oswaldo Cruz*, *103*, 236–243.

Mas-Coma, S., & Bargues, M. D. (2009). Populations, hybrids and the systematic concepts of species and subspecies in Chagas disease triatomine vectors inferred from nuclear ribosomal and mitochondrial DNA. *Acta Tropica*, *110*, 112–136.

Mazzotti, L. (1943). Estudios sobre *Triatoma hegneri*. *Rev. Inst. Salub. Enf. Trop, 4,* 53-56.

Meng, C., & Kubatko, L. S. (2009). Detecting hybrid speciation in the presence of incomplete lineage sorting using gene tree incongruence: A model. *Theoretical Population Biology*, *75*, 35–45.

Meyer, B. S., Matschiner, M., & Salzburger, W. (2016). Disentangling incomplete lineage sorting and introgression to refine species-tree estimates for lake Tanganyika cichlid fishes. *Systematic Biology*, syw069.

Mirarab, S., Reaz, R., Bayzid, M. S., Zimmermann, T., Swenson, M. S., Warnow, & T. 2014. ASTRAL: genome-scale coalescent-based species tree estimation. *Bioinformatics 30,* i541–i548.

Monteiro, F. A., Peretolchina, T., Lazoski, C., Harris, K., Dotson, E. M., Abad-Franch, F., … Marcet, P. L. (2013). Phylogeographic pattern and extensive mitochondrial DNA divergence disclose a species complex within the Chagas Disease vector *Triatoma dimidiata*. *PLoS ONE*, *8*, e70974.

Morrone, J. J., Escalante, T., & Rodríguez-Tapia, G. (2017). Mexican biogeographic provinces: map and shapefiles. *Zootaxa*, *4277*, 277.

Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V., Underwood, E. C., … Kassem, K. R. (2001). Terrestrial ecoregions of the world: a new map of life on earth: a new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity. *BioScience*, *51*, 933–938.

Quisberth, S., Waleckx, E., Monje, M., Chang, B., Noireau, F., & Brenière, S. F. (2011). —Andean" and —non-Andean" ITS-2 and mtCytB haplotypes of *Triatoma infestans* are observed in the Gran Chaco (Bolivia): population genetics and the origin of reinfestation. *Infection, Genetics and Evolution*, *11*(5), 1006–1014.

Peterson, A. T., Soberón, J., Pearson, R. G., Anderson, R. P., Martínez-Meyer, E., Nakamura, M., & Araújo, M. B. (Ed.). (2011). *Ecological niches and geographic distributions*. Princeton, N.J: Princeton University Press.

Pfeiler, E., Bitler, B. G., Ramsey, J. M., Palacios-Cardiel, C., & Markow, T. A. (2006). Genetic variation, population structure, and phylogenetic relationships of *Triatoma rubida* and *T. recurva* (Hemiptera: Reduviidae: Triatominae) from the Sonoran Desert, insect vectors of the Chagas' disease parasite *Trypanosoma cruzi*. *Molecular Phylogenetics and Evolution*, *41*, 209–221.

Ramsey, J. M., Peterson, A. T., Carmona-Castro, O., Moo-Llanes, D. A., Nakazawa, Y., Butrick, M., … Ibarra-Cerdeña, C. N. (2015). Atlas of Mexican Triatominae (Reduviidae: Hemiptera) and vector transmission of Chagas disease. *Memórias do Instituto Oswaldo Cruz*, *110*, 339–352.

Ritchie, A. M., Lo, N., & Ho, S. Y. W. (2016). The impact of the tree prior on molecular dating of data sets containing a mixture of inter- and intraspecies sampling. *Systematic Biology*, syw095.

Ruddiman, W. F. (2008). *Earth's climate: past and future* (2nd ed). New York: W.H. Freeman.

Salazar-Schettino, P. M., Rojas-Wastavino, G. E., Rosales-Piña, J. S., Vences-Blanco, M. O. V., & Cabrera-Bravo, M. (2013). *Triatoma mexicana* Herrich-Schaeffer (Hemiptera: Reduviidae: Triatominae) descripción de la genitalia externa del macho y morfología externa de la hembra. *EntomoBrasilis*, *6*, 68–73.

Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual* (2nd ed). New York: Cold Spring Harbor Laboratory Press.

Schofield, C. J. (1988). The biosystematics of Triatominae. In M. W. Service (Ed.) *Biosystematics of haematophagous Insects.* Clarenden, Oxford: Systematics Association Special Volume 37.

Stephens, C. R., Giménez-Heau, J., González, C., Ibarra-Cerdeña, C. N., Sánchez-Cordero, V., & González-Salazar, C. (2009). Using biotic interaction networks for prediction in biodiversity and emerging diseases. *PLoS ONE*, *4*, e5725.

Stephens, C., Sánchez-Cordero, V., & González-Salazar, C. (2017). Bayesian inference of ecological interactions from spatial data. *Entropy*, *19*, 547.

Tamay-Segovia, P., Alejandre-Aguilar, R., Martínez, F., Villalobos, G., de La Torre, P., Laclette, J. P., … Espinoza, B. (2008). Two *Triatoma dimidiata* clades (Chagas disease vector) associated with different habitats in southern Mexico and Central America. *The American Journal of Tropical Medicine and Hygiene*, *78*, 472–478.

Templeton, A. R. (1989). The meaning of species and speciation – a genetic perspective. In D. Otte, & J. A. Endler (eds.) *Speciation and its Consequences*. Sunderland, Massachusetts: Sinauer Associates.

Usinger, R. L. (1944). The Triatominae of North and Central America and the West Indies and their public health significance. *Public Health Bulletin, 288,* 1-83.

Usinger, R. L., Wygodzinsky, P., Ryckman, R. E. (1966). The biosystematics of Triatominae. *Annual Review of Entomology, 11,* 309-330.

Villalobos, G., Martínez-Hernández, F., de la Torre, P., Laclette, J. P., & Espinoza, B. (2011). Entomological indices, feeding sources, and molecular identification of *Triatoma phyllosoma* (Hemiptera: Reduviidae) one of the main vectors of Chagas disease in the istmo de Tehuantepec, Oaxaca, Mexico. The American Journal of Tropical Medicine and Hygiene, 85(3), 490–497. doi:10.4269/ajtmh.2011.10-0508

Wagner, F., Härtl, S., Vogt, R., & Oberprieler, C. (2017). ―Ⅺme another marguerite!": species delimitation in a group of intensively hybridizing lineages of ox-eye daisies (*Leucanthemum* Mill., Compositae-Anthemideae). *Molecular Ecology*, *26*, 4260–4283.

Wallis, G. P., Cameron-Christie, S. R., Kennedy, H. L., Palmer, G., Sanders, T. R., & Winter, D. J. (2017). Interspecific hybridization causes long-term phylogenetic discordance between nuclear and mitochondrial genomes in freshwater fishes. *Molecular Ecology*, *26*, 3116–3127.

Yu, Y., Dong, J., Liu, K. J., & Nakhleh, L. (2014). Maximum likelihood inference of reticulate evolutionary histories. *Proceedings of the National Academy of Sciences*, *111*, 16448–16453.

FIGURES

FIGURE 1 Detection of ancient hybridization in a phylogeny from an eco-biogeographycal approach.

FIGURE 2 Recent areas of distribution of *Phyllosoma* complex species and relatives. (Map source: EROS. 2010. https://databasin.org/datasets/d2198be9d2264de19cb93fe6a380b69c)

FIGURE 3 Hypotheses of potential interbreeding among species of *Phyllosoma* complex and relatives.

FIGURE 4 Maximum clade credibility gene trees of the complete data set of the *Phyllosoma* complex.

FIGURE 5 Maximum clade credibility gene trees of the nonintrogressed data set of the *Phyllosoma* complex.

FIGURE 6 Maximum clade credibility species tree of the nonintrogressed data set of the *Phyllosoma* complex.

FIGURE 1
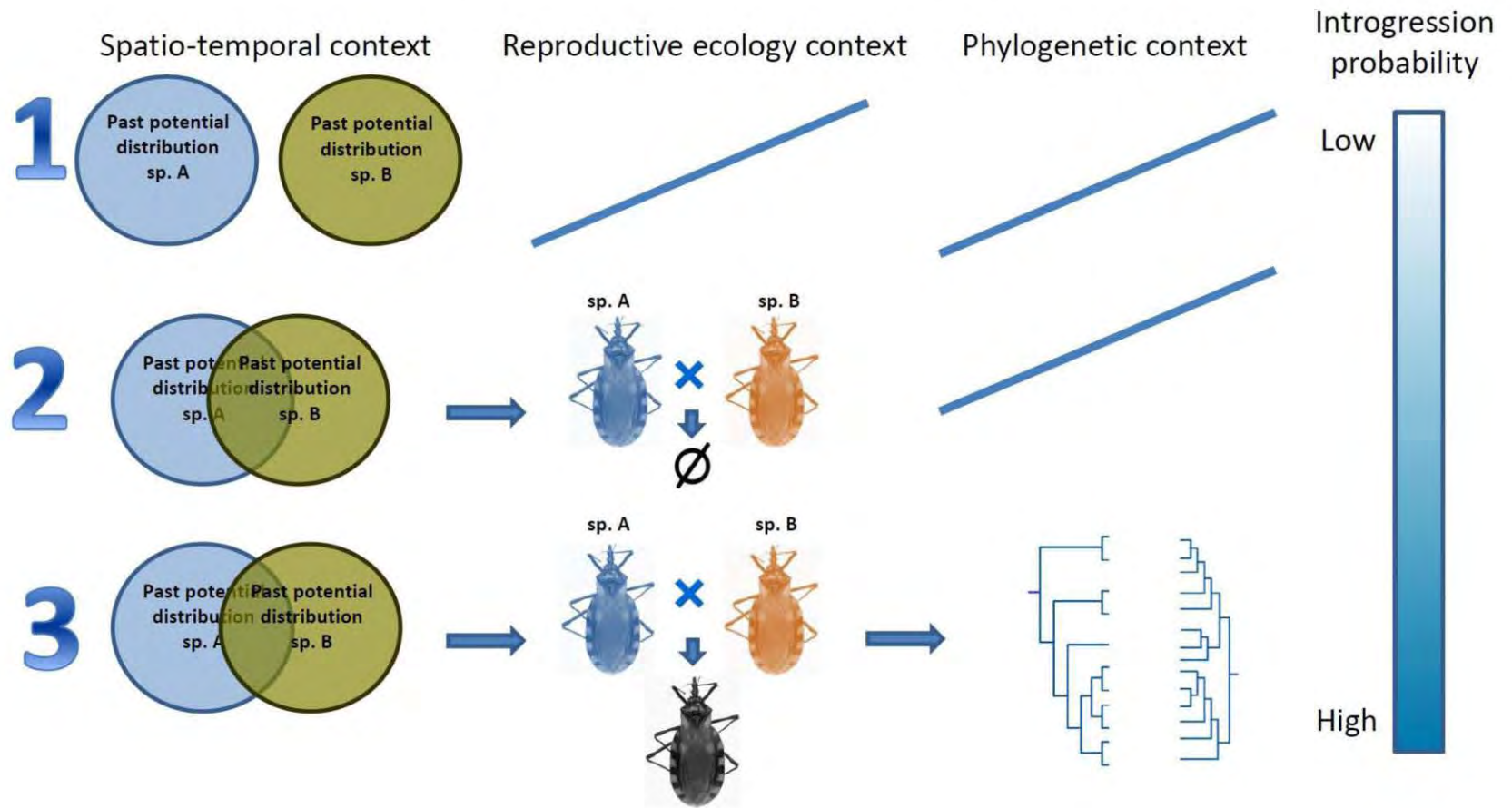
FIGURE 2

FIGURE 3

FIGURE 4

ITS-2

Cyt b

FIGURE 5

ITS-2

Cyt b
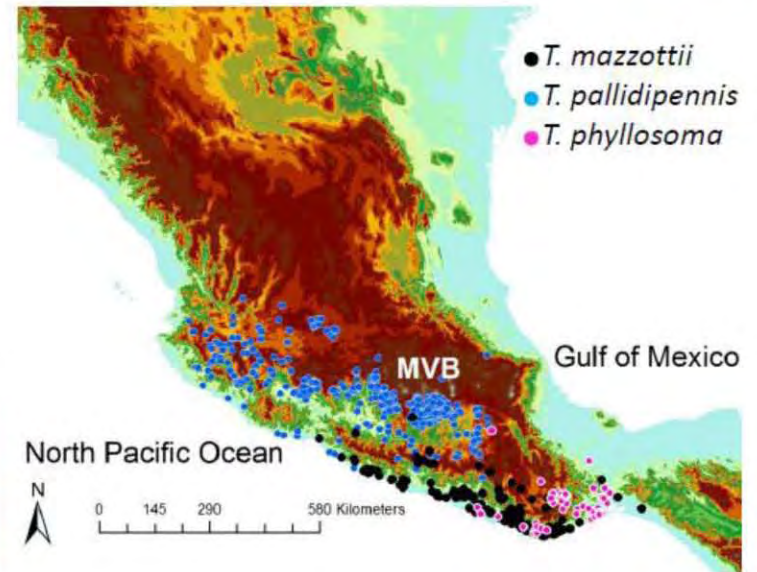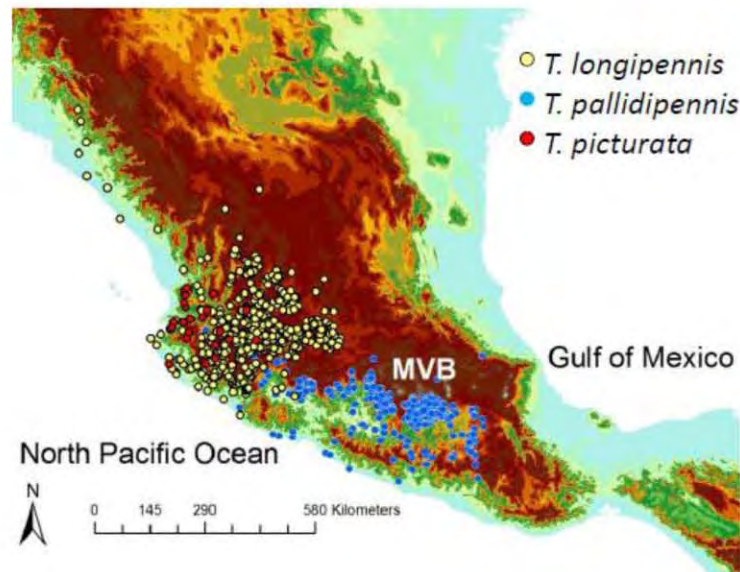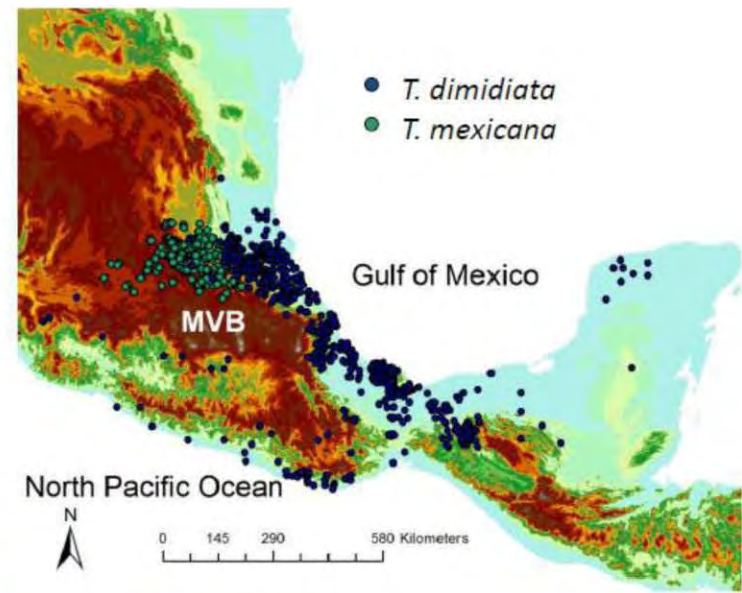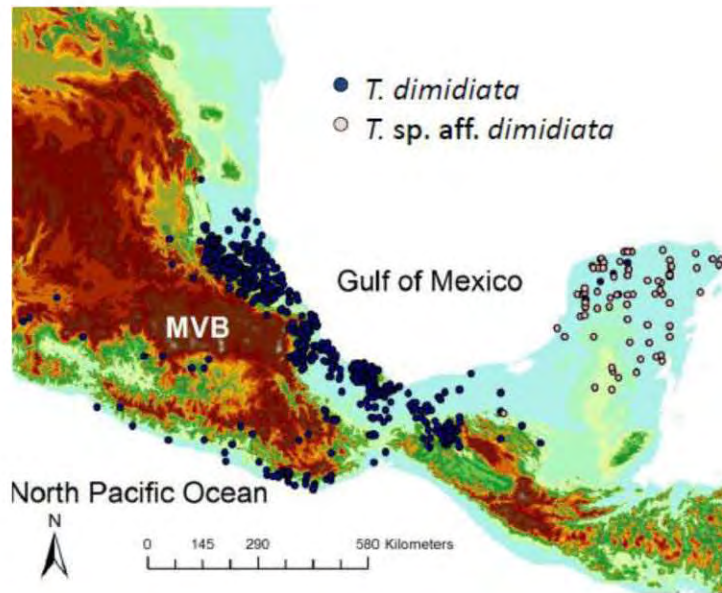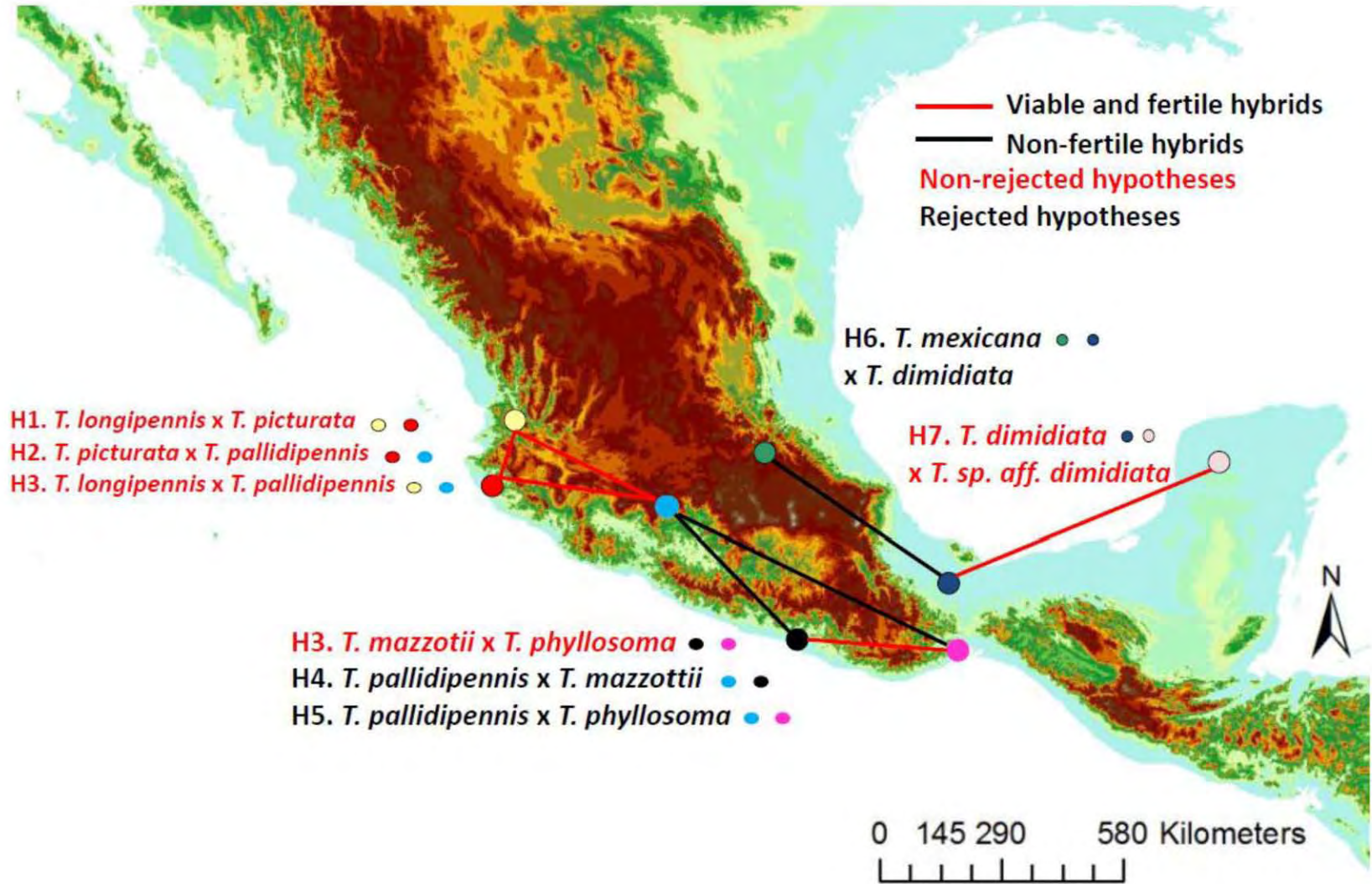
FIGURE 6

TABLES

TABLE 1 *Phyllosoma* complex specimens and relatives used for the sequencing.

TABLE 2 Experimental crosses of *Phyllosoma* complex specimens and relatives resulting in viable and fertile F1 (circles) and in non-viable or non-fertile F1 (triangles) specimens.

TABLE 1

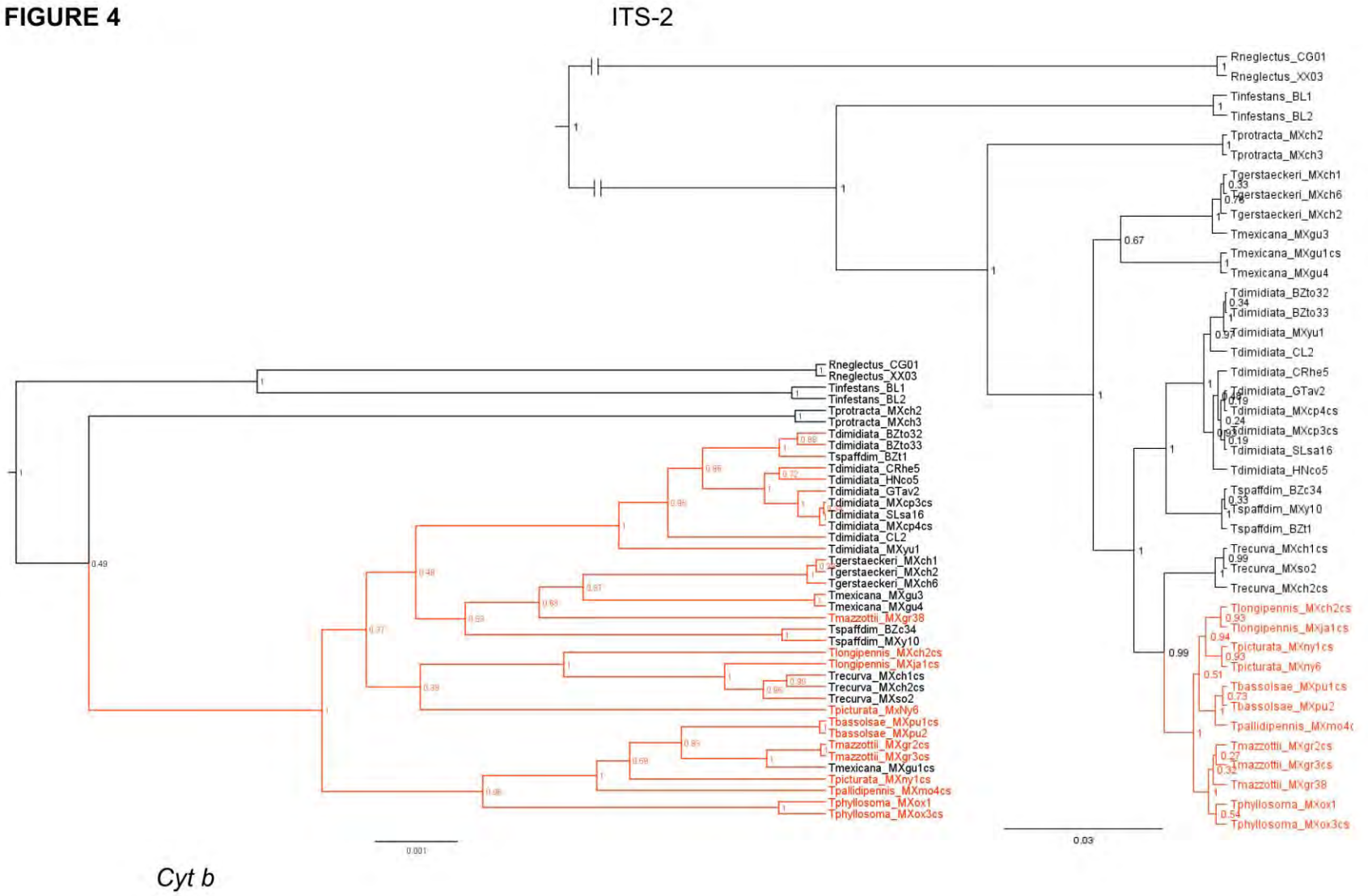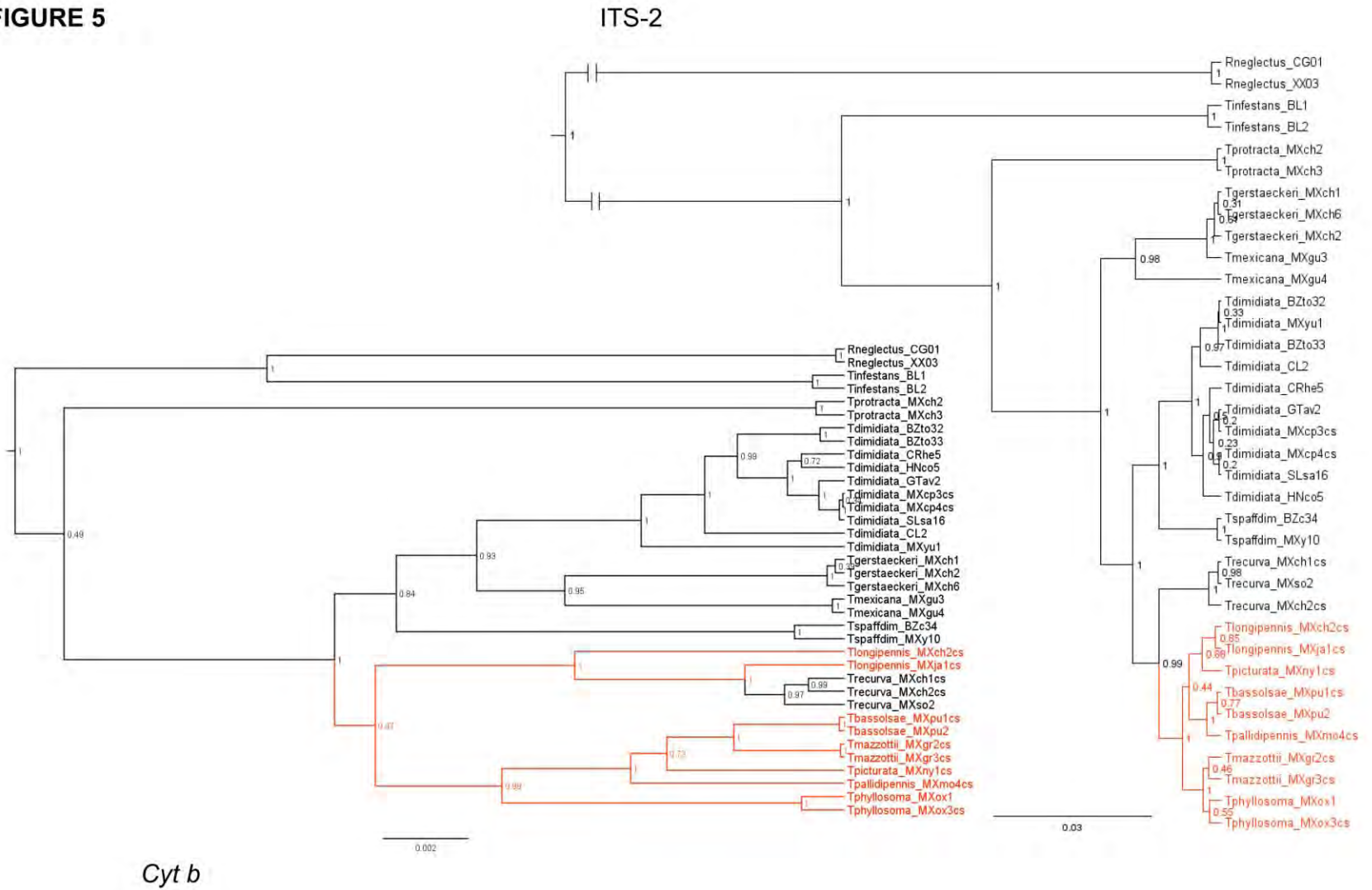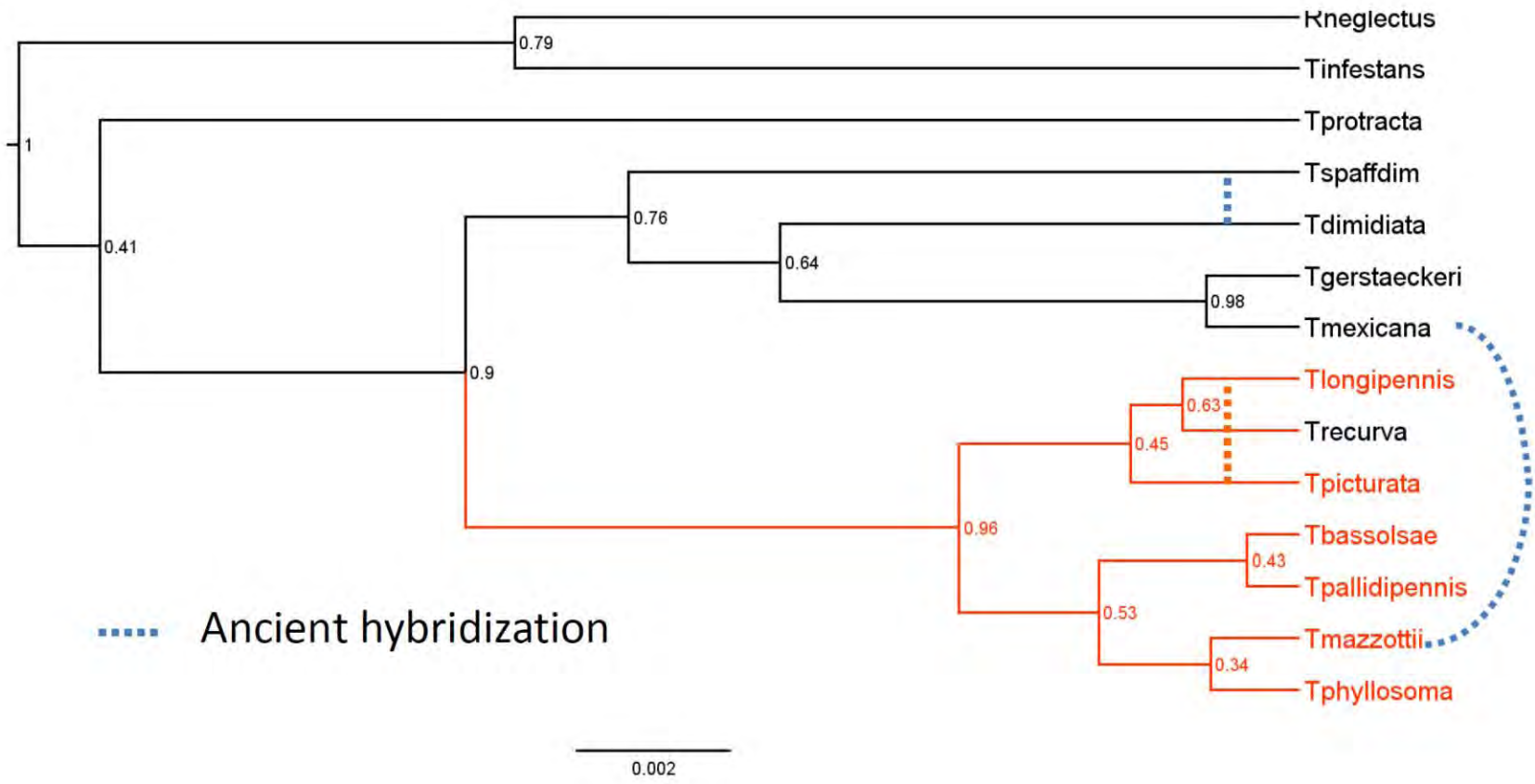| Group	Taxon	*Phyllosoma* complex	Sample code	Provenance	Location	ITS-2	*Cyt b*
Ingroup	*Triatoma bassolsae*	Carcavallo et al., 2000; de la Rúa et al., 2014	bassolsae.MXpu1cs	Colony	Mexico, Puebla	XXX	XXX
			bassolsae.Mxpu2cs	Colony	Mexico, Puebla	AY860402§	AY859410§
	Triatoma longipennis	Lent & Wygodzinsky, 1979; de la Rúa et al., 2014	longipennis.MXja1cs	Colony	Mexico, Jalisco	XXX	XXX
			longipennis.MXch2cs	Colony	Mexico, Chihuahua	XXX	XXX
	Triatoma mazzottii	Lent & Wygodzinsky, 1979; de la Rúa et al., 2014	mazzottii.MXgr38	Colony	Mexico, Guerrero	AY860392§	AY859421§
			mazzottii.MXgr2cs	Colony	Mexico, Guerrero	XXX	XXX
			mazzottii.MXgr3cs	Colony	Mexico, Guerrero	XXX	XXX
	Triatoma pallidipennis	Lent & Wygodzinsky, 1979; de la Rúa et al., 2014	pallidipennis.MXmo4cs	Field	Mexico, Morelos	XXX	XXX
	Triatoma phyllosoma	Lent & Wygodzinsky, 1979; de la Rúa et al., 2014	phyllosoma.MXox1	Field	Mexico, Oaxaca	HQ185165	HQ185139
			phyllosoma.MXox3cs	Colony	Mexico, Oaxaca	XXX	XXX
	Triatoma picturata	Lent & Wygodzinsky, 1979; de la Rúa et al., 2014	picturata.MXny1cs	Colony	Mexico, Nayarit	XXX	XXX
			picturata.MXny6	Colony	Mexico, Nayarit	AY860396§	AY859413§
Outgroup	*Triatoma dimidiata*	Schofield, 1988; de la Rúa et al., 2014	dimidiata.MXcp3cs	Field	Mexico, Chiapas	XXX	XXX
			dimidiata.MXcp4cs	Field	Mexico, Chiapas	XXX	XXX
			dimidiata.SLsa16	Field	El Salvador, Santa Ana	AM286693€	JN585881€
			dimidiata.CL2	Field	Colombia	AM286704€	KT998309€
			dimidiata.CRhe5	Field	Costa Rica, Heredia	KT874432€	KT998330€
			dimidiata.GTav2	Field	Guatemala, Alta Verapaz	AM286694€	JN585861€
			dimidiata.MXyu1	Field	Mexico, Yucatán	FJ197146€	FJ197157€
			dimidiata.BZto32	Field	Belize, Toledo	DQ871354¶	FJ197155¶
			dimidiata.BZto33	Field	Belize, Toledo	FJ197153¶	FJ197154¶
			dimidiata.HNco5	Field	Honduras, Copán	KT874434€	KT998325€
	Triatoma sp. aff. *dimidiata*	Bargues et al., 2008; de la Rúa et al., 2014	Tspaffdim.MXy10	Field	Mexico, Yucatán	KT874439€	KT998296€
			Tspaffdim.BZc34	Field	Belize, Cayo	FJ197152€	FJ197156€
			Tspaffdim.BZt1	Field	Belize, Toledo	FJ197147€	KT998312€
	Triatoma gerstaeckeri	Bargues et al., 2008; de la Rúa et al., 2014	gerstaeckeri.MXch1	Field	Mexico, Chihuahua	JQ282707‡	JQ282723‡
			gerstaeckeri.MXch2	Field	Mexico, Chihuahua	JQ282708‡	JQ282724‡

		gerstaeckeri.MXch6	Field	Mexico, Chihuahua	JQ282709‡	JQ282709‡
Triatoma infestans	-	infestans.BL1	Field	Bolivia	HQ333235†	HQ333214†
		infestans.BL2	Field	Bolivia	HQ333218†	HQ333211†
Triatoma mexicana	Martínez et al., 2006; de la Rúa et al., 2014	mexicana.MXgu3	Field	Mexico, Guanajuato	JQ282711‡	JQ282726‡
		mexicana.MXgu4	Field	Mexico, Guanajuato	JQ282712‡	JQ282727‡
		mexicana.MXgu1cs	Colony	Mexico, Guanajuato	XXX	XXX
Triatoma protracta	-	protracta.MXch2	Field	Mexico, Chihuahua	JQ282714‡	JQ282728‡
	-	protracta.MXch3	Field	Mexico, Chihuahua	JQ282715‡	JQ282729‡
Triatoma recurva	Pfeiler et al., 2006; de la Rúa et al., 2014	recurva.MXso2	Field	Mexico, Sonora	JQ282716‡	JQ282731‡
		recurva.MXch1cs	Colony	Mexico, Chihuahua	XXX	XXX
		recurva.MXch2cs	Colony	Mexico, Chihuahua	XXX	XXX
Rhodnius neglectus	-	Rneglectus.CG01	-	-	KT316941	KT317059
	-	Rneglectus.XX03	-	-	KT317029	KT317035

€ Dorn et al., 2016; ¶Dorn et al., 2009; ‡Espinoza et al., 2013; § Martínez et al., 2006; †Quisberth et al., 2011; ¥ Villalobos et al., 2012; XXX: this study.

TABLE 2

F1*F1	T. bassolsae	T. longipennis	T. mazzottii	T. phyllosoma	T. picturata	T. pallidipennis	T. recurva	T. sp. aff. dimidiata	T. dimidiata	T. mexicana
T. bassolsae	_	€	§	€	€	€	‡	-	-	¶
T. longipennis	O	_	§	€	§	¥	‡	-	-	¶
T. mazzottii	Δ	O	_	€	§	§	‡	-	-	¶
T. phyllosoma	Δ	O	O	_	€	€	‡	-	-	¶
T. picturata	O	O	Δ	O	_	¥	‡	-	-	¶
T. pallidipennis	O	O	Δ	Δ	O	_	‡	-	-	¶
T. recurva	Δ	Δ	Δ	O	Δ	O[1]	_	-	-	‡
T. sp. aff. dimidiata	?	?	?	?	?	?	?	_	†	-
T. dimidiata	?	?	?	?	?	?	?	O[2]	_	-
T. mexicana	Δ	Δ	Δ	Δ	Δ	Δ	Δ	?	?	_

[1] Backcross among *T. recurva* x *T. pallidipennis* F1 hybrids with *T. pallidipennis* specimens. [2] Hybrids of sympatric species from Northern Maya Lowlands. †Herrera-Aguilar et al., 2009; §Martínez-Ibarra et al., 2008; ¥(2009); ¶(2011a); ‡(2015).

DISCUSIÓN GENERAL

Esta tesis presenta un estudio de las consecuencias ecológicas y evolutivas de las interacciones entre las múltiples especies que participan en la transmisión de la enfermedad de Chagas. A diferencia de las interacciones entre pares de especies, el estudio de las interacciones entre múltiples especies bajo un determinado contexto espacio-temporal permite la detección de patrones relevantes para la comprensión de la transmisión de la enfermedad de Chagas. Desde una perspectiva ecosistémica, se infirió el patrón de transmisibilidad de *T. cruzi* en áreas no desérticas de México, gracias a la red de interacciones potenciales entre sus vectores (triatominos) y huéspedes silvestres (mamíferos). El descubrimiento de este patrón posibilitaría la generación de hipótesis sobre la ecología de la enfermedad de Chagas, que podrían refutarse en futuros estudios. Desde una perspectiva evolutiva, se infirió que la introgresión, como resultado de la hibridación antigua, es común entre los principales vectores de la enfermedad de Chagas en México. Dicha inferencia puede abrir paso al estudio de la relevancia del origen híbrido de los vectores para la transmisión eficaz de *T. cruzi*. A continuación, se mencionan las principales implicaciones de los patrones ecosistémicos y evolutivos detectados para las interacciones entre los huéspedes silvestres y vectores de la enfermedad de Chagas.

Inicialmente, este trabajo provee una aproximación a la transmisión del agente etiológico de la enfermedad de Chagas, *T. cruzi* (Rengifo-Correa et al., 2017; capítulo 1). *Trypanosoma cruzi* es un endoparásito que desarrolla todo su ciclo de vida en sus huéspedes y vectores, por lo que es obligatoria la interacción entre los mismos para la transmisión del parásito (Jansen y Roque, 2010; 2015). La transmisión puede ser mediada por el vector o independiente del vector (vía materna o vía alimentación con mamíferos infectados) (Jansen y Roque, 2010; 2015). Identificar cuál de estos dos mecanismos es más frecuente tiene implicaciones epidemiológicas inmediatas. Cuando predomina la transmisión independiente del vector, se espera que el parásito se mantenga en los huéspedes silvestres y la transmisión hacia humanos sea poco frecuente. Cuando predomina la transmisión mediada por el vector, se espera que la frecuencia de transmisión de *T. cruzi* hacia humanos dependa del grado de asociación

de las poblaciones de vectores a los ambientes antropogénicos. En este trabajo se encontró que la transmisión mediada por vector prevalece sobre la transmisión independiente del vector en áreas no desérticas de México. Más aún, las especies de vectores estudiadas se mantienen como focos de infestación domésticos, peridomésticos y naturales (Salazar-Schettino et al., 2010; Waleckx et al., 2015), con registros numerosos en zonas rurales y urbanas (Ramsey et al., 2015). Estas características de los vectores de *T. cruzi* implican una conexión latente entre los huéspedes silvestres y los huéspedes humanos. Asimismo, en este trabajo se infirió que la transmisión de *T. cruzi* puede ocurrir potencialmente entre nueve vectores (*Triatoma barberi*, *T. dimidiata* -*T. dimidiata* 2-, *T.* sp. *aff. dimidiata* -*T. dimidiata* 3-, *T. longipennis*, *T. mazzotii*, *T. mexicana*, *T. pallidipennis*, *T. phyllosoma* y *T. picturata*) y más de 100 especies de mamíferos registrados para áreas no desérticas de México. Las ecorregiones no desérticas en México se presentan en dos tercios del país (Olson et al., 2001). Por lo tanto, la transmisión de *T. cruzi* en zonas no desérticas de México dista de ser resultado de eventos aislados geográficamente y limitado a unos pocos huéspedes y vectores.

Previamente se ha presumido la importancia de la transmisión de *T. cruzi* vía vector, pero este trabajo constituye de los primeros intentos de inferencia desde un enfoque ecosistémico. Varios estudios han evidenciado la importancia de la transmisión de *T. cruzi* mediada por vector en algunos estados de México (Salazar-Schettino et al., 2010; Waleckx et al., 2015). Asimismo, la relevancia de la detección de patrones de transmisibilidad *T. cruzi* entre sus vectores y huéspedes silvestres en México ha sido resaltada previamente (Georgieva et al., 2017; Peterson et al., 2002; Ramsey et al., 2012; Ravinovich et al., 2011). Sin embargo, dada la naturaleza de los métodos utilizados en estos estudios, la transmisión solo puede ser percibida o inferida entre pares de especies. Como consecuencia, el alcance de dichos estudios es local y limitado a pocas especies, lo que a su vez lleva a una comprensión fragmentaria y segregada del sistema epidemiológico. Recientemente se ha inferido un panorama global de la transmisión de *T. cruzi* mediada por vector en México (Ibarra-Cerdeña et al., 2017; Rengifo-Correa et al., 2017). En este panorama se resalta que la conexión

(interacciones potenciales) entre los huéspedes silvestres y vectores arrojan señales sobre la magnitud de la transmisión de *T. cruzi*. Por ejemplo, dependiendo de la competencia de los vectores, la riqueza de sus interacciones con huéspedes potenciales (Rengifo-Correa et al., 2017) y el grado de asociación de los huéspedes silvestres a ambientes antropogénicos (Ibarra-Cerdeña et al., 2017) se espera un incremento en la transmisión de *T. cruzi*. Todas estas inferencias surgieron del ecosistema más probable de *T. cruzi* en México, es decir, la red compleja conformada por las interacciones inferidas entre vectores y huéspedes de *T. cruzi*. En síntesis, en este trabajo no solo se enumera las especies más probables involucradas en la transmisión de *T. cruzi*, sino que también se infieren sus interacciones, ayudando a la comprensión de la transmisión de la enfermedad de Chagas en el ecosistema más probable de su agente etiológico.

La inferencia del ecosistema más probable *T. cruzi* ayuda a la comprensión de la transmisión del parásito al facilitar la detección de patrones de interacción en un contexto espacial dado. A partir de dichos patrones se pueden generar diferentes hipótesis, por ejemplo, sobre el flujo (conectividad entre áreas geográficas) y la dinámica (pérdida o ganancia de conexiones) de la transmisión de *T. cruzi* en México (Ibarra-Cerdeña et al., 2017). Otras preguntas que pueden abordarse desde una perspectiva ecosistémica son: (1) ¿cómo afecta la estacionalidad climática de las diferentes áreas geográficas a la probabilidad de transmisión de *T. cruzi*? (Ramsey et al., 2012); (2) ¿cómo afecta la pérdida de huéspedes silvestres en ambientes altamente modificados por el hombre al grado de domiciliación de los vectores? (Waleckx et al., 2015); (3) ¿están las unidades discretas de tipificación de *T. cruzi* asociadas a especies de huéspedes o a áreas geográficas (Jansen et al., 2015)?; (4) ¿cuál es el efecto de la diversidad regional de huéspedes en la virulencia de *T. cruzi*? (Flores-Ferrer et al., 2017), entre muchas otras preguntas. Asimismo, el ecosistema inferido para *T. cruzi* en sí mismo constituye una hipótesis susceptible a ser refutada experimentalmente. El ecosistema más probable de *T. cruzi*, dadas las asociaciones espaciales de los mamíferos (huéspedes potenciales) y triatominos (vectores potenciales) evaluados, es falsable con los registros de huéspedes confirmados para el área de estudio (Ibarra-

Cerdeña et al., 2017; Rengifo-Correa et al., 2017). Alternativamente, una forma directa de refutar la red de interacciones potenciales entre mamíferos y triatominos es el uso combinado de métodos moleculares y marcaje con isótopos estables (Gómez-Días y Figuerola, 2010). Si además los esfuerzos de marcaje se enfocan en el parásito, se podría refutar de manera directa el ecosistema más probable de *T. cruzi* y las preguntas mencionadas anteriormente. En conclusión, es posible generar nueva información sobre la transmisión de *T. cruzi* en un contexto ecosistémico a partir de hipótesis experimentalmente refutables, emergentes de la red compleja de interacción inferida para vectores y huéspedes.

Dado que el método de redes complejas de interacción es de carácter inferencial, su aplicación puede extenderse a sistemas en los que no es posible obtener evidencia directa de las interacciones bióticas. Por ejemplo, es posible inferir interacciones bióticas entre especies que coexistieron en el pasado. Este trabajo además provee una inferencia de la hibridación antigua entre los principales vectores de la enfermedad de Chagas en México, el complejo *Phyllosoma* (capítulo 2). Así, fue posible detectar introgresión en las muestras de ADN mitocondrial estudiadas, probablemente como consecuencia de la hibridación antigua entre diferentes especies del complejo *Phyllosoma*. Esta inferencia se logró gracias a la integración del contexto espacio-temporal (método de redes complejas), con la información de la ecología reproductiva y filogenética de las especies en consideración.

Cabe resaltar que este estudio constituye el primer esfuerzo de involucrar explícitamente el contexto espacio-temporal y ecológico a la inferencia de hibridación en el ámbito filogenético. Bajo este ámbito el objetivo principal ha sido discernir entre la hibridación y la coalescencia profunda como causas de la incongruencia entre árboles de genes (Holder et al., 2001). Por ejemplo, se busca cuantificar la contribución relativa de la hibridación y la coalescencia profunda, suponiendo como hibridación todo lo que no puede ser explicado por la modelación de la coalescencia profunda (Meng y Kubatko, 2009). Como consecuencia, cabe esperarse una sobrestimación de los eventos de hibridación (Yu et al., 2014). A diferencia de esta aproximación, aquí la inferencia de la hibridación no dependió de la estimación de la coalescencia profunda.

Más bien, la inferencia de la hibridación dependió del patrón filogenético (incongruencia entre árboles de genes), el contexto ecológico y el contexto espacio-temporal de las especies consideradas. Así, mientras el contexto espacio-temporal permite desestimar la hibridación entre especies que no coocurren (ni coocurrieron en el pasado), el contexto ecológico permite desestimar la hibridación entre especies cuya coocurrencia espacial ha sido tan prolongada que derivó en aislamiento reproductivo. Entre las especies para las que la hibridación es de esperarse, de acuerdo con el contexto espacio-temporal y el ecológico, se puede detectar incongruencia entre el árbol de gen mitocondrial y el nuclear como señal de introgresión. Al reducir la dimensión de interacciones bióticas esperadas (casos de hibridación), el contexto espacio-temporal y el ecológico ayudan a una inferencia mesurada de hibridación en el sistema filogenético estudiado.

Gracias a esta aproximación ecológica – biogeográfica, se pudo detectar las huellas de hibridación antigua e inferir una hipótesis filogenética para las especies del complejo *Phyllosoma*. Así, se detectó introgresión potencial, probablemente como consecuencia de la hibridación antigua, para un conjunto de muestras plausible de las especies del complejo y cercanas. Por ejemplo, *Triatoma mazzottii*, *T. mexicana*, *T. picturata* y *T.* sp. *aff. dimidiata* fueron inferidas como especies que sufrieron introgresión tras eventos de hibridación antigua con otras especies. Esta información sugiere que la introgresión puede ser un proceso común entre las diferentes especies del complejo *Phyllosoma* y cercanas. Adicionalmente, la inferencia filogenética del complejo fue beneficiada por la consideración de la introgresión y la coalescencia profunda. Estos dos procesos son fuentes esperadas de la incongruencia entre árboles de genes para especies de evolución rápida o reciente (Heled y Drummond, 2010), como es el caso del complejo *Phyllosoma* (Bargues et al., 2000; Ibarra-Cerdeña et al., 2014). Cuando muestras que carecen de introgresión fueron analizadas, considerando la coalescencia profunda, los valores de soporte de la filogenia inferida fueron mayores a los de la filogenia inferida para todas las muestras. Es decir, la comprensión de la historia evolutiva del complejo se vio beneficiada de la consideración integral de los procesos que causan incongruencia entre los árboles de genes.

Se espera que la detección de la hibridación en la historia evolutiva del complejo *Phyllosoma* sea una contribución importante para el estudio ecológico de los vectores asociados con la enfermedad de Chagas. Previamente se ha reportado que híbridos entre *T. dimidiata* y *T.* sp. *aff. dimidiata* presentan mayor susceptibilidad a *T. cruzi* que sus parentales (Herrera-Aguilar et al. 2009). Esta observación ha abierto la discusión sobre la posibilidad de transmisión diferencial de *T. cruzi* por parte de los vectores y sus híbridos, siendo estos últimos aparentemente más eficaces. Considerando que la hibridación parece un proceso recurrente en la historia evolutiva de algunos vectores de la enfermedad de Chagas, también queda abierto el interrogante sobre las repercusiones evolutivas de la asociación entre *T. cruzi* con los vectores y sus híbridos.

CONCLUSIONES

- Se infirieron algunas consecuencias ecológicas y evolutivas de las interacciones entre las especies involucradas en la transmisión de *T. cruzi* en México. Para dicha inferencia se desarrolló una aproximación hipotético – deductiva que involucró la biogeografía ecológica y la evolutiva, la información empírica sobre las interacciones entre pares de especies y la información filogenética.

- Los vectores y huéspedes más probables de *T. cruzi* y las interacciones inferidas para los mismos constituyen el ecosistema más probable del parásito. Este ecosistema facilita la detección de patrones de transmisión de la enfermedad de Chagas, de los que a su vez pueden surgir hipótesis experimentalmente refutables.

- Se infirió la importancia de la transmisión de *T. cruzi* vía vector para áreas no desérticas de México desde un enfoque ecosistémico, concordando con observaciones previas. La transmisión de *T. cruzi* puede ocurrir potencialmente entre nueve vectores (*Triatoma barberi*, *T. dimidiata* -*T. dimidiata* 2-, *T.* sp. *aff. dimidiata* -*T. dimidiata* 3-, *T. longipennis*, *T. mazzotii*, *T. mexicana*, *T. pallidipennis*, *T. phyllosoma* y *T. picturata*) y más de 100 especies de mamíferos registrados para áreas no desérticas de México. Dada la extensión del área en consideración y el comportamiento de los vectores involucrados, la transmisión de *T. cruzi* dista de ser resultado de eventos aislados geográficamente y limitado a unos pocos huéspedes y vectores.

- El carácter inferencial del método de redes complejas de interacción puede extenderse a sistemas en los que no es posible obtener evidencia directa de las interacciones bióticas. Aquí se infirió hibridación entre especies que coexistieron en el pasado, gracias a la integración del contexto espacio-temporal (método de redes complejas), con la información de la ecología reproductiva y filogenética.

- La introgresión parece un proceso común entre las especies del complejo *Phyllosoma*. La introgresión se detectó en las muestras de ADN mitocondrial de *Triatoma mazzottii*, *T. mexicana*, *T. picturata* y *T.* sp. *aff. dimidiata*. La introgresión detectada probablemente fue consecuencia de hibridación antigua entre las especies del complejo *Phyllosoma*.

- La consideración del papel de la introgresión y la coalescencia profunda en la historia evolutiva del complejo *Phyllosoma* beneficia la inferencia de sus relaciones filogenéticas.

- El complejo *Phyllosoma sensu* Lent y Wygodzynsky es parafilético respecto a *T. recurva*, de acuerdo al análisis de coalescencia de los genes ITS-2 (nuclear) y Cyt b (mitocondrial). *Triatoma recurva* se incluye en el complejo *Phyllosoma* por su compatibilidad reproductiva, morfología y distribución aproximadamente alopátrida con las demás especies del complejo.

- Futuros estudios podrían esclarecer las repercusiones evolutivas de la asociación entre *T. cruzi* con los vectores y sus híbridos, considerando que la hibridación en la historia evolutiva de algunos vectores de la enfermedad de Chagas parece un proceso recurrente.

LITERATURA CITADA

Abad-Franch, F., 2016. A simple, biologically sound, and potentially useful working classification of Chagas disease vectors. Memórias do Instituto Oswaldo Cruz 111, 649–651.

Abad-Franch, F., Diotaiuti, L., Gurgel-Gonçalves, R., Gürtler, R.E., 2013. Certifying the interruption of Chagas disease transmission by native vectors: *cui bono*? Memórias do Instituto Oswaldo Cruz 108, 251–254.

Agosta, S.J., Janz, N., Brooks, D.R., 2010. How specialists can be generalists: resolving the "parasite paradox" and implications for emerging infectious disease. Zoologia 27, 151–162.

Baack, E.J., Rieseberg, L.H., 2007. A genomic view of introgression and hybrid speciation. Current Opinion in Genetics & Development 17, 513–518.

Bargues, M.D., Marcilla, A., Ramsey, J.M., Dujardin, J.P., Schofield, C.J., Mas-Coma, S., 2000. Nuclear rDNA-based molecular clock of the evolution of Triatominae (Hemiptera: Reduviidae), vectors of Chagas disease. Memórias do Instituto Oswaldo Cruz 95, 567–573.

Baum, D.A., Smith, S.D. 2013. Tree thinking: an introduction to phylogenetic biology. Roberts and company publishers, Colorado.

Ceballos, G., Arroyo-Cabrales, J., 2012. Lista actualizada de los mamíferos de México 2012. Revista Mexicana de mastozoología Nueva época 2, 27–80.

Clayton, J., 2010. Chagas disease: pushing through the pipeline. Nature 465, S12–S15.

Coura, J.R., 2013. Chagas disease: control, elimination and eradication. Is it possible? Memórias do Instituto Oswaldo Cruz 108, 962–967.

De Fuentes-Vicente, J.A., Cabrera-Bravo, M., Enríquez-Vara, J.N., Bucio-Torres, M.I., Gutiérrez-Cabrera, A.E., Vidal-López, D.G., Martínez-Ibarra, J.A., Salazar-Schettino, P.M., Córdoba-Aguilar, A., 2017. Relationships between altitude, triatomine (*Triatoma dimidiata*) immune response and virulence of *Trypanosoma cruzi*, the causal agent of Chagas' disease. Medical and Veterinary Entomology 31, 63–71.

Flores-Ferrer, A., Marcou, O., Waleckx, E., Dumonteil, E., Gourbière, S., 2017. Evolutionary ecology of Chagas disease; what do we know and what do we need? Evolutionary Applications 1-18.

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.

Georgieva, A.Y., Gordon, E.R.L., Weirauch, C., 2017. Sylvatic host associations of Triatominae and implications for Chagas disease reservoirs: a review and new host records based on archival specimens. PeerJ 5, e3826.

Gómez-Díaz, E., Figuerola, J., 2010. New perspectives in tracing vector-borne interaction networks. Trends in Parasitology 26, 470–476.

Heled, J., Drummond, A.J., 2010. Bayesian inference of species trees from multilocus data. Molecular Biology and Evolution 27, 570–580.

Herrera-Aguilar, M., Be-Barragán, L.A., Ramirez-Sierra, M.J., Tripet, F., Dorn, P., Dumonteil, E., 2009. Identification of a large hybrid zone between sympatric sibling

species of *Triatoma dimidiata* in the Yucatan peninsula, Mexico, and its epidemiological importance. Infection, Genetics and Evolution 9, 1345–1351.

Holder, M.T., Anderson, J.A., & Holloway, A.K., 2001. Difficulties in detecting hybridization. Systematic Biology 50, 978–982.

Ibarra-Cerdeña, C.N., Valiente-Banuet, L., Sánchez-Cordero, V., Stephens, C.R., Ramsey, J.M., 2017. *Trypanosoma cruzi* reservoir—triatomine vector co-occurrence networks reveal meta-community effects by synanthropic mammals on geographic dispersal. PeerJ 5, e3152.

Ibarra-Cerdeña, C.N., Zaldívar-Riverón, A., Peterson, A.T., Sánchez-Cordero, V., Ramsey, J.M., 2014. Phylogeny and niche conservatism in North and Central American triatomine bugs (Hemiptera: Reduviidae: Triatominae), vectors of Chagas' disease. PLoS Neglected Tropical Diseases 8, e3266.

Jansen, A.M., Roque, A.L.R., 2010. Domestic and wild mammalian reservoirs, in: American trypanosomiasis. Elsevier, pp. 249–276.

Jansen, A.M., Xavier, S.C.C., Roque, A.L.R., 2015. The multiple and complex and changeable scenarios of the *Trypanosoma cruzi* transmission cycle in the sylvatic environment. Acta Tropica 151, 1–15.

Lent, H., Wygodzinsky, P., 1979. Revision of the Triatominae (Hemiptera, Reduviidae) and their significance as vectors of Chagas disease. Bulletin of the American Museum of Natural History 163, 123–520.

Meng, C., Kubatko, L.S., 2009. Detecting hybrid speciation in the presence of incomplete lineage sorting using gene tree incongruence: A model. Theoretical Population Biology 75, 35–45.

Morrone, J.J., 2009. Evolutionary biogeography: an integrative approach with case studies. Columbia University Press, New York.

Morrone, J.J., Escalante, T. 2016. Introducción a la biogeografía. Universidad Nacional Autónoma de México. México.

Nouvellet, P., Dumonteil, E., Gourbière, S., 2013. The Improbable Transmission of *Trypanosoma cruzi* to Human: the missing link in the dynamics and control of Chagas disease. PLoS Neglected Tropical Diseases 7, e2505.

Olson, D.M., Dinerstein, E., Wikramanayake, E.D., Burgess, N.D., Powell, G.V., Underwood, E.C., D'amico, J.A., Itoua, I., Strand, H.E., Morrison, J.C., Loucks, C.J., Allnutt, T.F., Ricketts, T.H., Kura, Y., Lamoreux, J.F., Wettengel, W.W., Hedao, P., Kassem, K.R., 2001. Terrestrial ecoregions of the world: a new map of life on earth. BioScience 51, 933–938.

Peterson, A.T., Sánchez-Cordero, V., Beard, C.B., Ramsey, J.M., 2002. Ecologic niche modeling and potential reservoirs for Chagas disease, Mexico. Emerging infectious diseases 8, 662.

Peterson, A.T., Soberón, J., 2012. Species distribution modeling and ecological niche modeling: getting the concepts right. Natureza & Conservação 10, 102–107.

Rabinovich, J.E., Kitron, U.D., Obed, Y., Yoshioka, M., Gottdenker, N., Chaves, L.F., 2011. Ecological patterns of blood-feeding by kissing-bugs (Hemiptera: Reduviidae: Triatominae). Memorias do Instituto Oswaldo Cruz 106, 479–494.

Ramírez-Pulido, J., González-Ruiz, N., Gardner, A.L., Arroyo-Cabrales, J. 2014. List of recent land mammals of Mexico, 2014. Special Publications, Museum of Texas Tech University, no. 63.

Ramsey, J.M., Gutiérrez-Cabrera, A.E., Salgado-Ramírez, L., Peterson, A.T., Sánchez-Cordero, V., Ibarra-Cerdeña, C.N., 2012. Ecological connectivity of *Trypanosoma cruzi* reservoirs and *Triatoma pallidipennis* hosts in an anthropogenic landscape with endemic Chagas disease. PLoS One 7, e46013.

Ramsey, J.M., Peterson, A.T., Carmona-Castro, O., Moo-Llanes, D.A., Nakazawa, Y., Butrick, M., Tun-Ku, E., la Cruz-Félix, K. de, Ibarra-Cerdeña, C.N., 2015. Atlas of Mexican Triatominae (Reduviidae: Hemiptera) and vector transmission of Chagas disease. Memórias do Instituto Oswaldo Cruz 110, 339–352.

Rengifo-Correa, L., Stephens, C.R., Morrone, J.J., Téllez-Rendón, J.L., González-Salazar, C., 2017. Understanding transmissibility patterns of Chagas disease through complex vector–host networks. Parasitology 144, 760–772.

Salazar-Schettino, M.P., 2011. A revision of thirteen species of Triatominae (Hemiptera: Reduviidae) vectors of Chagas disease in Mexico. Journal of the Selva Andina Research Society 1, 57–80.

Salazar-Schettino, P.M., Cabrera-Bravo, M., Vazquez-Antona, C., Zenteno, E., Alba-Alvarado, M.D., Gutierrez, E.T., Gomez, Y.G., Perera-Salazar, M.G., Torre, G.G. de la, Bucio-Torres, M.I., 2016. Chagas disease in Mexico: report of 14 cases of chagasic cardiomyopathy in children. The Tohoku Journal of Experimental Medicine 240, 243–249.

Stephens, C., Sánchez-Cordero, V., González Salazar, C., 2017. Bayesian inference of ecological interactions from spatial data. Entropy 19, 547.

Stephens, C.R., Heau, J.G., González, C., Ibarra-Cerdeña, C.N., Sánchez-Cordero, V., González-Salazar, C., 2009. Using biotic interaction networks for prediction in biodiversity and emerging diseases. PLoS ONE 4, e5725.

Tarleton, R.L., Reithinger, R., Urbina, J.A., Kitron, U., Gürtler, R.E., 2007. The challenges of Chagas disease— Grim Outlook or Glimmer of Hope? PLoS Medicine 4, e332.

Waleckx, E., Gourbière, S., Dumonteil, E., 2015. Intrusive versus domiciliated triatomines and the challenge of adapting vector control practices against Chagas disease. Memórias do Instituto Oswaldo Cruz 110, 324–338.

World Health Organization, 2002. Control of Chagas disease: second report of the WHO Expert Committee, Control of Chagas disease. WHO, Geneva.

World Health Organization, 2015. Chagas disease in Latin America: an epidemiological update based on 2010 estimates. WHO Weekly Epidemiological Record 90. WHO, Geneva.

Yu, Y., Dong, J., Liu, K.J., Nakhleh, L., 2014. Maximum likelihood inference of reticulate evolutionary histories. Proceedings of the National Academy of Sciences 111, 16448–16453.