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*

(*Received 27 July 2016; revised 19 October 2016; accepted 9 November 2016; first published online 12 January 2017*)

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

Parasitology (2017), **144**, 760–772.
doi:10.1017/S0031182016002468

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/

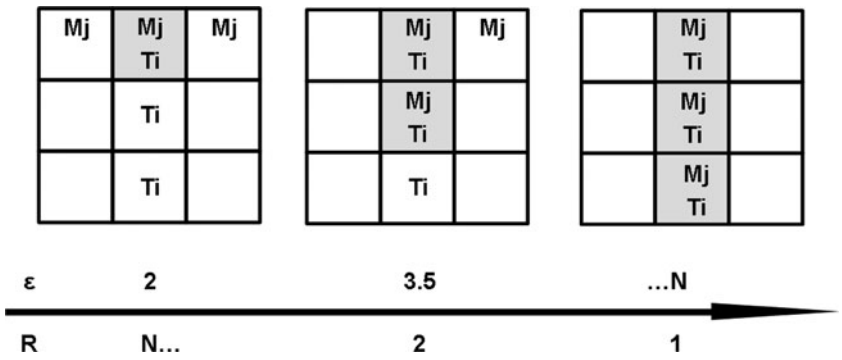

Fig. 1. Co-occurrence pattern of a triatomine (*Ti*) species and a mammal species (*Mj*). Epsilon (ε) values increase when the overlap between mammal and triatomine distributions increases. Relative rank (R) values are high for species with high ε values.

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:

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

where $P(T_i|M_j) = N_{Ti}$ and $_{Mj}/N_{Mj}$ with N_{Ti} and M_j 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

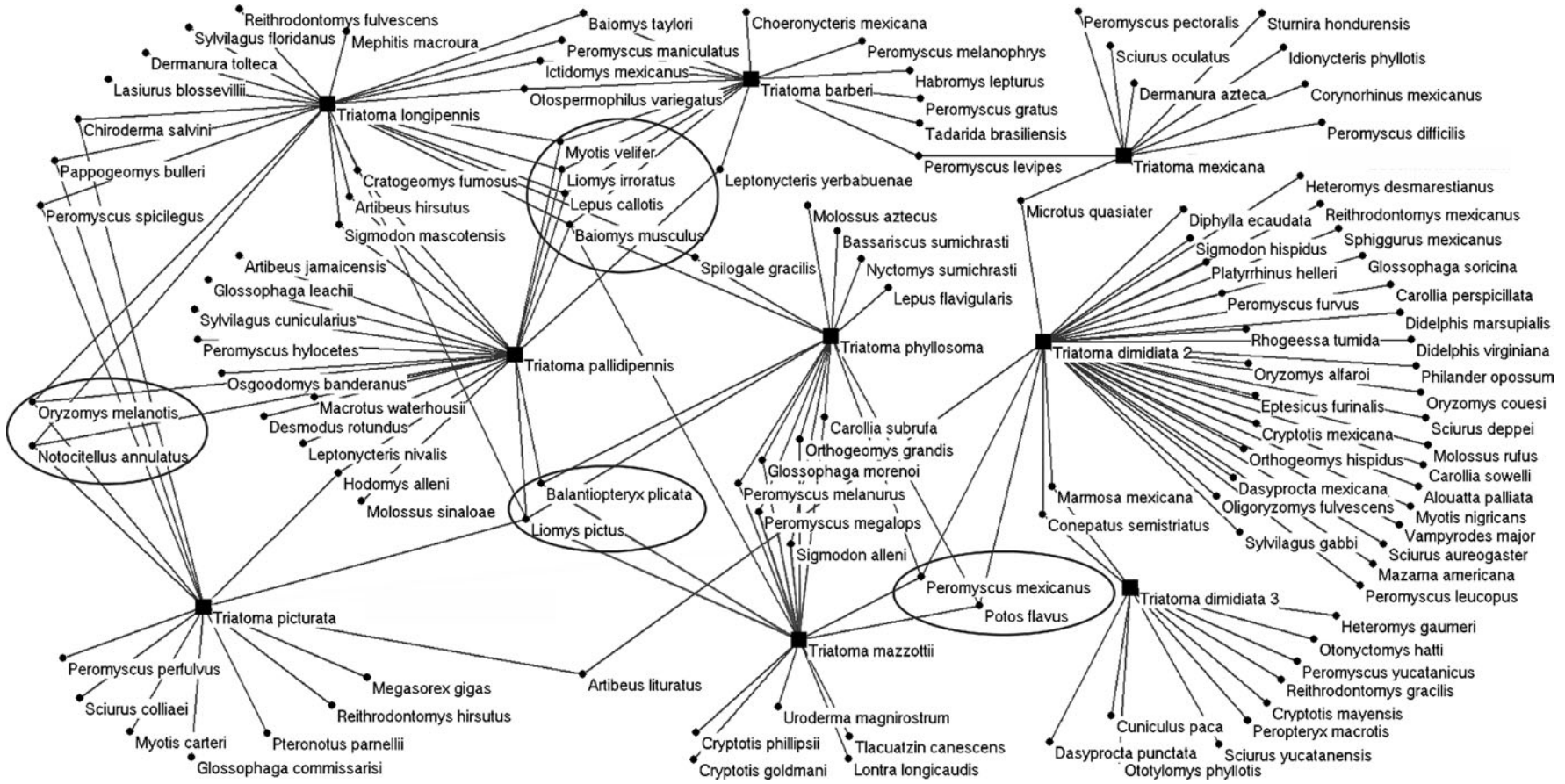

Fig. 2. Inferred vector–host network of *Trypanosoma cruzi* from Triatominae (squares) and wild mammals (points) species in non-desert areas of Mexico. Circled mammals potentially interact with three or more triatominae species.

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.

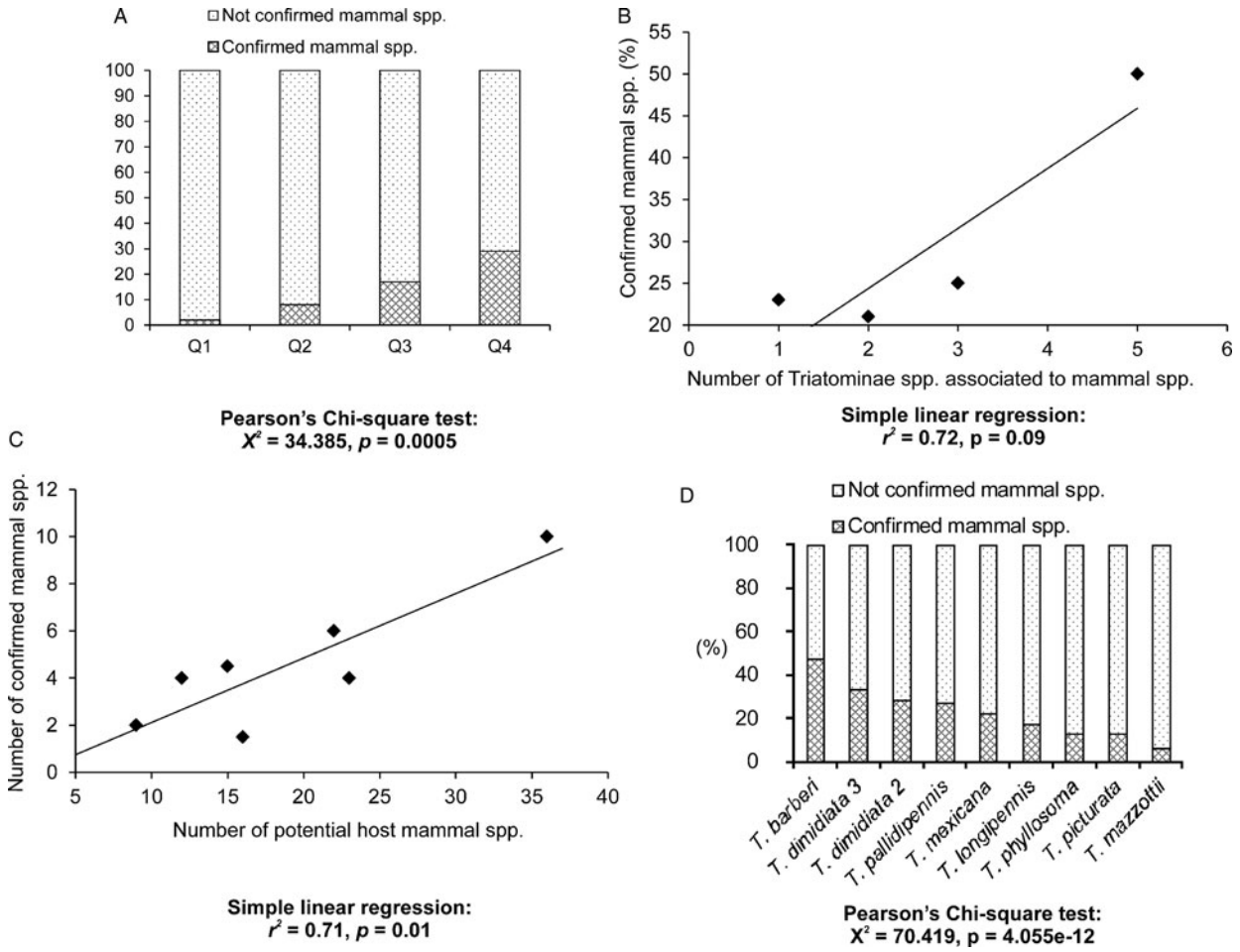

Fig. 3. Hypotheses of ecological epidemiology of Chagas disease in Mexico. Hypotheses were examined considering the wild mammal species confirmed as *Trypanosoma cruzi* hosts in Mexico. (A) H1. The probability for a mammal species to be confirmed for *T. cruzi* is different among different quartiles. (B) H2. The probability for a mammal species to be confirmed is not correlated 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 ($\varepsilon > 1{\cdot}96$, $\varepsilon > 4$ 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 $\varepsilon > 1{\cdot}96$ Q3 (17%) ($\chi^2 = 6{\cdot}7$, $P = 0{\cdot}03$), and with threshold $\varepsilon > 4$, the true positives percentage was higher in Q4 (33%), than with threshold $\varepsilon > 1{\cdot}96$ Q4 (29%) ($\chi^2 = 24{\cdot}2$, $P < 0{\cdot}0001$). 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 ($\varepsilon > 1{\cdot}96$). 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

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 ($\varepsilon > 1{\cdot}96$) 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* ($\chi^2 = 34{\cdot}385$, $P = 0{\cdot}0005$, 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 $P = -3{\cdot}648 + 0{\cdot}235 \times \varepsilon$, with a P-value $< 0{\cdot}001$ on the regression coefficient. This

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 ($r^2 = 0{\cdot}72$, $P = 0{\cdot}09$, 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 ($r^2 = 0{\cdot}72$, $P = 0{\cdot}01$, 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 ($\chi^2 = 70{\cdot}419$, $P = 4{\cdot}055 \times 10^{-12}$), 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

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

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.

ACKNOWLEDGEMENTS

We thank the anonymous referees for very helpful and constructive comments. We thank Paz María Salazar-Schettino and Gloria Rojas for providing valuable literature. LRC thanks the Posgrado en Ciencias Biológicas of the Universidad Nacional Autónoma de México (UNAM) for academic training and logistic support.

FINANCIAL SUPPORT

LRC thanks the Consejo Nacional de Ciencia y Tecnología for a doctoral scholarship. CRS acknowledges support from DGAPA-UNAM grant IN113414 and from Fronteras de la Ciencia grant 1093. CGS is grateful for financial support from the Centro de Ciencias de la Complejidad and the Laboratorio Nacional de Ciencias de la Complejidad.

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., Rodríguez-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., Martínez-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.

Martínez-Ibarra, J. A. and Novelo-López, M. (2004). Blood meals to molt, feeding time and postfeeding defecation delay of *Meccus pallidipennis* (Stål, 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 mitocondrial 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., Rodríguez-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.