CONABIO

Predictive modeling of zoonoses using spatial data

SPECIES Sistema Para la ExploraCión de Informacion ESpacial

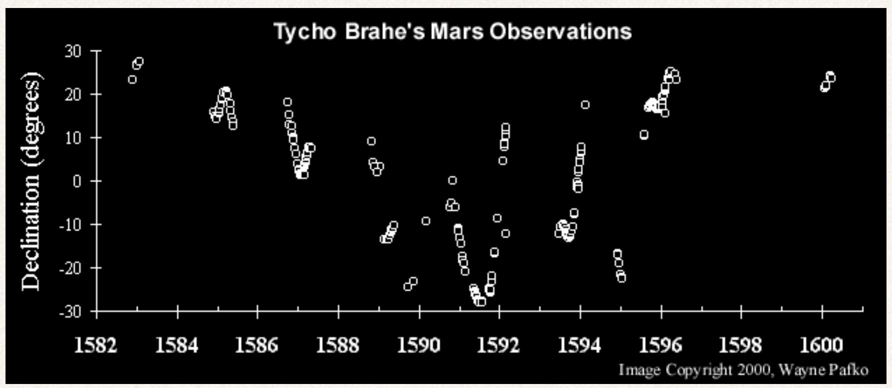
Chris Stephens
C3-Centro de Ciencias de la Complejidad y Instituto de Ciencias Nucleares, UNAM
Workshop on Arboviruses - INSP 09/02/2016


But… Just how many interactions can we directly observe?
Can we infer ecological interactions without direct observation?
Importancia médica
T. Infestans
T. barberi
T. pallidipennis
T. longipenis
T. recurva
T. neotomae
Culex quinquefasciatus
gettyimages
Frank Lukasseck
SCIENCEphotoLIBRARY

Inferring Interactions from Spatial Data…

A famous historical antecedent

Data —> Phenomenology —> Taxonomy —> Theory



Data → Phenomenology

Kepler's Laws

1. The orbit of a planet is an ellipse with the Sun at one of the two foci.
2. A line segment joining a planet and the Sun sweeps out equal areas during equal intervals of time.
3. The square of the orbital period of a planet is proportional to the cube of the semi-major axis of its orbit.

Isaac Newton computed the acceleration of a planet moving according to Kepler's first and second law.

1. The *direction* of the acceleration is towards the Sun.
2. The *magnitude* of the acceleration is inversely proportional to the square of the planet's distance from the Sun (the *inverse square law*).

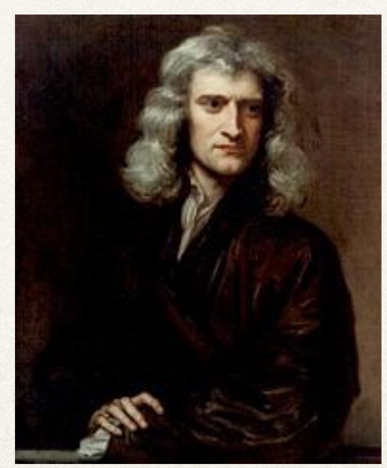
This implies that the Sun may be the physical cause of the acceleration of planets. Newton defined the force acting on a planet to be the product of its mass and the acceleration. So:

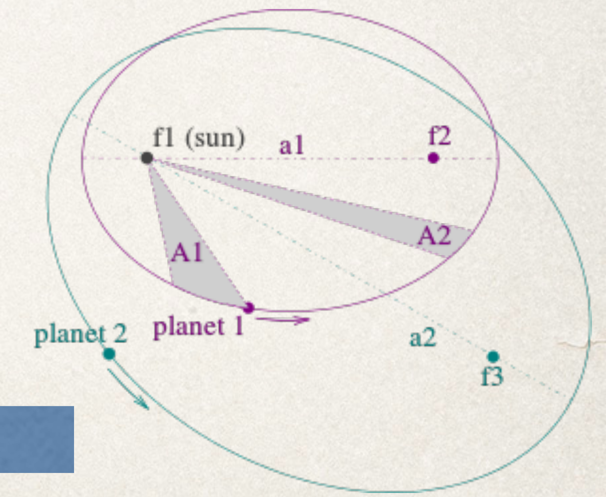
1. Every planet is attracted towards the Sun.
2. The force acting on a planet is in direct proportion to the mass of the planet and in inverse proportion to the square of its distance from the Sun.

The Sun plays an unsymmetrical part, which is unjustified. So he assumed, in Newton's law of universal gravitation:

1. All bodies in the solar system attract one another.
2. The force between two bodies is in direct proportion to the product of their masses and in inverse proportion to the square of the distance between them.

As the planets have small masses compared to the Sun, the orbits conform approximately to Kepler's laws. Newton's model fits actual observations more accurately.





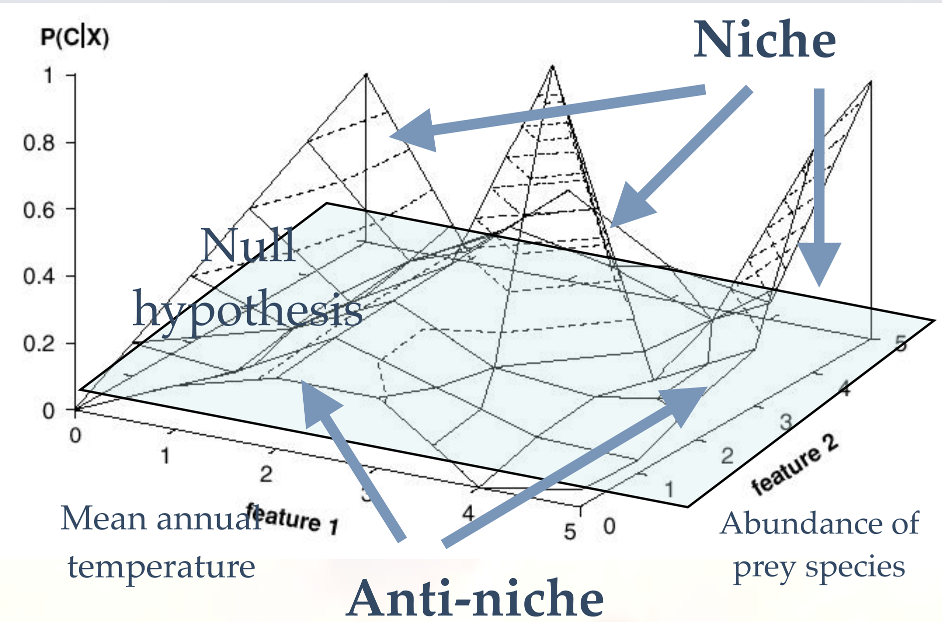
$$F = ma$$

$$F = GMm / r^2$$

Theory

"Keplerian" Ecological models

CONABIO



$P(C|\mathbf{X})$

What affects it?
The "niche"
$\mathbf{X} = (X1, X2, X3, \ldots, XM)$

What do we want to predict?
$C = (C1, C2, C3, \ldots, CN)$
the presence, or abundance, or,… of one or more populations or taxa

$S(C|\mathbf{X})$
Risk score

A large part of the complexity is in the multi-factoriality of both C and X. Adaptation is inherent in the fact that $P(C|X)$ can change in time.

$$\mathbf{X} = X(sd)+X(se)+X(n)+X(ev)+X(g)+X(af)+X(hm)+X(i)+X(sp)+\ldots$$

Macro-Climactic factors

Micro-Climatic factors

Hydrography

Prey species

Human activity

Behavioural characteristics

Phenotypic characteristics

Competitor species

Predator species

Problems of co-dependence and causality

And the data? Where are the "Brahes"? There's lots of them!

Normally data mining takes place in a "categorical" space (the equivalent in ecology is a niche space). However, most ecological data is spatio-temporal at multiple scales. Spatial data mining is much less developed than standard data mining.

- Collection data ← SNIB, CONABIO
- Ecological niche data
- Ecological niche model data
- Socio-economic data
- Socio-demographic data
- Phenotypic data
- Vegetable and crop cover
- Geographical data
- Medical and public health data…

Problems with spatial data:

Different sources

Different location, data base, access,…

Different data types

categorical, metric, continuous, discrete,…

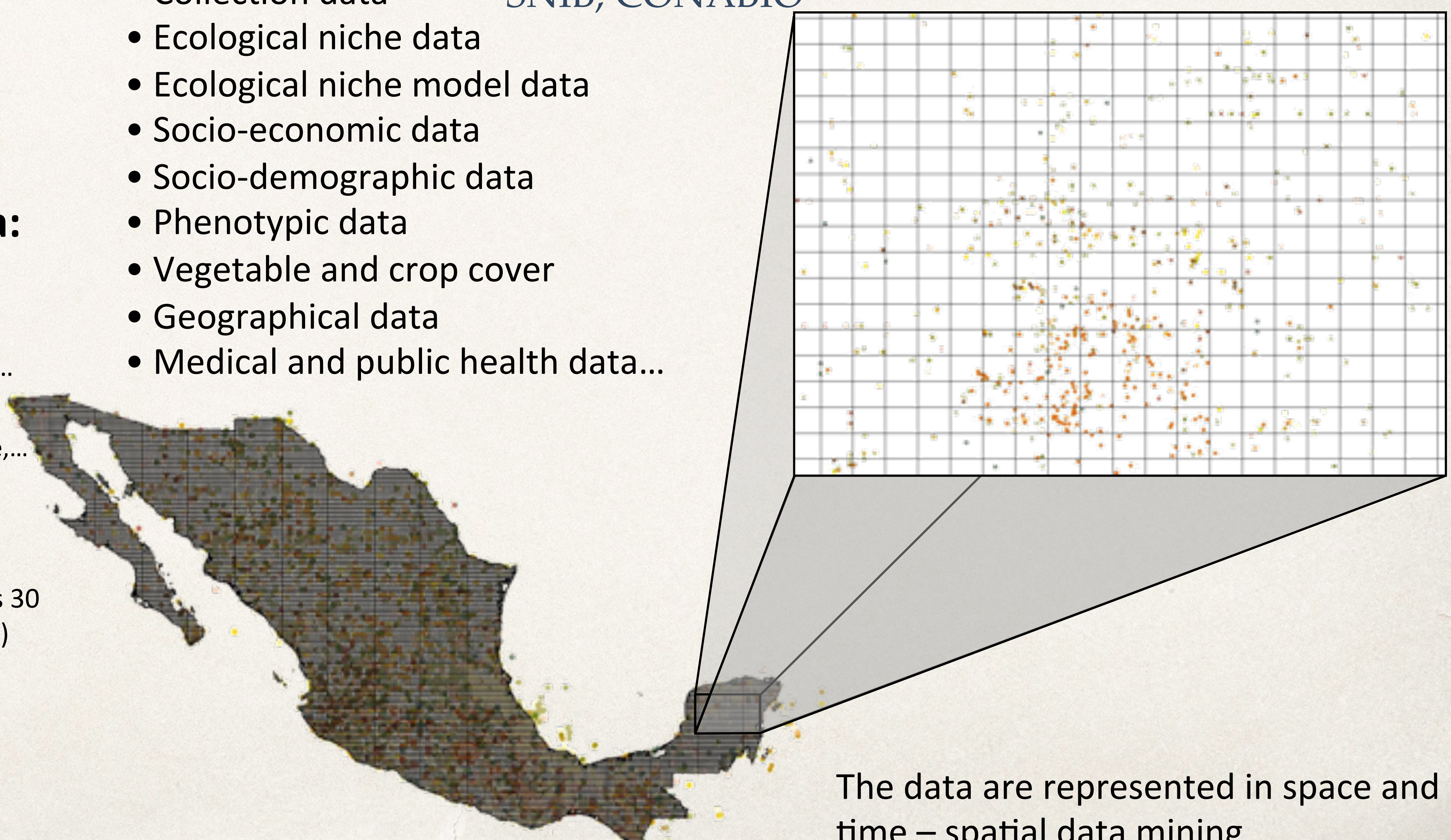
Different spatial resolution

Explicit – e.g., pixel by pixel in environmental layers

Implicit – 30,000,000 data points versus 30

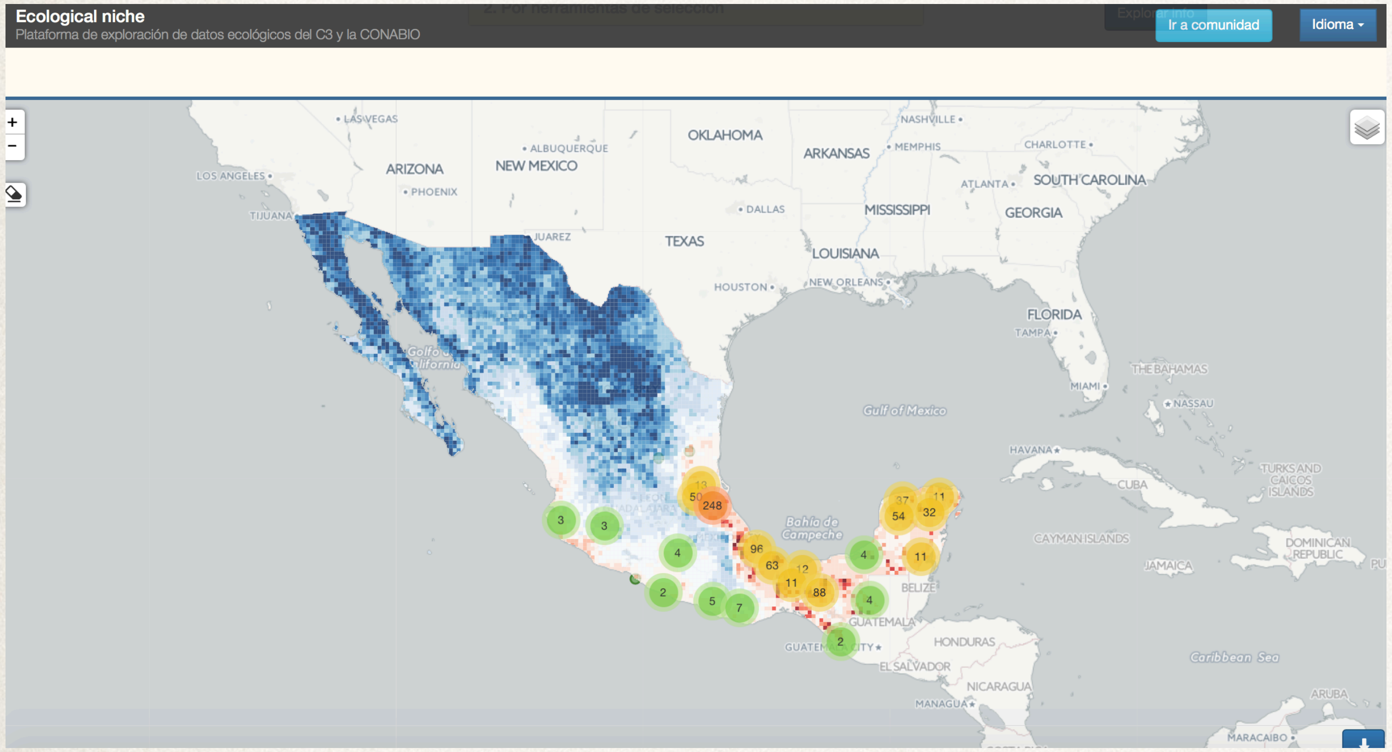
"Quality" (e.g. Phenotypic characteristic) versus "quantity"

Abiotic versus biotic



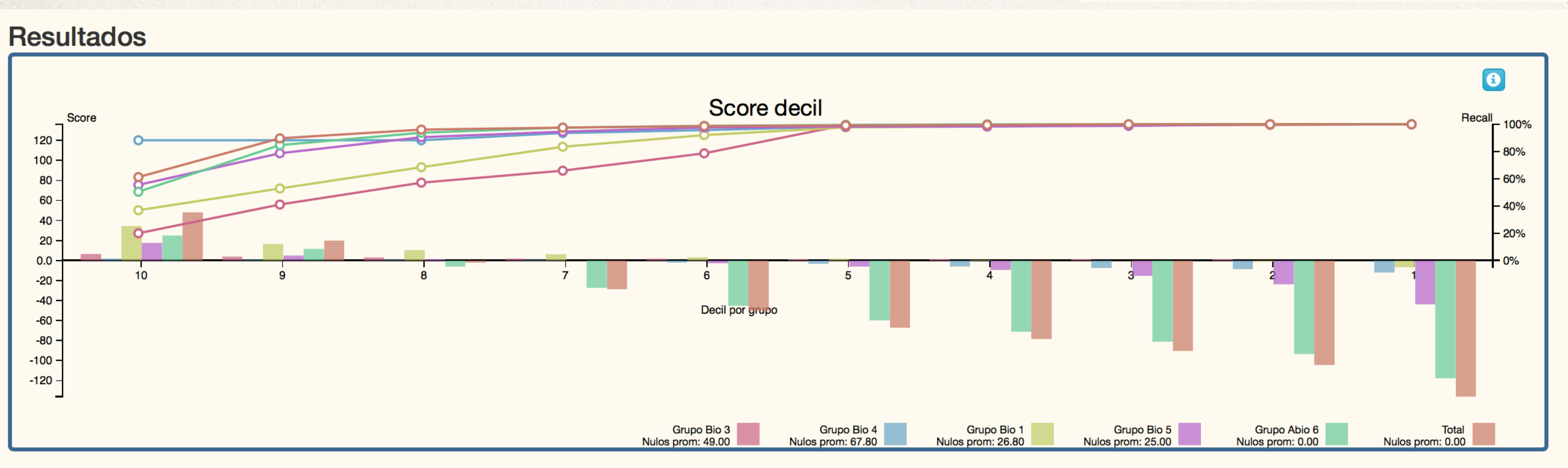
The data are represented in space and time – spatial data mining

Ecological Niche for vectors, hosts, pathogens, cases,… from SPECIES



Niche and anti-niche for Triatoma Dimidiata

Decil	Variable	Epsilon	Score	% Occ por decil
10	Carollia brevicauda	14.49	2.18	45.16%
10	Glossophaga soricina	14.20	1.62	32.43%
10	Artibeus jamaicensis	14.20	1.59	31.70%
10	Molossus rufus	13.89	2.20	48.55%
10	Myotis keaysi	13.41	2.51	61.54%
10	Eptesicus furinalis	13.27	2.75	51.85%
10	Diphylla ecaudata	12.98	2.69	60.71%
10	Mimon cozumelae	12.95	3.61	75.00%
10	Myotis nigricans	12.70	2.38	62.35%
10	Desmodus rotundus	12.59	1.52	31.22%
10	Dermanura tolteca	12.30	1.85	46.08%
10	Rhogeessa tumida	12.07	2.41	51.35%
10	Sturnira parvidens	11.59	1.55	33.82%

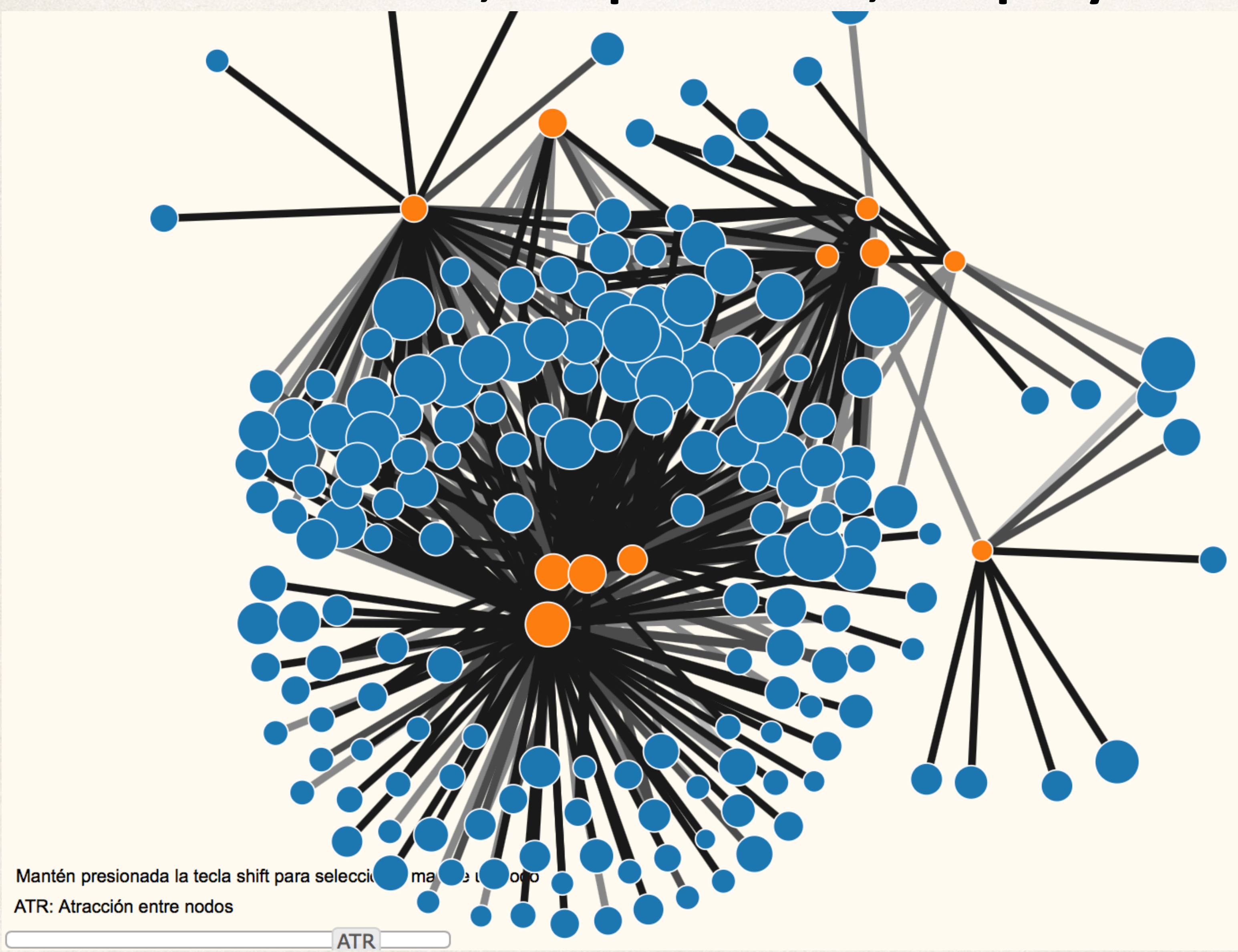


Both abiotic and biotic factors included

Now for Communities…

You can judge a man by his "friends"

or his "enemies", or "parasites", or "prey" or "predators" or…



Vector-host network for Chagas

CONABIO

From networks to predictive models: Leishmaniasis

	Mammals	Epsilon	Conf.
1	Eira barbara	10.1683	
2	Rhogeessa aeneus	9.3649	
3	Artibeus intermedius	9.1628	
4	Reithrodontomys gracilis	8.8921	Yes
5	Carollia sowelli	8.8303	
6	Heteromys gaumeri	8.8000	Yes
7	Peromyscus mexicanus	8.7859	
8	Heteromys desmarestianu	8.7164	Yes
9	Molossus rufus	8.6277	
10	Glossophaga soricina	8.5713	
11	Carollia perspicillata	8.5030	
12	Orthogeomys hispidus	8.3468	
13	Pteronotus parnellii	8.1632	
14	Desmodus rotundus	8.1519	
15	Dasyprocta mexicana	8.1128	
16	Sturnira lilium	8.0290	
17	Dermanura phaeotis	8.0055	
18	Dasyprocta punctata	7.9678	
19	Oryzomys couesi	7.7253	
20	Potos flavus	7.7246	
21	Conepatus semistriatus	7.6879	
22	Ototylomys phyllotis	7.5587	Yes
23	Ateles geoffroyi	7.4787	
24	Cryptotis magna	7.4207	
25	Cuniculus paca	7.3220	
26	Lampronycteris brachyotis	7.2852	
27	Sigmodon hispidus	7.2805	Yes
28	Peromyscus yucatanicus	7.2486	Yes
29	Oryzomys chapmani	7.1242	
30	Didelphis virginiana	7.1150	
31	Peromyscus melanocarpu	7.0260	
32	Microtus umbrosus	6.9630	
33	Thyroptera tricolor	6.9630	
34	Nasua narica	6.8953	
35	Megadontomys cryophilus	6.6830	
36	Oryzomys alfaroi	6.6816	
37	Sorex veraepacis	6.6797	
38	Carollia subrufa	6.6316	
39	Peromyscus aztecus	6.6173	
40	Didelphis marsupialis	6.4390	Yes
41	Sciurus yucatanensis	6.3865	
42	Philander opossum	6.2546	
43	Habromys ixtlani	6.1120	
44	Microtus waterhousii	6.1120	
45	Pteronotus rubiginosus	6.1120	
46	Reithrodontomys microdor	6.0967	
47	Coendou mexicanus	6.0268	
48	Centurio senex	6.0076	
49	Artibeus jamaicensis	5.9786	
50	Glossophaga morenoi	5.8847	

	Mammals	Epsilon	Conf.
51	Molossus sinaloae	5.8518	
52	Artibeus lituratus	5.8422	
53	Mormoops megalophylla	5.8374	
54	Habromys lepturus	5.7848	
55	Myotis keaysi	5.6148	
56	Chiroderma villosum	5.5562	
57	Tamandua mexicana	5.4845	
58	Tylomys nudicaudus	5.4510	
59	Saccopteryx bilineata	5.2984	
60	Macrotus mexicanus	5.2472	
61	Sciurus aureogaster	5.2267	
62	Baiomys musculus	5.2092	
63	Rhogeessa tumida	5.1950	
64	Sciurus deppei	5.1414	
65	Dermanura watsoni	5.1338	
66	Otonyctomys hatti	5.1338	
67	Orthogeomys grandis	5.0556	
68	Alouatta palliata	5.0457	
69	Choeroniscus godmani	5.0457	
70	Peropteryx macrotis	5.0457	
71	Pteronotus personatus	5.0266	
72	Lontra longicaudis	4.9330	
73	Reithrodontomys mexicanu	4.9120	
74	Oryzomys rostratus	4.8681	
75	Mimon cozumelae	4.8327	
76	Pteronotus davyi	4.7943	
77	Herpailurus yagouaroundi	4.7100	
78	Glossophaga leachii	4.6849	
79	Rhogeessa gracilis	4.6317	
80	Sylvilagus brasiliensis	4.6317	
81	Hodomys alleni	4.5155	
82	Leopardus wiedii	4.4420	
83	Peromyscus simulatus	4.4195	
84	Sigmodon alleni	4.3707	
85	Bassariscus sumichrasti	4.3110	
86	Oryzomys fulvescens	4.3110	
87	Diphylla ecaudata	4.3013	
88	Oryzomys melanotis	4.2907	Yes
89	Micronycteris microtis	4.2338	
90	Mazama americana	4.2274	
91	Microtus oaxacensis	4.2061	
92	Rheomys thomasi	4.2061	
93	Oryzomys saturatior	4.2061	
94	Myotis elegans	4.2024	
95	Oligoryzomys fulvescens	4.1984	
96	Natalus stramineus	4.0626	
97	Balantiopteryx io	4.0522	
98	Nyctinomops laticaudatus	4.0522	
99	Tlacuatzin canescens	4.0119	
100	Odocoileus virginianus	3.9265	

	Mammals	Epsilon	Conf.
101	Balantiopteryx plicata	3.8590	
102	Peromyscus leucopus	3.7994	
103	Sturnina ludovici	3.7888	
104	Enchisthenes hartii	3.6929	
105	Vampyrodes caraccioli	3.6929	
106	Eptesicus furinalis	3.6453	
107	Liomys pictus	3.6107	
108	Glossophaga commissaris	3.4861	
109	Lonchorhina aurita	3.4781	
110	Phyllostomus discolor	3.4781	
111	Peromyscus gymnotis	3.4516	
112	Anoura geoffroyi	3.4201	
113	Platyrrhinus helleri	3.3586	
114	Eumops bonariensis	3.3398	
115	Sciurus variegatoides	3.3398	
116	Uroderma bilobatum	3.3373	
117	Lasiurus intermedius	3.2197	
118	Lasiurus ega	3.1739	
119	Peromyscus megalops	3.1410	
120	Eumops glaucinus	3.0564	
121	Urocyon cinereoargenteus	2.9697	
122	Procyon lotor	2.9502	
123	Hylonycteris underwoodi	2.9343	
124	Rhynchonycteris naso	2.8580	
125	Eptesicus brasiliensis	2.8106	
126	Myotis albescens	2.8106	
127	Lophostoma evotis	2.8106	
128	Tapirus bairdii	2.8106	
129	Vampyrum spectrum	2.8106	
130	Marmosa mexicana	2.7731	Yes
131	Peromyscus furvus	2.7731	
132	Myotis velifera	2.5757	
133	Spilogale putorius	2.5411	
134	Microtus mexicanus	2.5268	
135	Dasypus novemcinctus	2.4725	
136	Myotis nigricans	2.4704	
137	Lophostoma brasiliense	2.4407	
138	Diclidurus albus	2.4407	
139	Sciurus niger	2.4407	
140	Leptonycteris curasoae	2.4268	
141	Nyctomys sumichrasti	2.4026	
142	Sigmodon mascotensis	2.3815	
143	Alouatta pigra	2.3374	
144	Peromyscus melanophrys	2.2204	
145	Dermanura tolteca	2.1920	
146	Trachops cirrhosus	2.1663	
147	Bauerus dubiaquercus	2.1612	
148	Spilogale pygmaea	2.1612	
149	Leptonycteris nivalis	2.1402	
150	Sylvilagus floridanus	2.1002	

- Only approximately 50 (2.5%) mammals from the Americas have been identified as hosts of Leishmania
- In México only 8 of 419 (2.1%) had been identified as hosts before our work

Model Validation

Species	ε	Negative	Positive	Total	% positive	Confidence (95%)
Carollia sowelli	8.83	43	2	45	4.4	-1 - 14
*Heteromys gaumeri**	8.8	5	0	5	0	-15 – 29
Peromyscus mexicanus	8.79	115	6	121	5	2 - 11
*Heteromys desmarestianus**	8.72	30	0	30	0	-2 - 16
Molossus rufus	8.63	1	0	1	0	-42 - 56
Glossophaga soricina	8.57	19	7	26	26.9	-3 - 16
Carollia perspicillata	8.5	8	0	8	0	-11 - 24
Pteronotus parnellii	8.16	4	0	4	0	-18 - 31
Desmodus rotundus	8.15	13	1	14	7.1	-6 - 20
Sturnira lilium	8.03	56	7	63	11.1	1 - 13
Artibeus phaeotis	8.01	35	1	36	2.8	-1 - 15
Oryzomys couesi	7.73	2	0	2	0	-28 - 41
*Ototylomys phyllotis**	7.56	9	1	10	10	-9 - 22
*Sigmodon hispidus**	7.28	36	4	40	10	-1 - 14
*Peromyscus yucatanicus**	7.25	3	0	3	0	-22 - 35
Didelphis virginiana	7.12	3	0	3	0	-22 - 30
Didelphis marsupialis	6.44	11	0	11	0	-8 - 21
Philander opossum	6.25	6	1	7	14.3	-12 - 25
Centurio senex	6.01	1	0	1	0	-42 - 56
Artibeus jamaicensis	5.98	81	5	86	5.8	1 - 12
Artibeus lituratus	5.84	38	3	41	7.3	-1 - 14
Myotis keaysi	5.61	2	0	2	0	-28 - 41
Chiroderma villosum	5.56	5	0	5	0	-15 - 29
Saccopteryx bilineata	5.3	1	0	1	0	-42 - 56
Sciurus aureogaster	5.23	71	8	79	7.3	1 - 12
Baiomys musculus	5.21	2	0	2	0	-28 - 41
Artibeus watsoni	5.13	2	0	2	0	-28 - 41
Choeroniscus godmani	5.05	10	3	13	23.1	-7 - 20
Pteronotus personatus	5.03	3	1	4	25	-18 - 31
Reithrodontomys mexicanus	4.91	1	0	1	0	-42 - 56
Oryzomys rostratus	4.87	22	1	23	4.3	-4 - 17
Micronycteris microtis	4.23	1	0	1	0	-42 - 56
Oligoryzomys fulvescens	4.2	6	0	6	0	-13 - 27
Peromyscus leucopus	3.8	22	4	26	15.4	-3 - 16
Sturnira ludovici	3.79	24	1	25	4	-3 - 17
Vampyrodes caraccioli	3.69	1	0	1	0	-42 - 56
Liomys pictus	3.61	47	1	48	2.1	0 - 14
Glossophaga commissarisi	3.49	2	6	8	75	-11 - 24
Lonchorhina aurita	3.48	1	0	1	0	-42 - 56
Phyllostomus discolor	3.48	0	1	1	100	-42 - 56
Platyrrhinus helleri	3.36	5	0	5	0	-22 - 35
Uroderma bilobatum	3.34	4	0	4	0	-18 - 31
Urocyon cinereoargenteus	2.97	1	0	1	0	-42 - 56
Procyon lotor	2.95	1	0	1	0	-42 - 56
Myotis velifer	2.58	3	0	3	0	-18 - 31
Microtus mexicanus	2.53	16	0	16	0	-6 - 19
Myotis nigricans	2.47	2	0	2	0	-28 - 41
Leptonycteris yerbabuenae	2.43	1	1	2	50	-28 - 41
Reithrodontomys fulvescens	2.08	20	0	20	0	-4 - 18
Neotoma mexicana	1.99	5	0	5	0	-15 - 29
Eptesicus fuscus	1.82	1	0	1	0	-42 - 56
Peromyscus levipes	1.34	1	0	1	0	-42 - 56
Sorex saussurei	1.29	3	0	3	0	-22 - 35
Osgoodomys banderanus	1.21	9	0	9	0	-10 - 23
Liomys irroratus	1.16	8	0	8	0	-11 - 24
Myotis auriculus	0.22	2	0	2	0	-28 - 41
Tadarida brasiliensis	-0.09	1	0	1	0	-42 - 56
Peromyscus hylocetes	-0.28	2	0	2	0	-28 - 41
Antrozous pallidus	-0.34	1	0	1	0	-42 - 56
Peromyscus zarhynchus	-0.46	2	0	2	0	-28 - 41
Chaetodipus hispidus	-0.71	4	0	4	0	-18 - 31
Peromyscus pectoralis	-0.73	2	0	2	0	-28 - 41
Neotomodon alstoni	-0.9	17	0	17	0	-5 - 19
Baiomys taylori	-1.16	10	3	13	23.1	-7 - 20
Chaetodipus nelsoni	-1.24	3	0	3	0	-22 - 35
Neotoma micropus	-1.27	16	0	16	0	-6 - 19
Peromyscus maniculatus	-1.37	58	2	60	3.3	0 - 13
Peromyscus eremicus	-1.41	0	1	1	100	-42 - 56
Perognathus flavus	-1.52	1	0	1	0	-42 - 56
Dipodomys merriami	-2.01	1	0	1	0	-42 - 56

- 922 individuals from 70 species were collected over a period of 18 months
- We predicted and confirmed 22 new species of host of Leishmania in México
- 13 of them are bats, identified for the first time in México
- Squirrels identified as hosts
- 34% of the species collected confirmed as hosts
- Average prevalence was 6.7%
- No species could be rejected as a host at the 95% confidence level
- Changes radically the panorama for control of Leishmania;
- Leishmania and Lutzomyias are generalists
- Linnean classification is not ecologically relevant

From networks to predictive models: Chagas

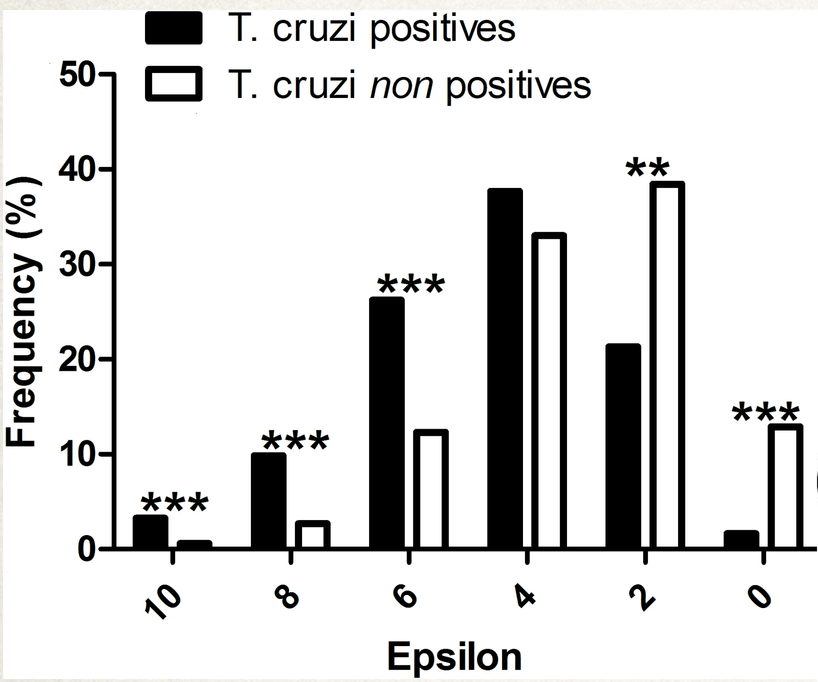
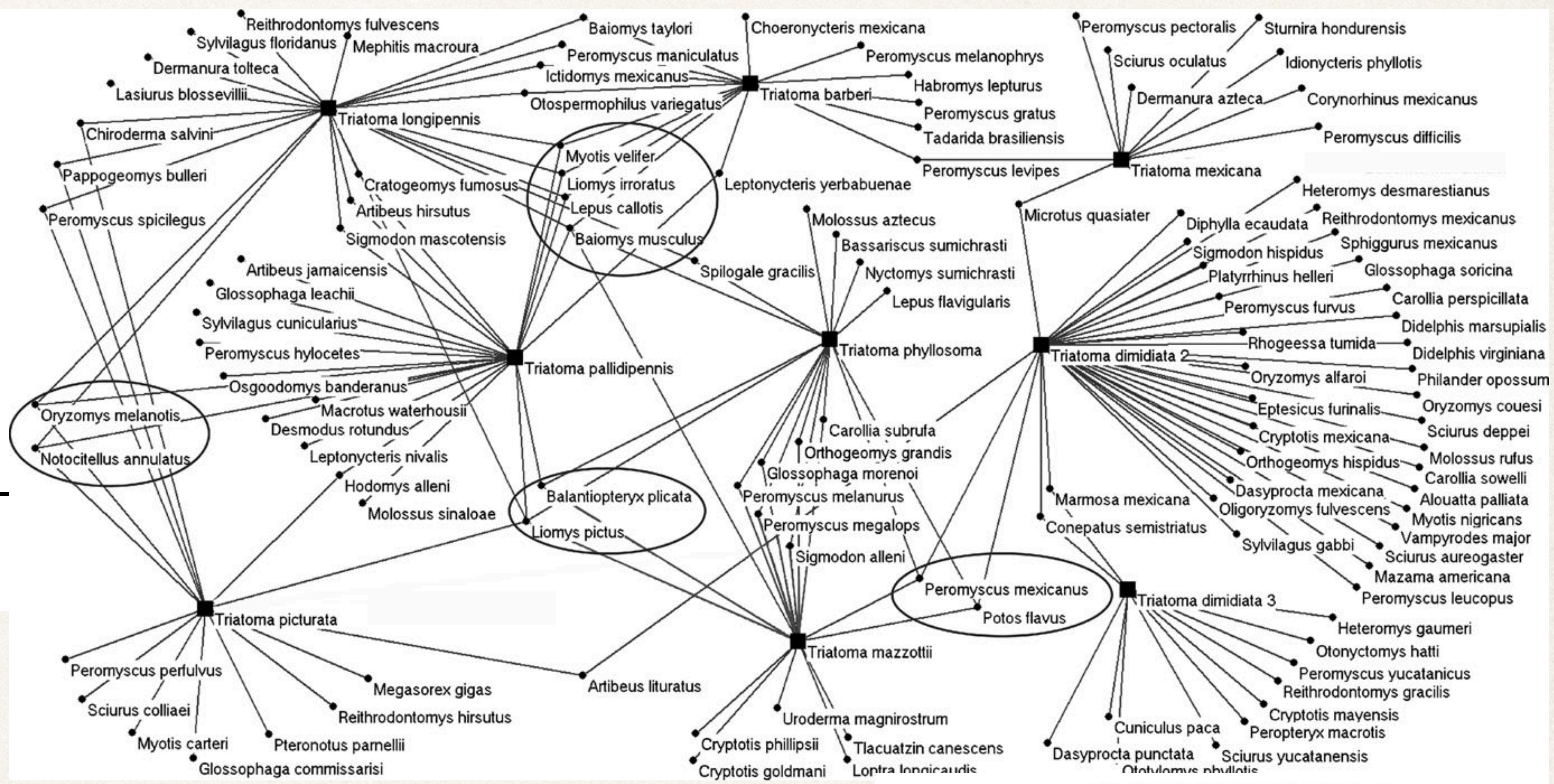
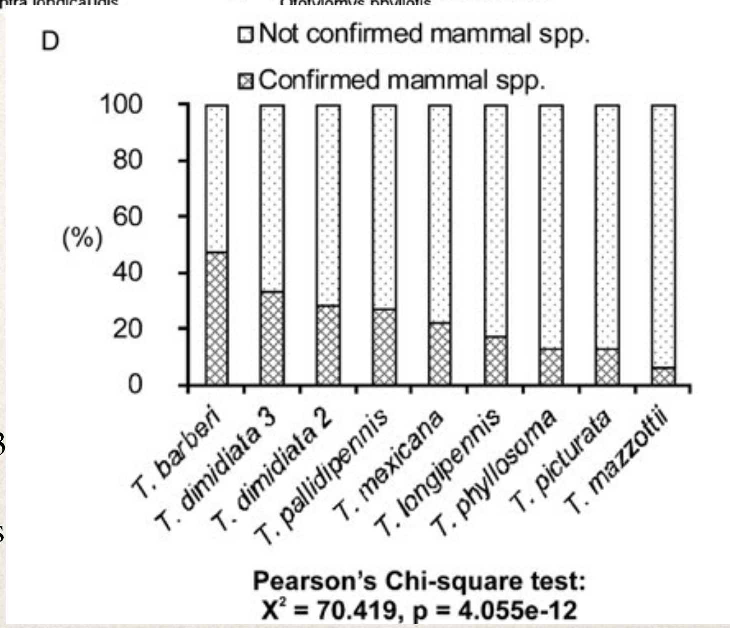


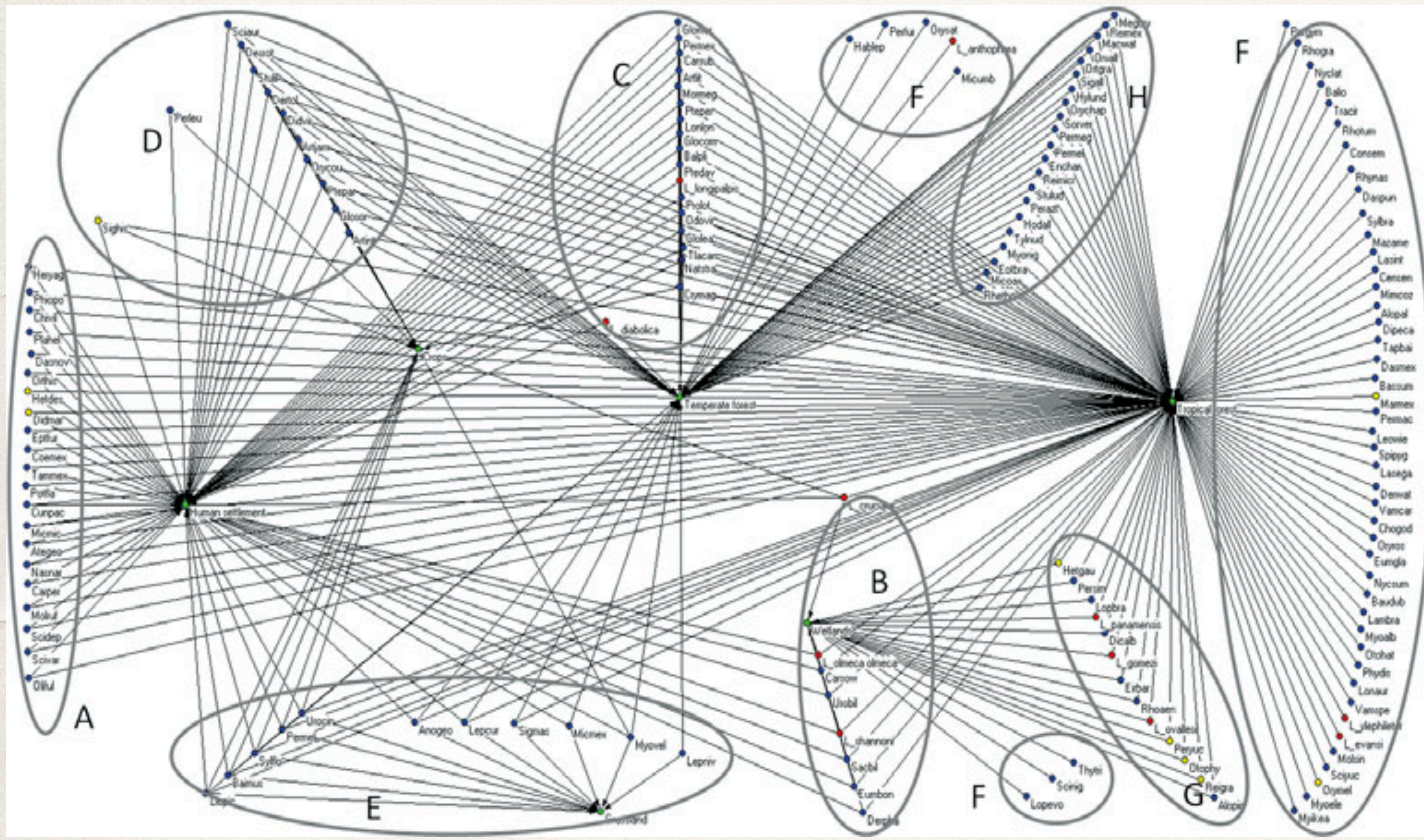
Figure 2 Evaluation of the performance of the interaction model: T. cruzi potential host species are those mammals that were independently reported in the literature as testing positive for natural infections by T. cruzi. mean P-values < 0.001, and are P-values < 0.01.



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).



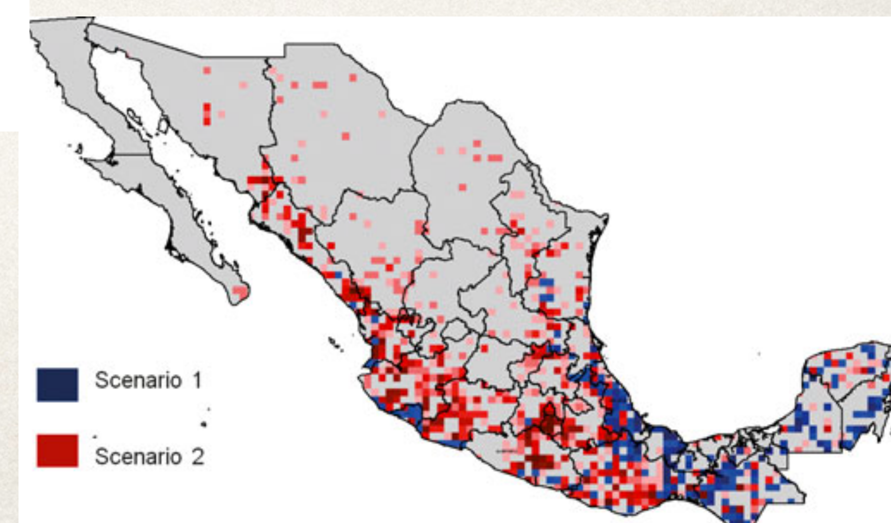
Dispersion of zoonoses



Assemblages of vectors and hosts of Leishmaniasis associated with different habitats

Can determine which potential hosts have overlap with habitats favoured by the vector without having direct overlap between vector and host

—> Disperal scenarios



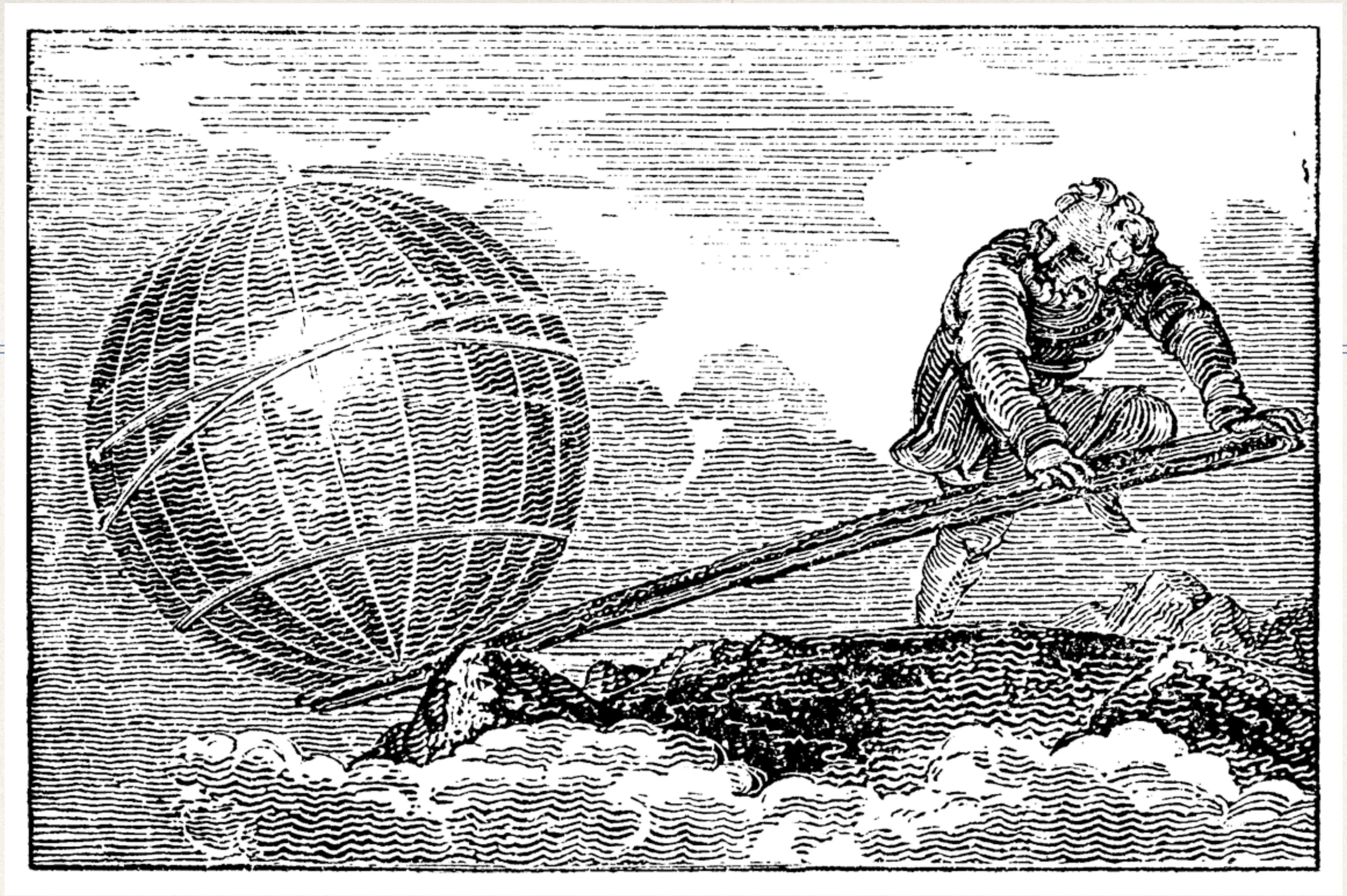
Conclusions

- Zoonoses are Complex Adaptive Systems
  - There are far too many relevant ecological interactions associated with a zoonosis to be experimentally identified and quantified
- The data associated with where and when things "are" (position and time), and what "things" (vectors, hosts, cases,…) are, can potentially tell us an enormous amount about "ecological" interactions
  - The methodology we have developed and the SPECIES platform allow us to infer/predict:
    - The full ecological niche (and anti-niche) of "things"
    - The network of the community/ecosystem in which they interact
- The labels for "things" - family, genus, species, DTU, population, sylvatic, peri-domestic, competent, guild, cases, biomarkers etc. - allow us to detect heterogeneity in the interactions
  - The more labels we have the more we can compare different hypotheses
  - More labels means more data: clinical cases, phylogenetic,…
  - Data has to be incorporated in the SPECIES platform

CONABIO

Conclusions

- Field and laboratory work is necessary to
  - Validate the predictive models
    - New hosts, new vectors,…
  - Validate data
    - Detect and correct data biases
- Leishmaniasis and Chagas are “generalists” - multi-vector, multi-host
  - Consistent with genetically very plastic pathogens
  - Genotypic variation has to leave phenotypic footprints
- There exists an “INFECTOME”
  - Many diseases are very “multi” - multi-pathogen (e.g., DTUs), multi-vector, multi-host etc.
    - Test for multiple pathogens
    - What pathogens are associated with which hosts?
    - How do pathogens interact?
- It is important to maintain development of the SPECIES platform
  - Develop further its functionality
  - Add more data layers: public health, socio-economic/socio-demographic,



δῶς μοι πᾶ στῶ καὶ τὰν γᾶν κινάσω

Give me a place to stand on and I´ll move the earth

Give me enough data and I´ll predict anything

The Data Revolution will revolutionise our ability to model and understand disease

Publications

Grupo de Trabajo

C3 - Centro de Ciencias de la Complejidad, UNAM; Instituto de Biología, UNAM; CONABIO; Facultad de Medicina, UNAM

1.- Dr. Christopher R. Stephens
2.- Dr. Raúl Sierra Alcocer
4.- Dr. Constantino González Salazar
5.- M. en C. Enrique del Callejo
6.- M. en C. Everardo Robredo
7.- Lic. Juan Carlos Salazar Carrillo

Competitive interactions between felid species may limit the southern distribution of bobcats Lynx rufus
V Sánchez-Cordero, D Stockwell, S Sarkar, H Liu, CR Stephens, ...
Ecography 31 (6), 757-764, 2008

Using biotic interaction networks for prediction in biodiversity and emerging diseases
CR Stephens, JG Heau, C González, CN Ibarra-Cerdeña, ...
PLoS One 4 (5), e5725, 2009

Exploratory analysis of the interrelations between co-located boolean spatial features using network graphs
R Sierra, CR Stephens
International Journal of Geographical Information Science 26 (3), 441-468, 2012

Constructing ecological networks: a tool to infer risk of transmission and dispersal of Leishmaniasis
C González-Salazar, CR Stephens
Zoonoses and public health 59 (s2), 179-193, 2012

Comparing the relative contributions of biotic and abiotic factors as mediators of species' distributions
C González-Salazar, CR Stephens, PA Marquet
Ecological Modelling 248, 57-70, 2013

Leishmania (L.) mexicana Infected Bats in Mexico: Novel Potential Reservoirs
M Berzunza-Cruz, Á Rodríguez-Moreno, G Gutiérrez-Granados, ...
PLoS neglected tropical diseases 9 (1), e0003438-e0003438, 2015

Predicting the potential role of non-human hosts in Zika virus maintenance
C González-Salazar, CR Stephens and V. Sanchez-Cordero
accepted for publication in Eco-health

UNDERSTANDING TRANSMISSIBILITY PATTERNS OF CHAGAS DISEASE THROUGH COMPLEX VECTOR-HOST NETWORKS
Laura Rengifo-Correa, Constantino González-Salazar, Juan J. Morrone, Juan Luis Téllez-Rendón, Christopher Stephens, Parasitology (2017) Jan 12:1-13. doi: 10.1017/S0031182016002468

Can you judge a disease host by the company it keeps? Predicting disease hosts and their relative importance using complex networks
CR Stephens et al, PLoS Neglected Tropical diseases, (2016)10(10): e0005004. doi:10.1371/journal.pntd.0005004