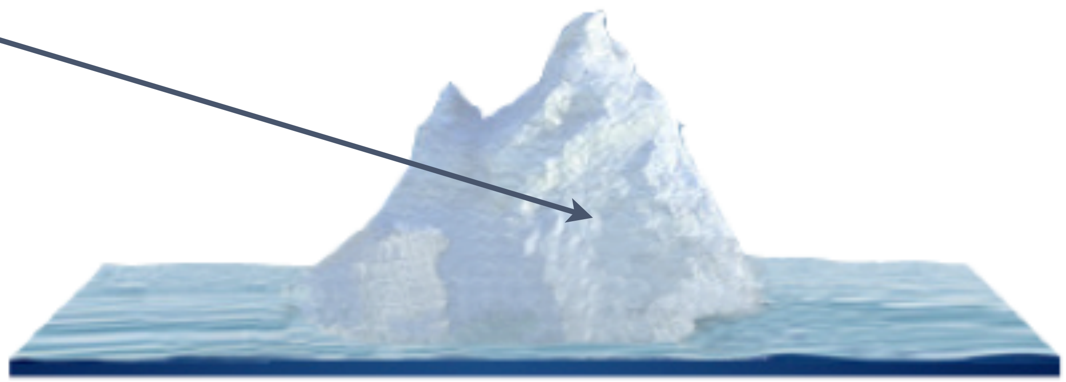
Can you Judge a Disease Host by its "Friends"?

Identifying biotic interactions through Co-occurrence Data

Chris Stephens
C3-Centro de Ciencias de la Complejidad y Instituto de Ciencias Nucleares, UNAM
Simposium "Ecología y evolución de enfermedades emergentes y reemergentes" VII CONGRESO DE LA SOCIEDAD CIENTÍFICA MEXICANA DE ECOLOGÍA, Querétero 02/10/19

What do we know…?
Known cases
Known vectors
Known hosts
Known risk factors



How do we infer what's under the water?

We need new methods to model these data…

Integrating multiple sources and multiple formats

Adaptive Systems

What do we know…?

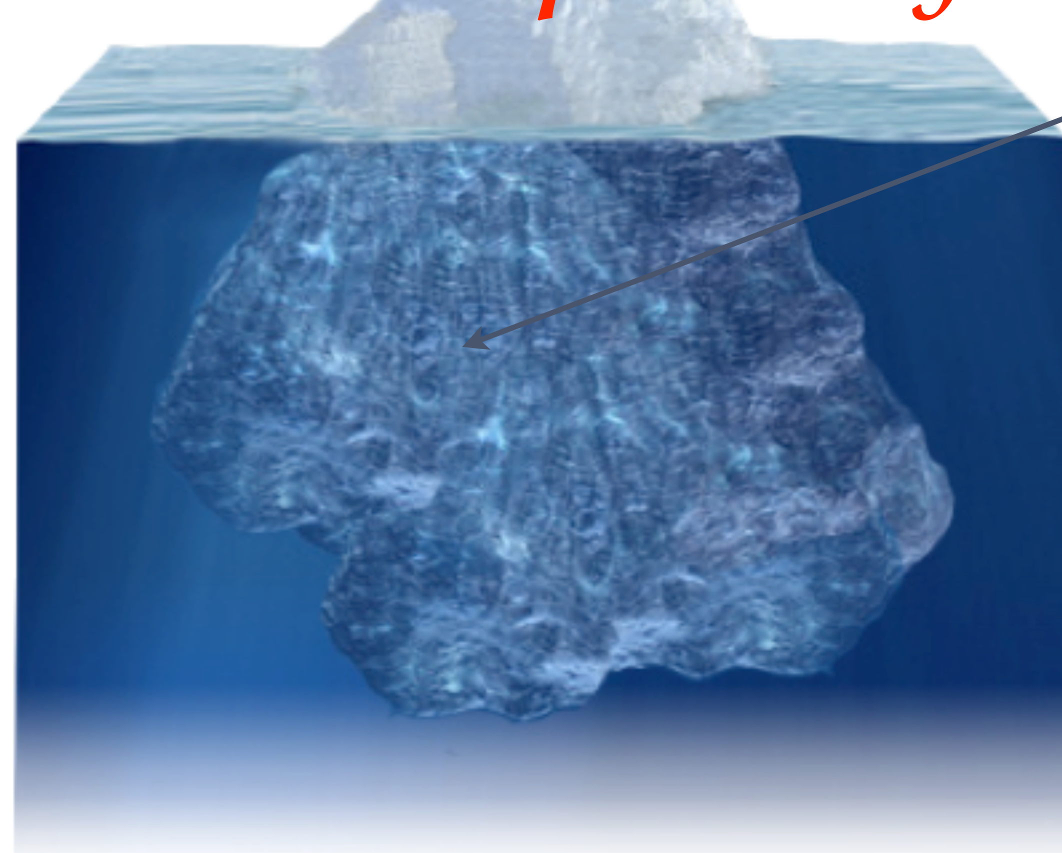
Known cases

Known vectors

Known hosts

Known risk factors

Known interactions



What don't we know…?

Unknown cases

Unknown vectors

Unknown hosts

Unknown risk factors

Unknown interactions

Can we infer ecological interactions without direct observation?

Just how many interactions can we directly observe?

What sort of data is necessary and sufficient?

Co-occurrence and Interaction

- The concept of "co-occurrence" in space and/or time is universal as an indicator/measure of interaction across all scientific areas. **It is a NECESSARY condition for interaction!**

- If two objects co-occur more or less than some non-interacting null hypothesis then we take it as a practical definition of "interaction".

$$\varepsilon(C|X_\alpha) = \frac{N_{X_\alpha}(P(C|X_\alpha) - P(C))}{\sqrt{(N_{X_\alpha}P(C)(1-P(C)))}}$$

- This tells us **nothing** though about the micro nature of the interaction

- For that we must consider the relevant **labels** (carnivore/herbivore/fructivore/insect/bat/…) for the participants in the interaction

- The concept of niche and community are vital for linking the micro to the macro. Niche gives the rationale for why an organism is where it is and why, micro interactions being the motivator for why a particular factor is a niche/anti-niche dimension

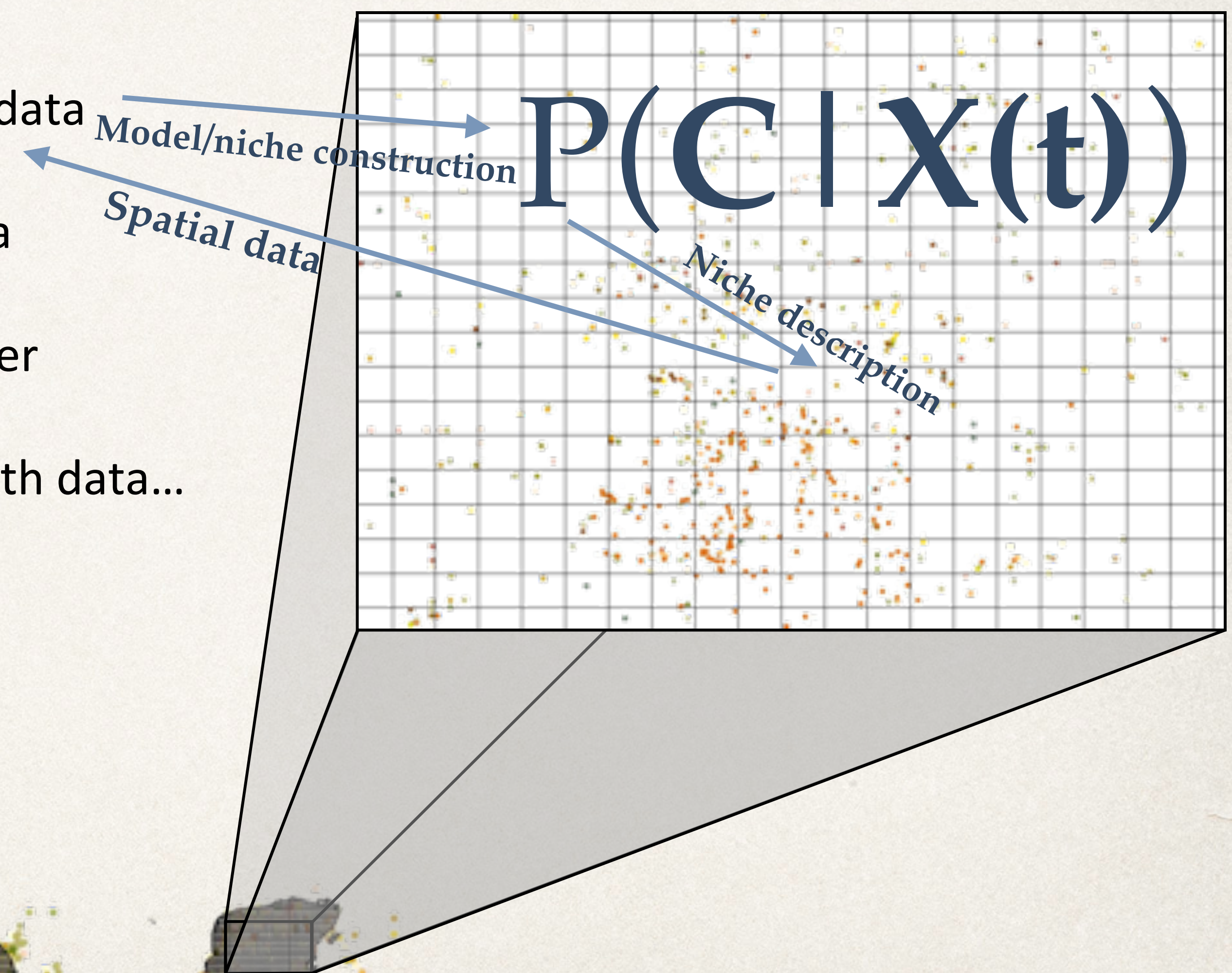
What is the niche of an EID?

The Data Revolution has provided a unique opportunity to construct multi-factorial ecological models for emerging diseases. However, most ecological data is spatio-temporal at multiple scales. Spatial data mining is much less developed than standard data mining.

$$P(C \mid \mathbf{X}(t))$$

Co-occurrence between EID occurrence, C, and niche variables, **X**(t) ecological interaction

- Collection data
- 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 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

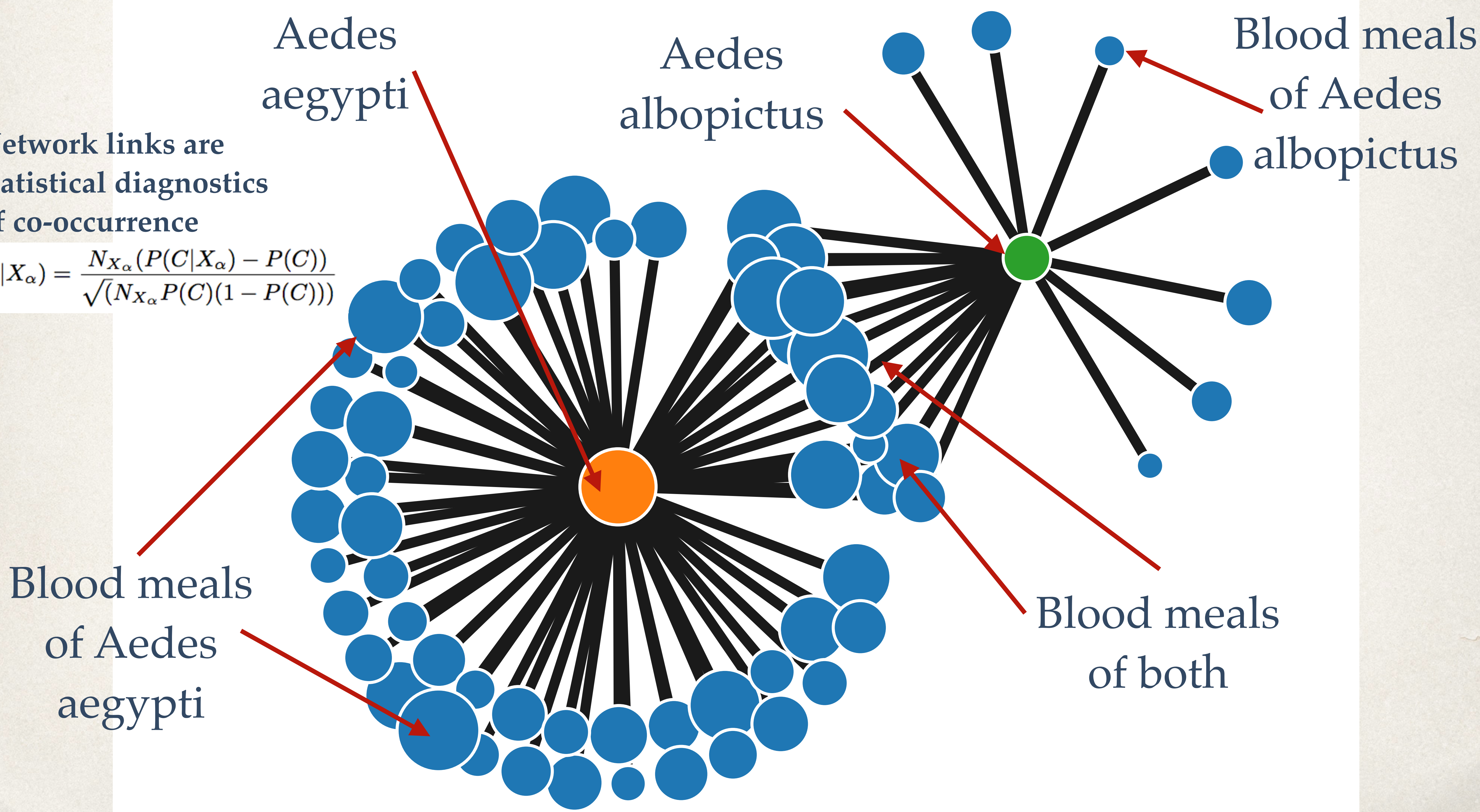
Abiotic versus biotic

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

What is a community of an EID?

Network links are statistical diagnostics of co-occurrence

$$\varepsilon(C|X_\alpha) = \frac{N_{X_\alpha}(P(C|X_\alpha) - P(C))}{\sqrt{(N_{X_\alpha}P(C)(1 - P(C)))}}$$

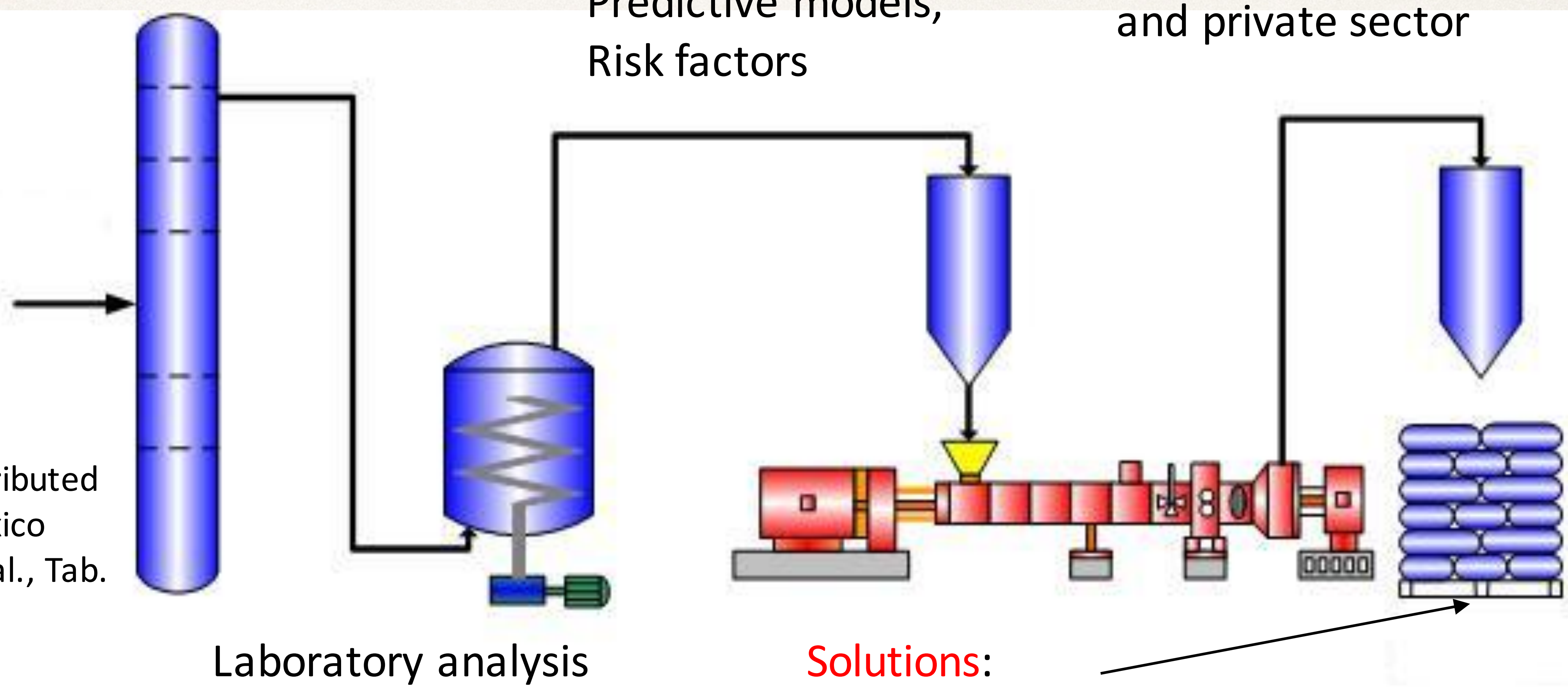


To Link Data-Predictions-Experiment

The Emerging Disease "production line"

Data Mining:
Predictive models,
Risk factors

Close relationship with
public health authorities
and private sector



Field work :
samples

Five groups distributed
throughout Mexico
DF, Chiap., NL, Jal., Tab.

Laboratory analysis

Two laboratories

Solutions:

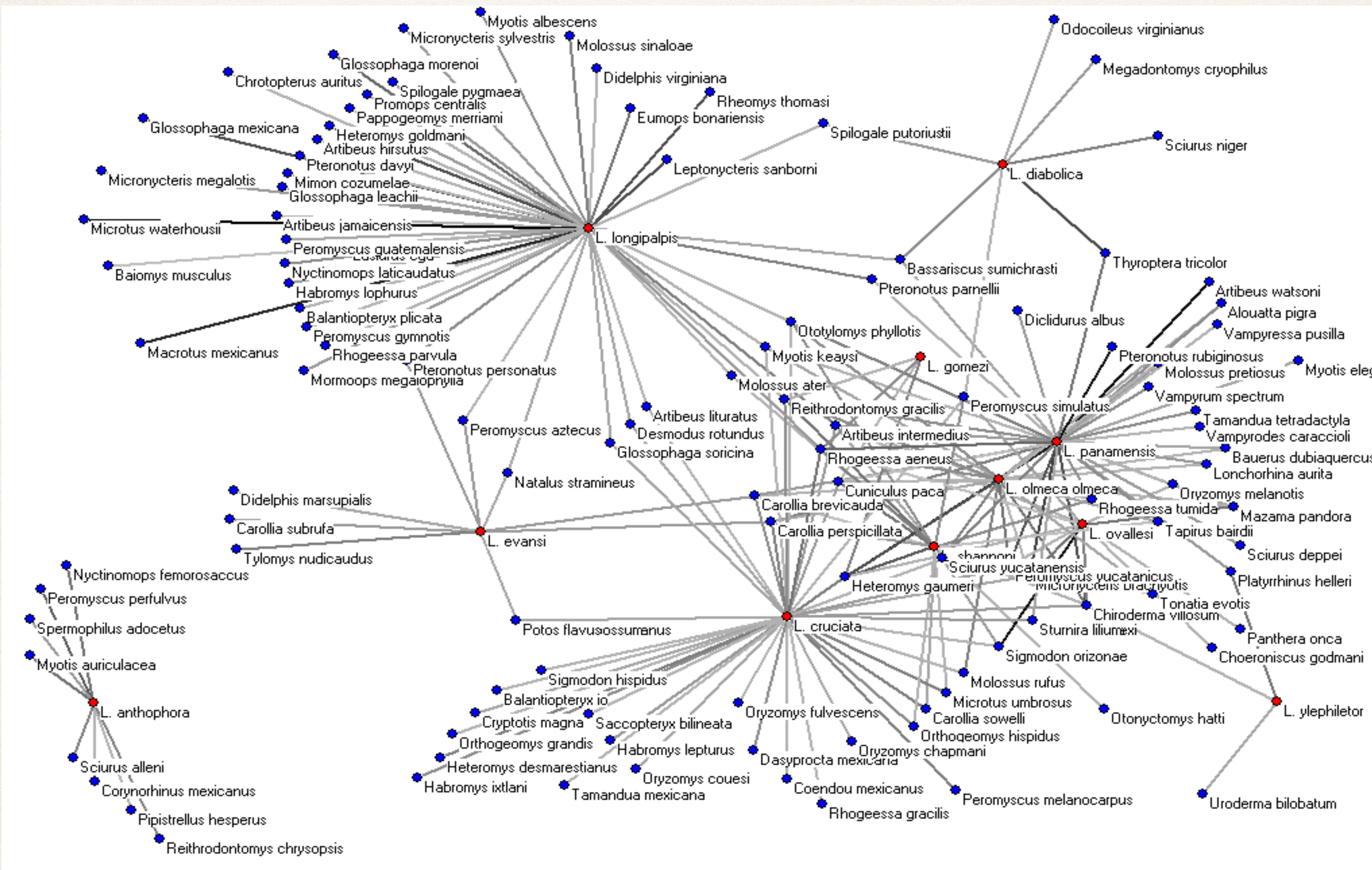
- Decision support systems
- Treatments
- Intelligent software

Over 2000 mammals
collected from over
110 species

We are always looking for collaboration: field work, lab work, modelling

Data-Predictions

Test zoonosis - Leishmaniasis



Predictions-Experiment
Test zoonosis - Leishmaniasis

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

Formatte

- Only about 50 (2.5%) of mammals on the American continent have been identified as hosts of Leishmania
- In Mexico only 8 out of 419 (2.1%) had been identified as hosts
- We collected 922 individuals from 70 species
- Predicted and confirmed 21 new species of mammal as carriers of Leishmania in Mexico
- 13 of them are bats, identified for the first time in Mexico
- Squirrels identified as carriers
- 33% of collected species were confirmed as hosts
- Overall infection rate was 6.7%
- No species could be rejected as a host at this infection rate at the 95% confidence level
- Changes the picture for control of Leishmania totally;
- Leishmania and Lutzomyias are eclectic in their host source.
- Linnean classification is NOT ecologically relevant

Data-Predictions-Experiment

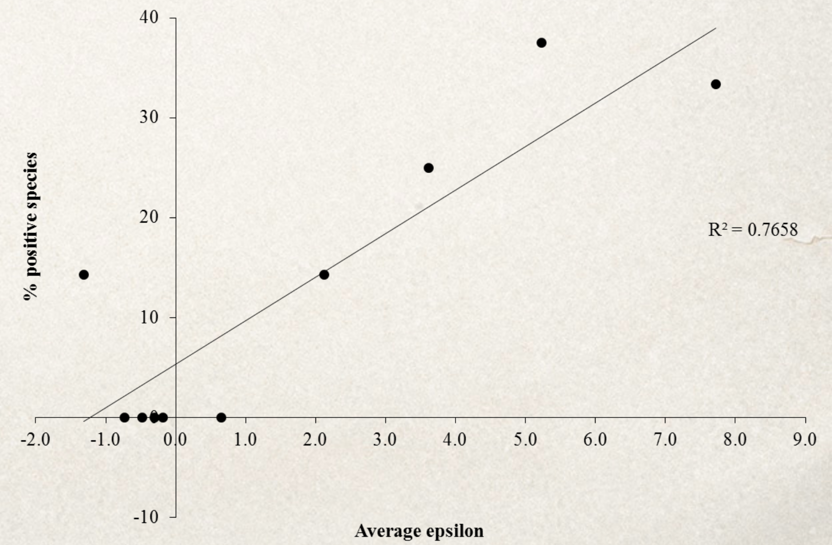
Test zoonosis - Leishmaniasis

	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	

Biotic facilitation seems to be the norm. Species are not distributed randomly



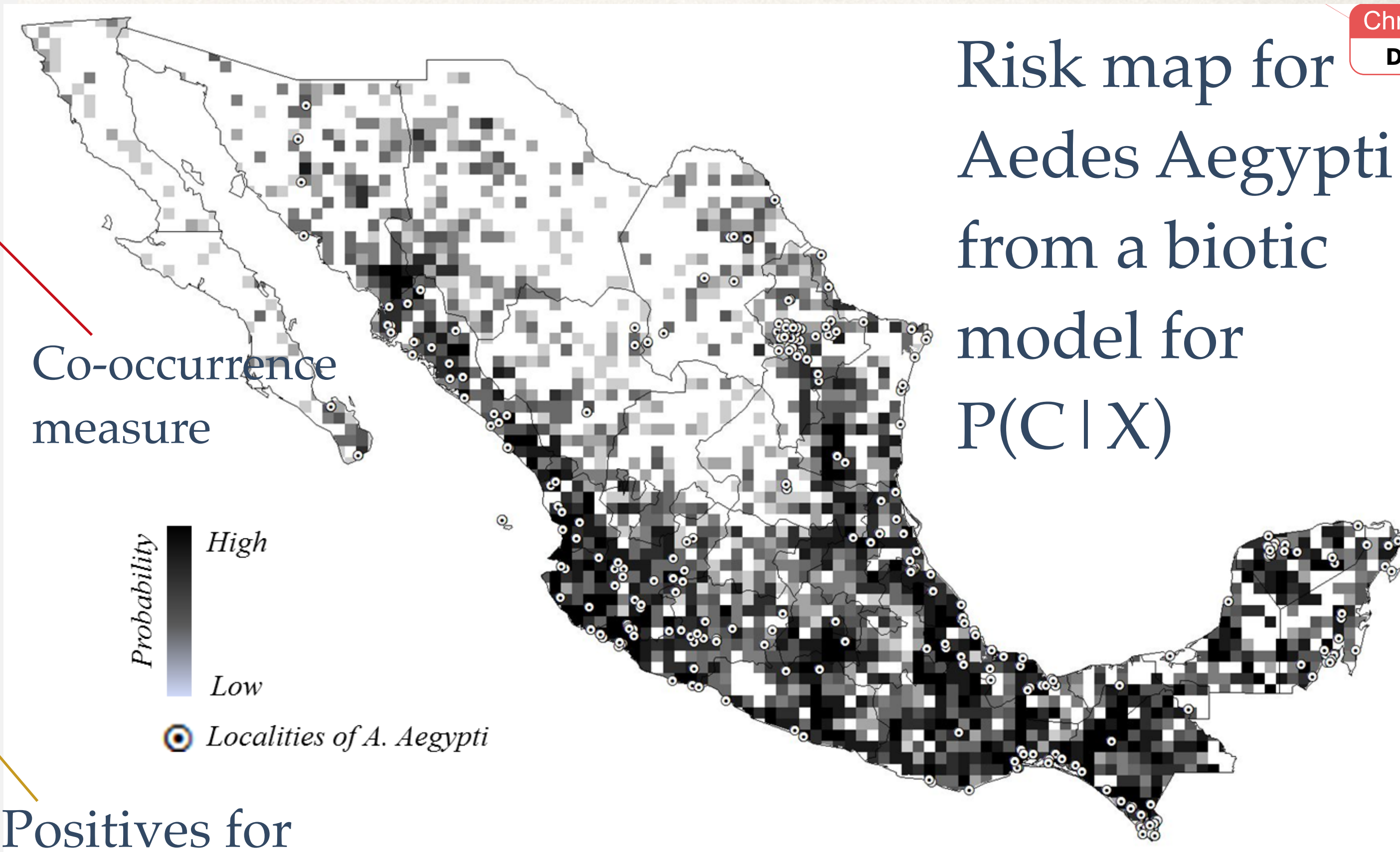
Predictive Model for potential hosts of ZIKV

Mammals with most statistically significant geographic overlap with Aedes Aegypti

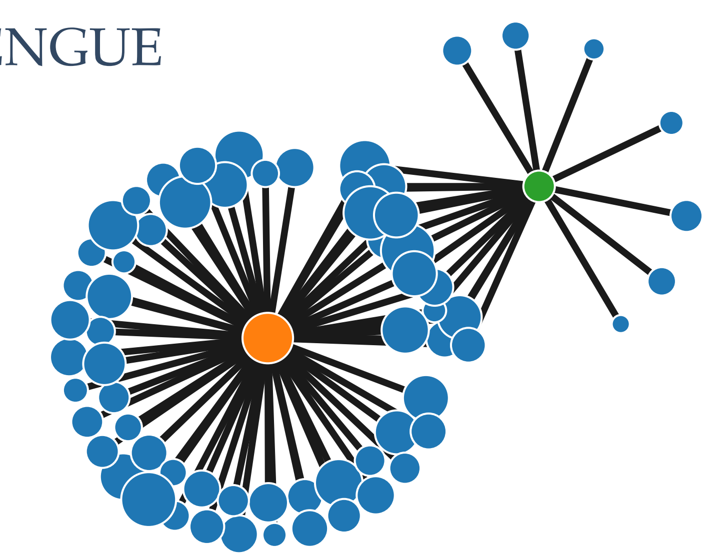
Rank	Mammal	epsilon	Rank	Mammal	epsilon
1	*Glossophaga soricina*	12.78	38	*Dasypus novemcinctus*	7.11
2	*Molossus rufus*	11.99	39	*Sigmodon hispidus*	7.02
3	*Artibeus jamaicensis**	11.68	40	*Uroderma bilobatum*	6.82
4	*Liomys pictus*	11.06	41	*Leptonycteris curasoae*	6.75
5	*Oryzomys couesi*	11.04	42	*Carollia perspicillata*	6.71
6	*Carollia subrufa*	10.49	43	*Centurio senex*	6.61
7	*Sturnira lilium*	10.28	44	*Sciurus colliaei*	6.59
8	*Artibeus lituratus**	9.91	45	*Lontra longicaudis*	6.49
9	*Choeroniscus godmani*	9.42	46	*Didelphis marsupialis*	6.49
10	*Liomys salvini*	9.33	47	*Cratogeomys bulleri*	6.35
11	*Oligoryzomys fulvescens*	9.15	48	*Carollia sowelli**	6.27
12	*Dermanura phaeotis*	9.12	49	*Myotis elegans*	6.12
13	*Rhogeessa tumida*	9.06	50	*Myotis nigricans**	6.06
14	*Pteronotus personatus*	9.05	51	*Sigmodon arizonae*	6.00
15	*Baiomys musculus*	8.97	52	*Rhynchonycteris naso*	5.95
16	*Glossophaga commissarisi*	8.80	53	*Tlacuatzin canescens*	5.87
17	*Didelphis virginiana*	8.58	54	*Leopardus pardalis*	5.84
18	*Pteronotus parnellii**	8.58	55	*Caluromys derbianus*	5.78
19	*Orthogeomys hispidus*	8.53	56	*Molossus molossus*	5.76
20	*Sciurus aureogaster*	8.52	57	*Oryzomys rostratus*	5.76
21	*Molossus sinaloae*	8.51	58	*Osgoodomys banderanus*	5.76
22	*Desmodus rotundus*	8.23	59	*Myotis carteri*	5.66
23	*Saccopteryx bilineata*	8.22	60	*Micronycteris microtis*	5.52
24	*Lasiurus intermedius*	8.15	61	*Sylvilagus brasiliensis*	5.47
25	*Phyllostomus discolor*	8.12	62	*Sylvilagus floridanus*	5.37
26	*Philander opossum*	8.10	63	*Spermophilus annulatus*	5.36
27	*Peromyscus gymnotis*	7.90	64	*Peromyscus leucopus*	5.30
28	*Balantiopteryx plicata*	7.81	65	*Conepatus leuconotus*	5.30
29	*Eptesicus furinalis*	7.69	66	*Chaetodipus pernix*	5.27
30	*Pteronotus davyi*	7.55	67	*Sciurus yucatanensis*	5.23
31	*Dermanura tolteca*	7.48	68	*Sigmodon mascotensis*	5.13
32	*Sciurus variegatoides*	7.48	69	*Eira barbara*	5.12
33	*Mormoops megalophylla*	7.45	70	*Ateles geoffroyi*	5.11
34	*Oryzomys melanotis*	7.42	71	*Neotoma phenax*	5.07
35	*Artibeus intermedius*	7.40	72	*Noctilio leporinus*	5.06
36	*Chaetodipus artus*	7.20	73	*Reithrodontomys fulvescens*	4.95
37	*Nasua narica*	7.18			

Co-occurrence measure

Positives for DENGUE



Risk map for Aedes Aegypti from a biotic model for P(C | X)



Complex Inference Network for Aedes aegypti and Aedes albopictus

SPECIES

Sistema Para la ExploraCión de Información ESpacial

Colaboración con el CONABIO

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
8.- Ma. Juan Barrios
9.- Ing. Raúl Jiménez *
10.- Pedro Romero Martínez

http://species.conabio.mx

http://species.conabio.gob.mx/candidate/

http://species.conabio.gob.mx/dbdev/

Conclusions

- Prediction of risk factors and interventions for EIDs is of huge public health importance
- EIDs are Complex Adaptive Systems
    - Multi-factorial, multi-scale, multi-discipline —> multi-interaction
    - There are too many interactions to observe directly
    - Standard mathematical techniques model only a few factors
- The Data Revolution has made available large amounts of data with which their complex, adaptive nature may be better modelled
    - Spatio-temporal data about organisms, relative to each other (biotic) and relative to the environment (abiotic), can be used to deduce the nature of their interactions
    - This can be done at the niche level (one to many) and at the community level (many to many)
    - Obtaining and integrating data is a huge challenge - political and technical
- The optimal use of this data requires innovation in modelling using multiple techniques - from SIR-type models to agent-based modelling and the use of advanced machine learning and AI techniques.
- Our work on various zoonosis show the utility of innovative approaches that use data of arbitrary spatial resolution and format, such as predicting host range.
    - Importance of a Data-Predictions-Experiment production line approach to emerging diseases
    - Importance of a multi-pathogen, multi-vector, multi-host approach