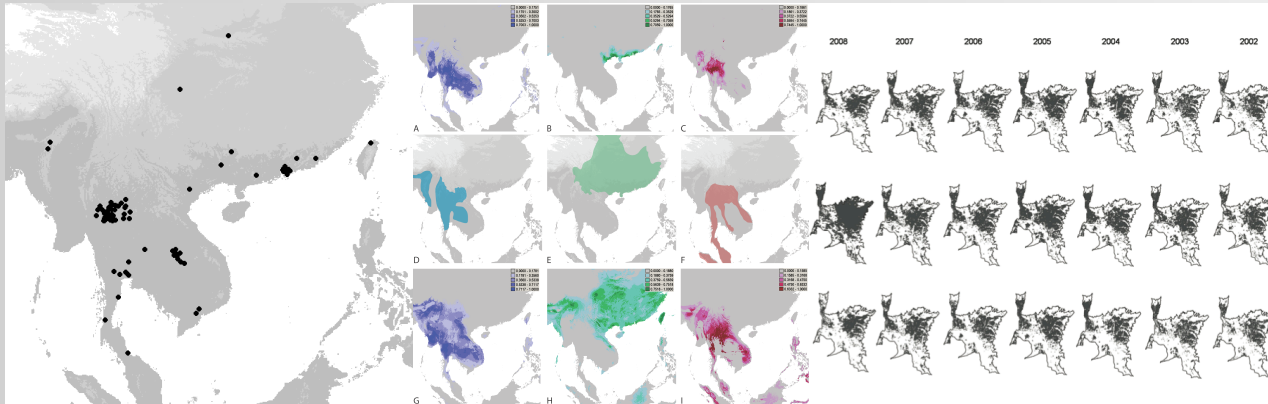
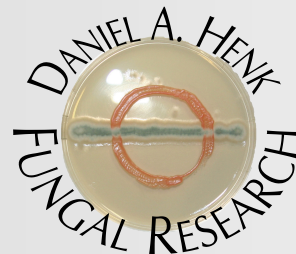


ECOLOGICAL NICHE MODELING AND INFECTIOUS DISEASE: CONSIDERING SPACE, TIME, AND EVOLUTION

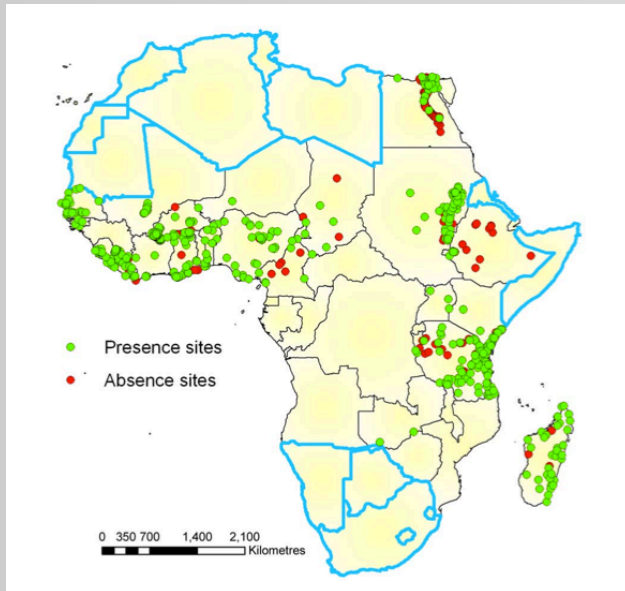


UNIVERSITY OF
BATH

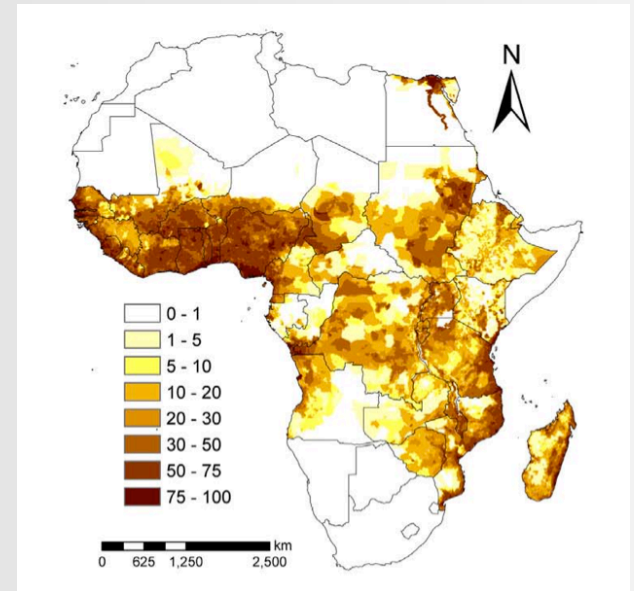


WHAT IS ECOLOGICAL NICHE MODELING?

Estimate the complete or possible geographic distribution for a species using correlation of 'environmental data' with observed occurrences.



ENM
+
Environmental data



OUTLINE

- The current tools
- The applications for infectious disease
 - Emerging disease
 - Environmental epidemiology
 - Forecasting under climate change
 - Predicting evolutionary trends
- The Rumsfeld status report (unknowns and known unknowns)
 - Temporal considerations
 - Sample space considerations
 - Genetic variation*

OUTLINE

- The current tools
- The applications for infectious disease
 - Emerging disease
 - Environmental epidemiology
 - Forecasting under climate change
 - Predicting evolutionary trends
- The Rumsfeld status report (unknowns and known unknowns)
 - Temporal considerations
 - Sample space considerations
 - Genetic variation*

I WILL ARGUE THAT THIS LAST POINT IS A LITTLE CONSIDERED HINGE POINT FOR ADVANCING MANY USES OF NICHE MODELS FOR MICROBES

THE TOOLS

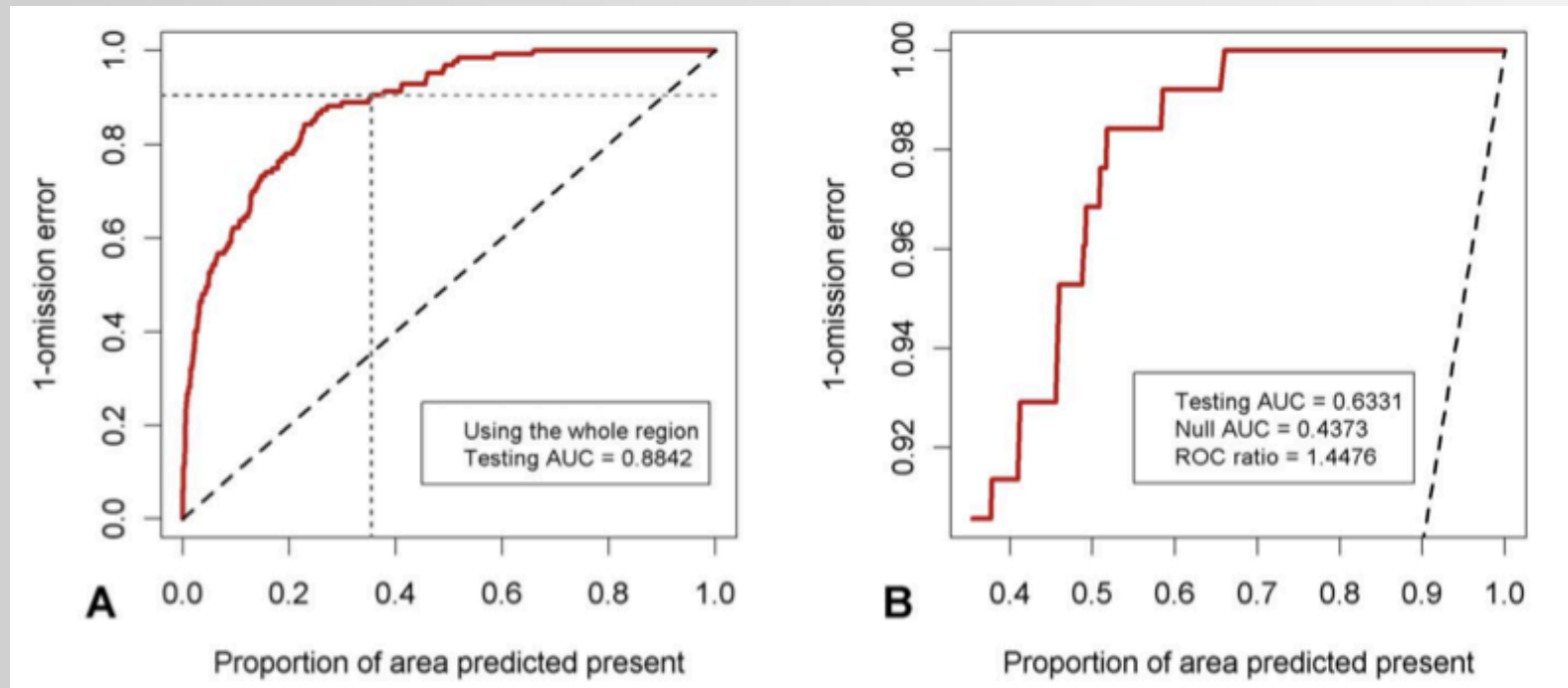
- Regression models
 - Generalized linear models (GLM)
- Machine learning models
 - Genetic algorithm for rule set prediction (GARP)
 - Maximum entropy (MaxEnt)
- Hybrid distribution-transmission dynamic models

- Regression models are common in disease prediction and in distribution estimation – they are the oldest of the methods – (24,000 records with search terms “niche model” and “generalized linear model”)
 - + Generally offers the best fit to any dataset
 - + Highly flexible inclusion of parameters (surrounding areas)
 - Requires true absence data

- GARP
 - Uses a genetic algorithm to gradually improve the fit of a rule set for the environmental data to the occurrence points
 - First accessible software and results in a smashing wave of studies (2180 results from 1999 – 2012)
 - + Does not require true absence data
 - + Easy to use with relatively straightforward cross dataset parameters
 - Poor performance generally with high error rates

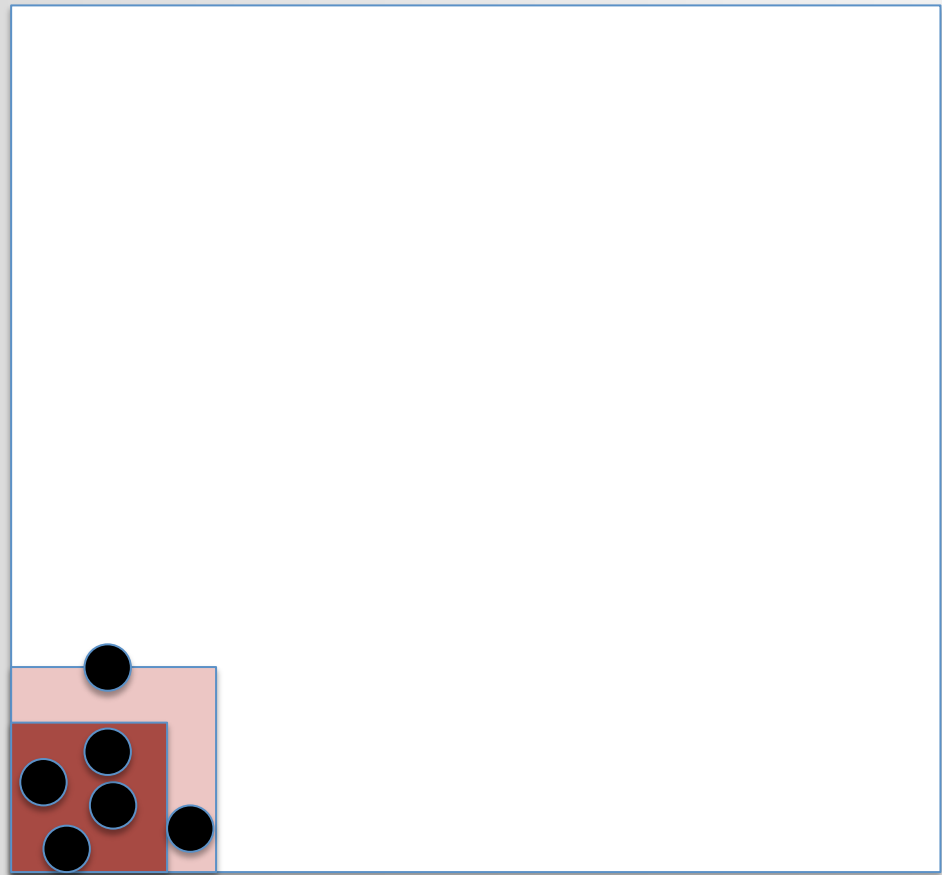
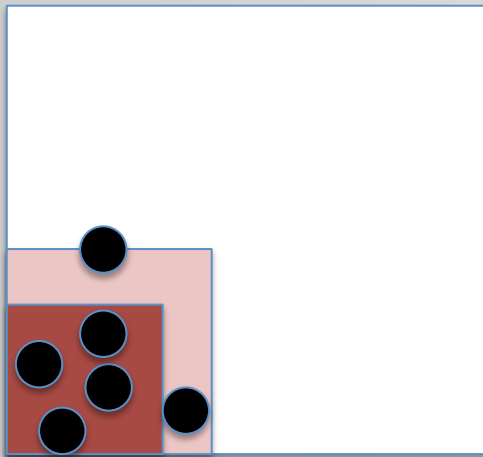
- MaxEnt
 - Attempts to estimate the probability distributions of the 'environmental data' (vectors of data points) that describe the observed occurrences while remaining nearest uniform and the background mean
 - Fast easy to use software has made it hugely popular despite the difficult underlying statistics and rationale (from 2006 to 2012 over 1940 published studies using this method)
- + Presence only data
- Complex estimators

- All of the these methods depend heavily on receiver operating characteristic curves to assess how well they are working.

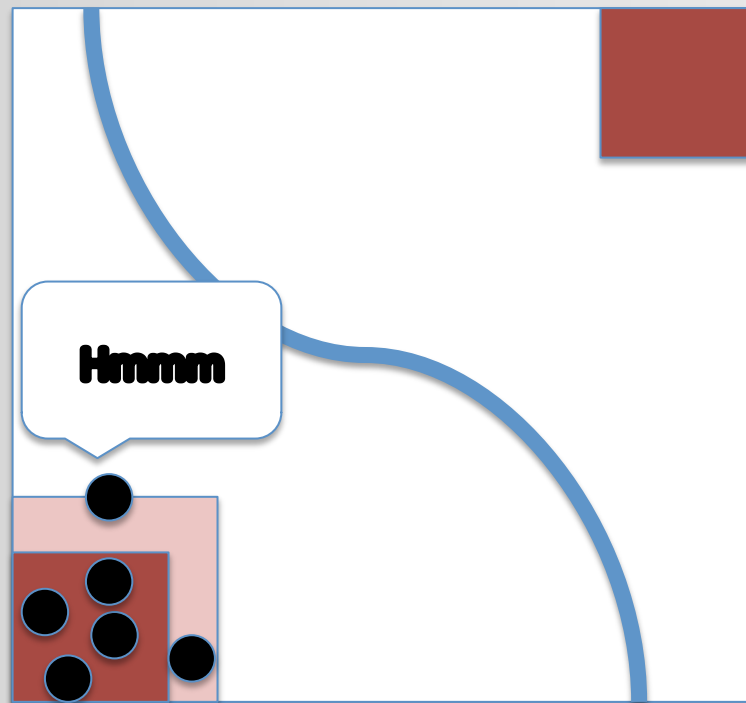


- Even this simple metric can be gamed unintentionally!

Adding white space composed of 'background' results in inflated AUC



- All of the these methods make extreme assumptions about equilibrium between location data and environmental variables and niches without any explicit consideration of time or dispersal.



- Hybrid distribution-dynamic models
 - Attempts to combine the niche-models with disease transmission models and avoid some of the assumptions about equilibrium. Can use dispersal kernels and can include standard SIR information.
 - Very few actual implementations
- + Accounts for change in distributions
 - Both data and model intensive – no easily generalizable approach and requires multiple layers of analysis
 - Still does not account for change associated with emergence

APPLICATIONS FOR INFECTIOUS DISEASE

Emerging disease – These applications generally attempt to assess the the potential distribution of diseases that are currently limited in range by dispersal factors.

Review Article

***Cryptococcus gattii*: Emergence in Western North America: Exploitation of a Novel Ecological Niche**

Kausik Datta,¹ Karen H. Bartlett,² and Kieren A. Marr¹

APPLICATIONS FOR INFECTIOUS DISEASE

Obvious problems related to the dynamic nature of emerging disease

Probably can only be assessed accurately with dynamic multidisciplinary approaches rather than simple niche model approaches alone.

Volume 19, Number 2—February 2013

Research

**Phylogenetic and Ecologic Perspectives of a Monkeypox
Outbreak, Southern Sudan, 2005**

APPLICATIONS FOR INFECTIOUS DISEASE

Environmental epidemiology

A standard approach for environmental diseases. Potentially identifying the sources and drivers of disease in the unsampled environment

These studies generally assume that where disease organism is located is correlated with where infections appear (clinical cases contribute to the niche model)

Ecological Niche Modeling to Estimate the Distribution of Japanese Encephalitis Virus in Asia

[Robin H. Miller](#),^{1*} [Penny Masuoka](#),¹ [Terry A. Klein](#),² [Heung-Chul Kim](#),³ [Todd Somer](#),⁴ and [John Grieco](#)¹


APPLICATIONS FOR INFECTIOUS DISEASE

Forecasting under climate change

Special issue in that future habitats have not been seen by the training model.

Retrospective studies show that niches change through time

**Climate Change and Risk of Leishmaniasis in North America:
Predictions from Ecological Niche Models of Vector and Reservoir
Species**

Camila González, Ophelia Wang, Stavana E. Strutz, Constantino González-Salazar, Víctor Sánchez-Cordero, Sahotra Sarkar 

APPLICATIONS FOR INFECTIOUS DISEASE

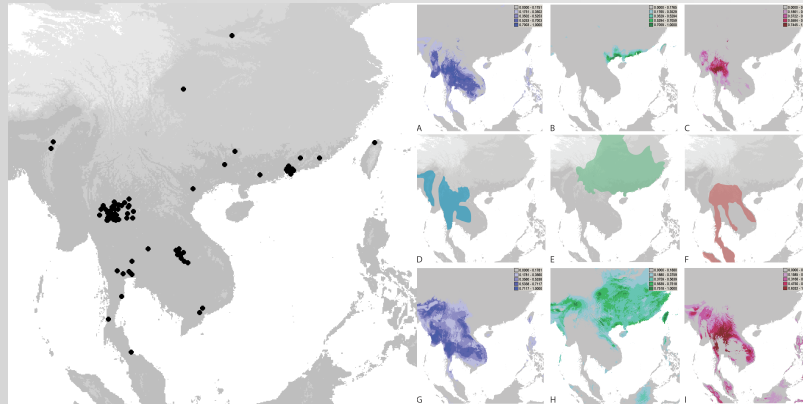
Evolutionary trends

Are niches phylogenetically conserved?

When aren't they?

How have species niches evolved through time

phylogeny examples - new 'cryptic species' new niche



THE RUMMY REPORT



START WITH THE KNOWN UNKNOWNNS

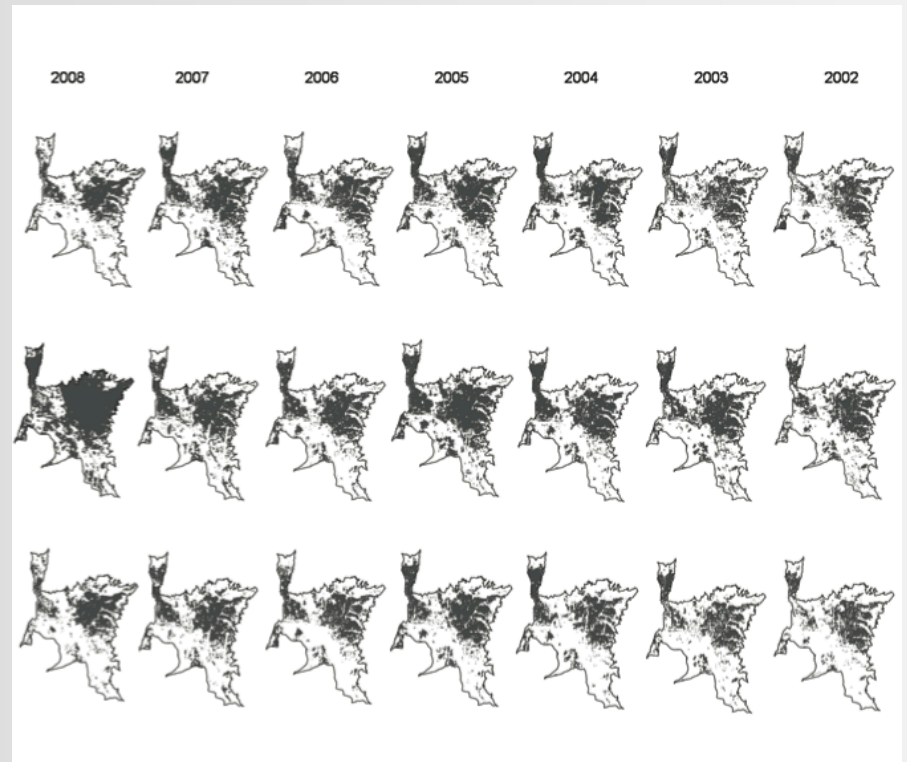


- Temporal Considerations
 - Are the occurrence data contemporary with the environmental data?
 - Does habitat use change with time?
 - Is distribution likely expanding or contracting?
 - Does a niche model transfer to other temporal space?

START WITH THE KNOWN UNKNOWNNS



- Can temporal cross validation answer the transfer question and the rate of change question? Or are they really mutually exclusive?

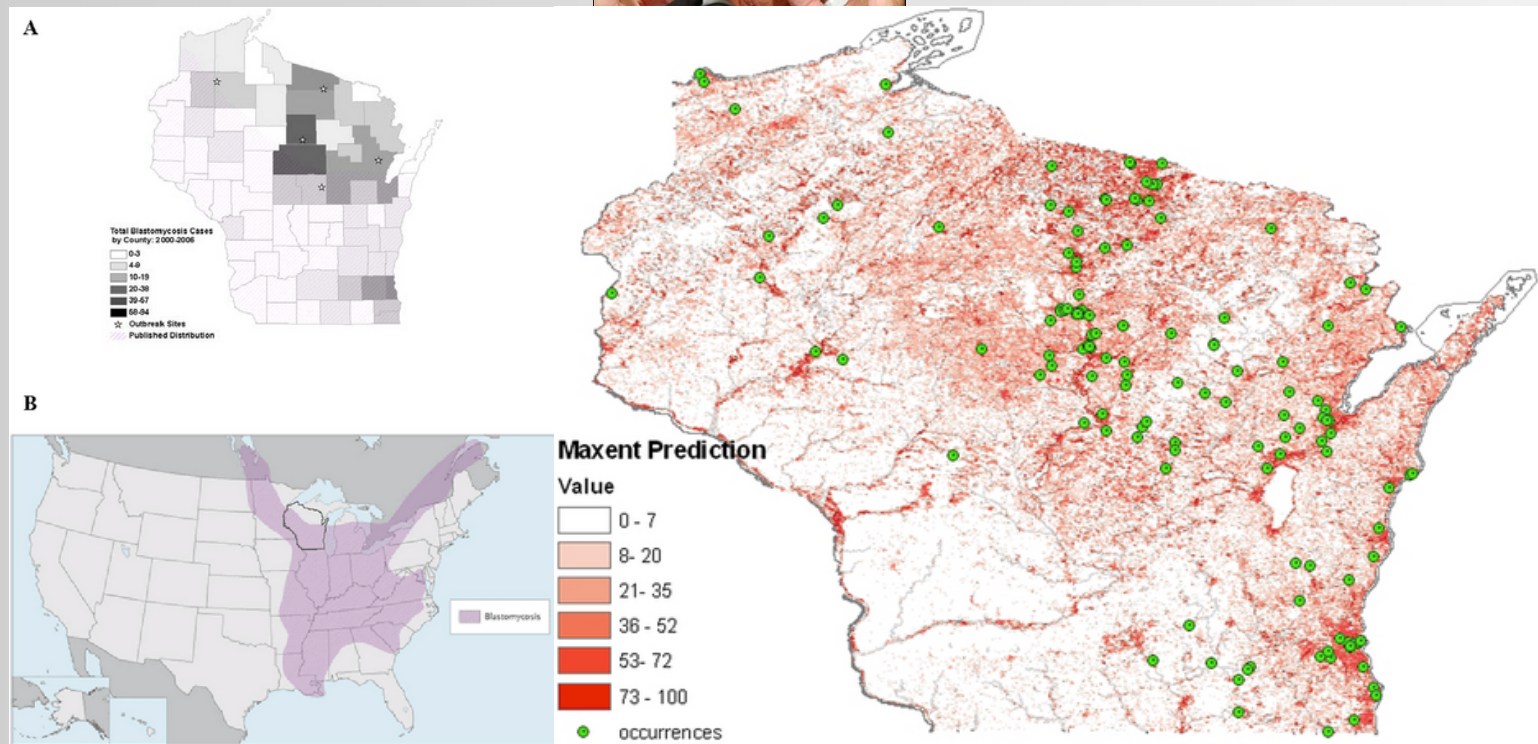


START WITH THE KNOWN UNKNOWNNS



- Spatial considerations
 - Probably the best studied of the known unknowns.
 - Multiple techniques to improve functionality of algorithms.
 - Must be tailored to the question at hand, but even then it is difficult to assess uncertainty.

START WITH THE KNOWN UNKNOWNNS



START WITH THE KNOWN UNKNOWNNS

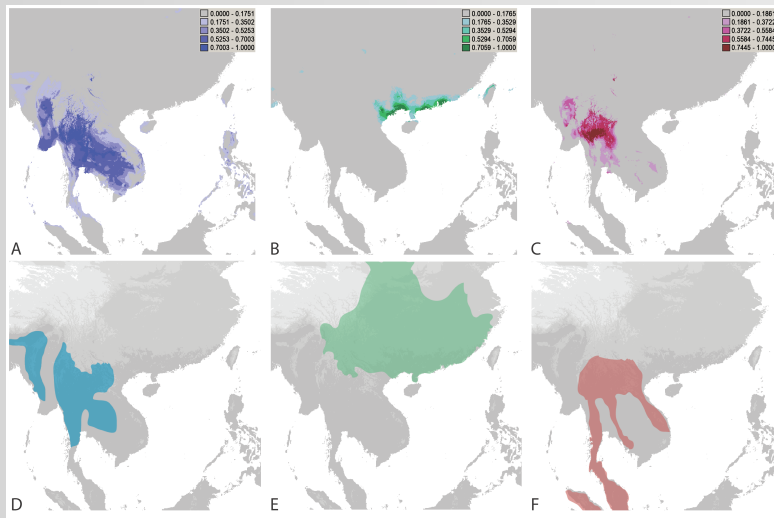


Main conclusions The large inter-specific variability observed among the 54 study species underlines the need to consider more than a few species to test properly the transferability of species distribution models. The pronounced asymmetry in transferability between the two study regions may be due to peculiarities of these regions, such as differences in the ranges of environmental predictors or the varied impact of land-use history, or to species-specific reasons like differential phenotypic plasticity, existence of ecotypes or varied dependence on biotic interactions that are not properly incorporated into niche-based models. The lower variation between internal and external evaluation of GLMs compared to GAMs further suggests that overfitting may reduce transferability. Overall, a limited geographical transferability calls for caution

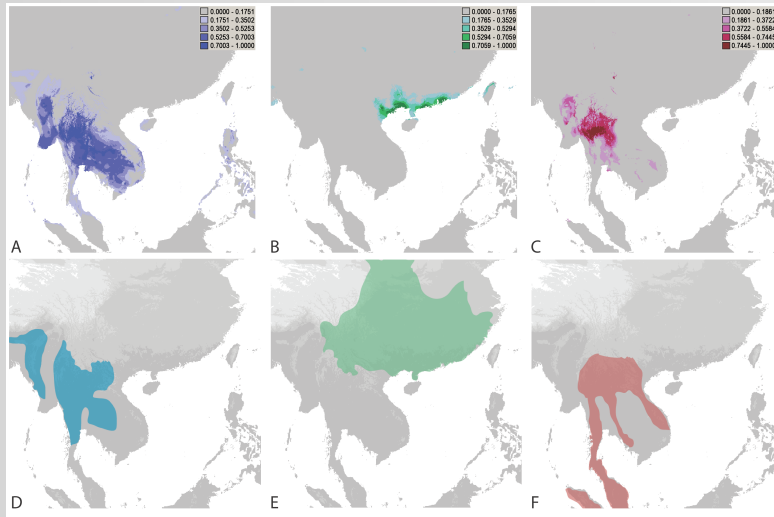
THE UNKNOWNNS



- Genetic Variation
 - Almost never considered explicitly within a given model but implicit in all of the models.
 - Evidence from the model users themselves that this variable is important



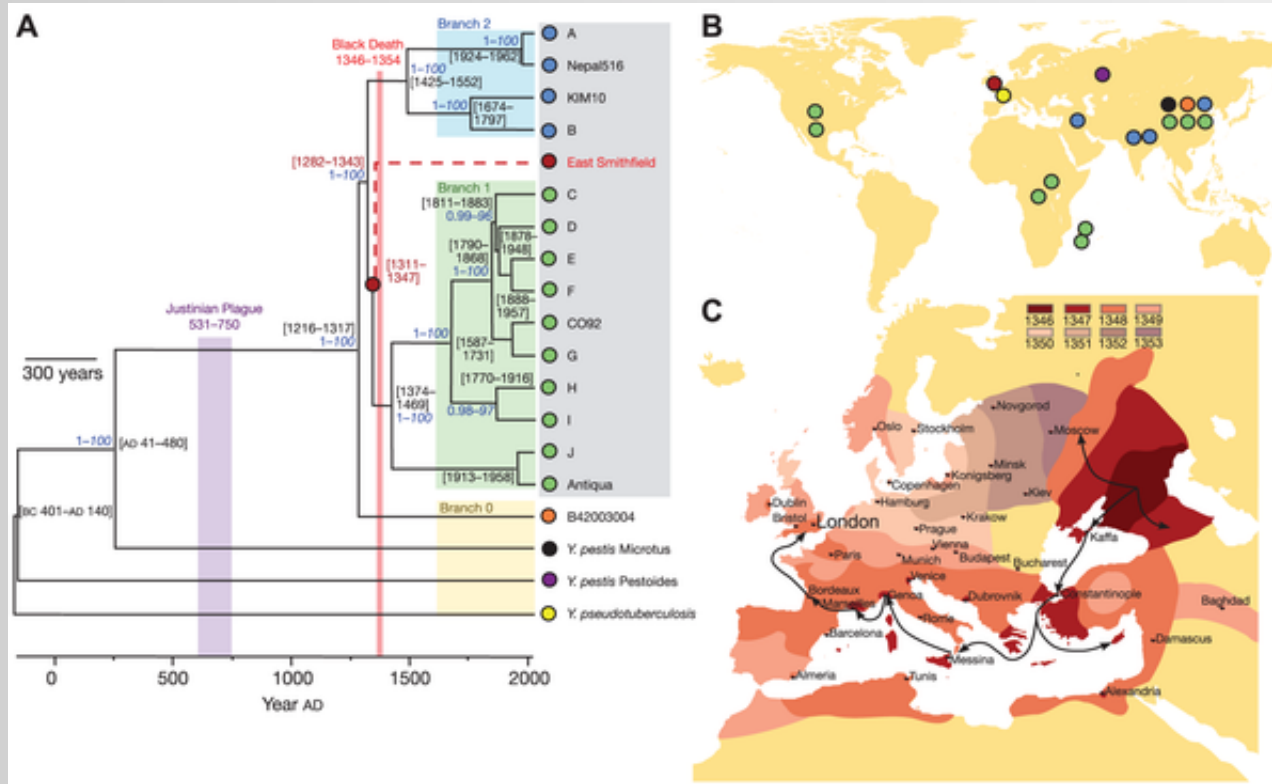
- Many niche models show that even before a majority of the genome no longer has polymorphism segregating between species, the ‘species’ have divergent niches. This is true for many taxa, but is it more important for microbes?



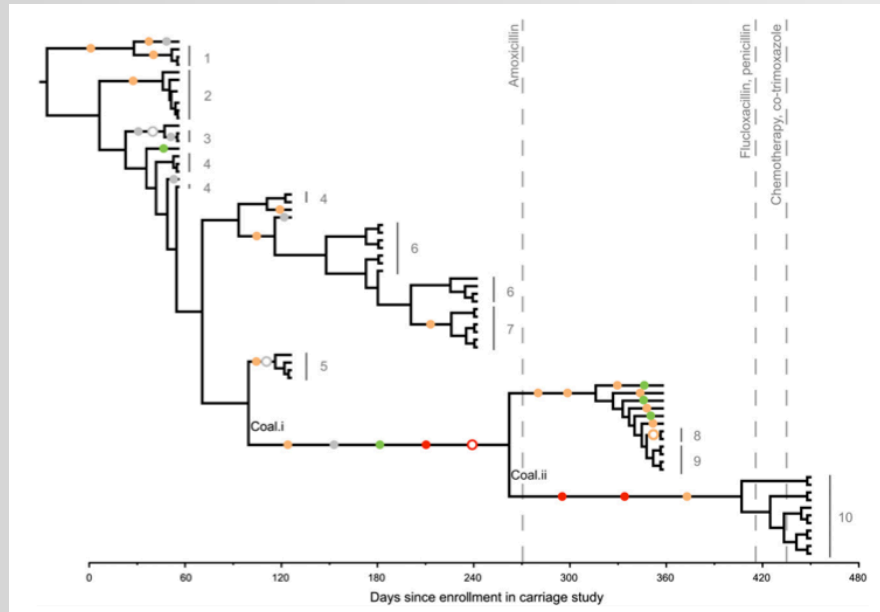
- Many niche models show that even before a majority of the genome no longer has polymorphism segregating between species, the ‘species’ have divergent niches. This is true for many taxa, but is it more important for microbes?

They can diversify quickly into CRYPTIC species that can misinform models.

Why else is this so relevant to microbial pathogens?



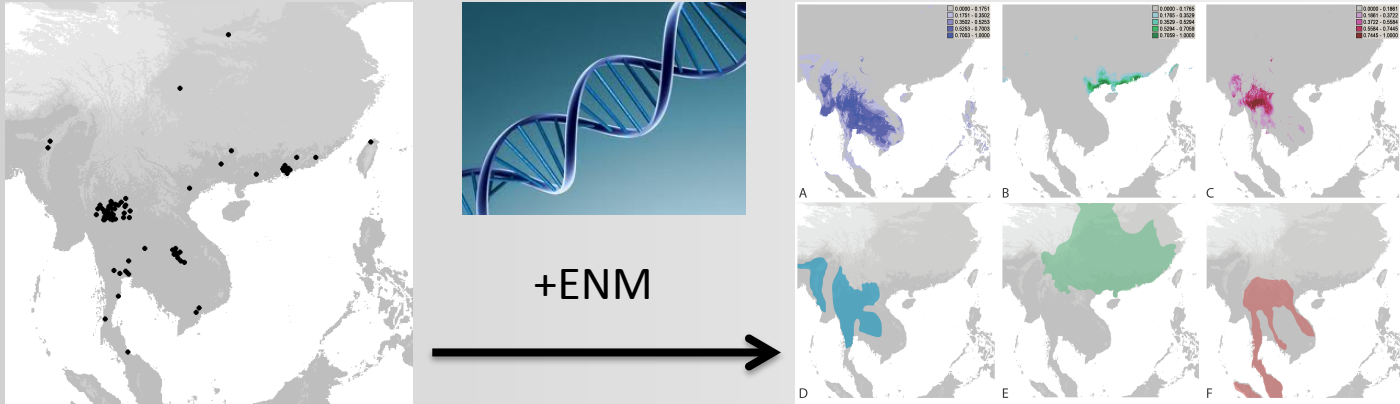
Why is this so relevant to microbial pathogens?



They change fast and in a biased way.

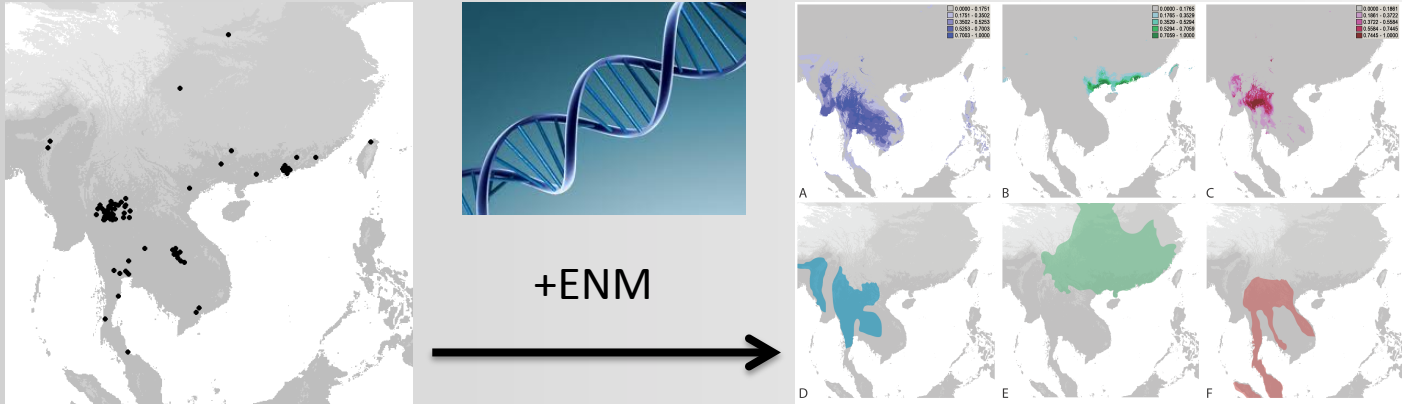
They adapt!

Can genetic data be incorporated into niche models?



- Option 1 – use hindsight of temporal transfer to estimate trends in niche change (very much along the lines of current evolutionary approaches undertaken at phylogenetic scale).
 - However, these models still suffer from the ‘species’ identifiability issue. How can one clearly identify ‘species’ when they are in flux.

Can genetic data be incorporated into niche models?



- Option 2 – Pile all of the genetic data into a multi-genetic model?