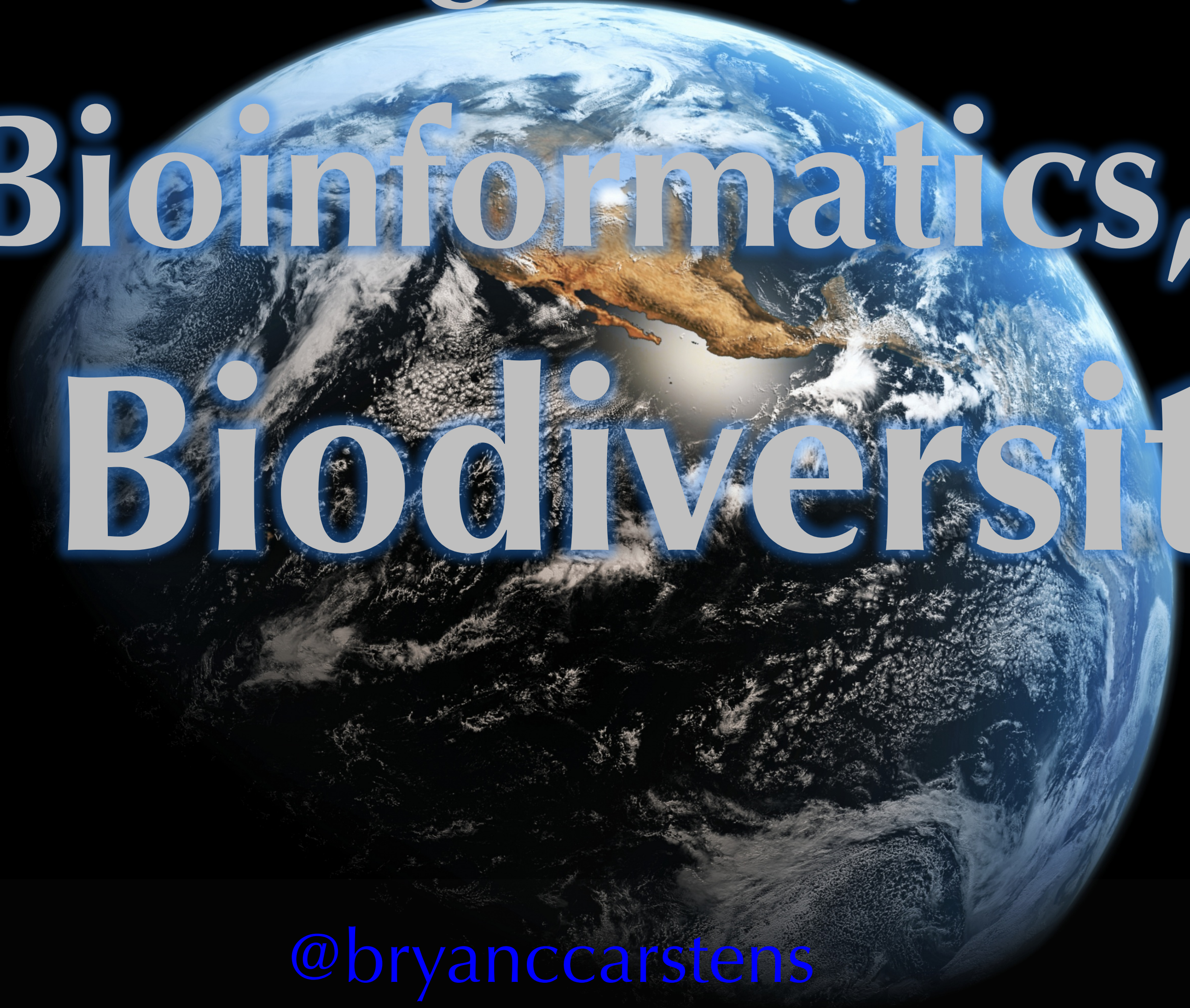
Big data, Bioinformatics, & Biodiversity.



@bryanccarstens

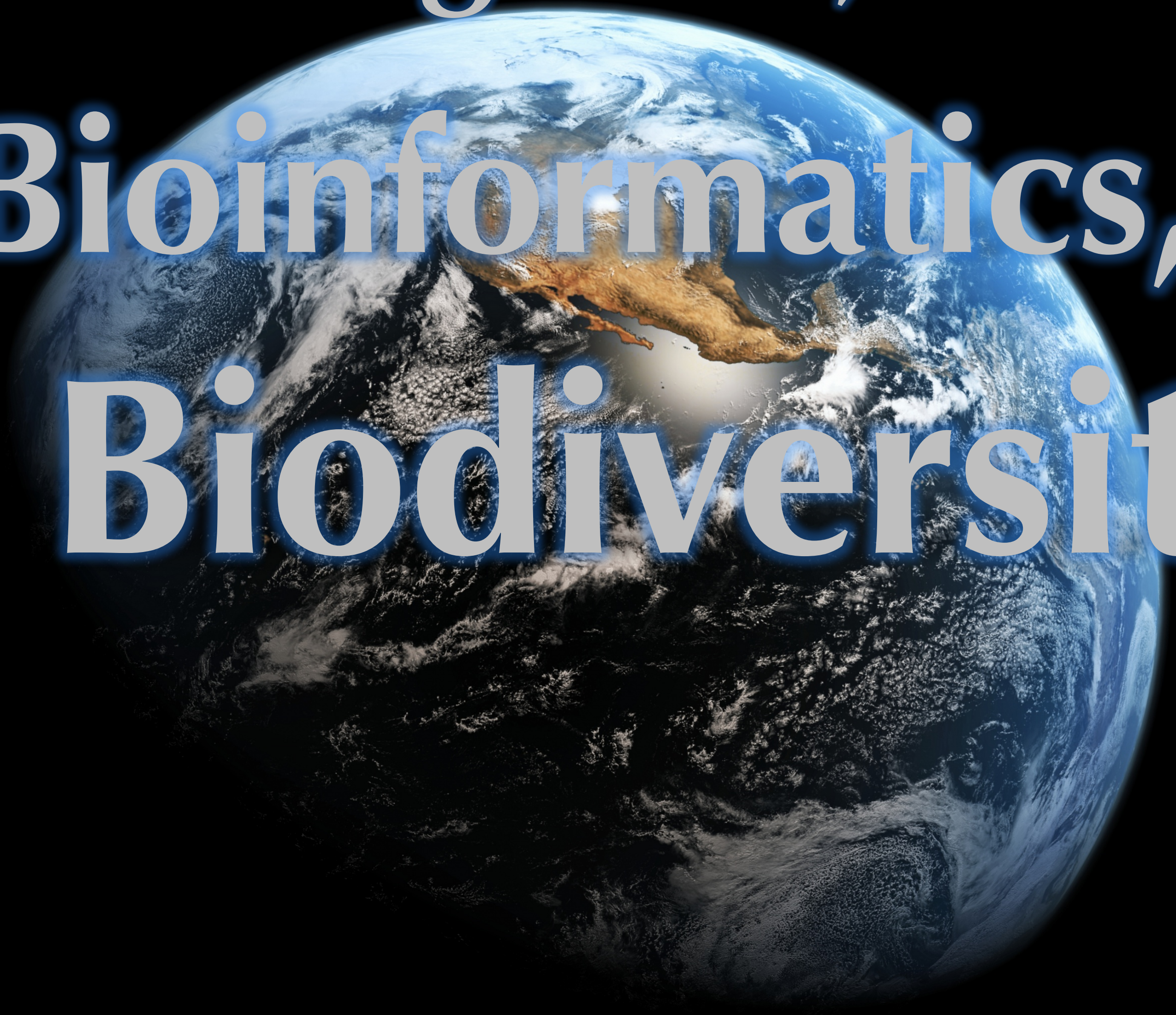
carstens.12@osu.edu

https://carstenslab.osu.edu

DEPARTMENT OF
EVOLUTION,
ECOLOGY AND
ORGANISMAL
BIOLOGY
eeob.osu.edu

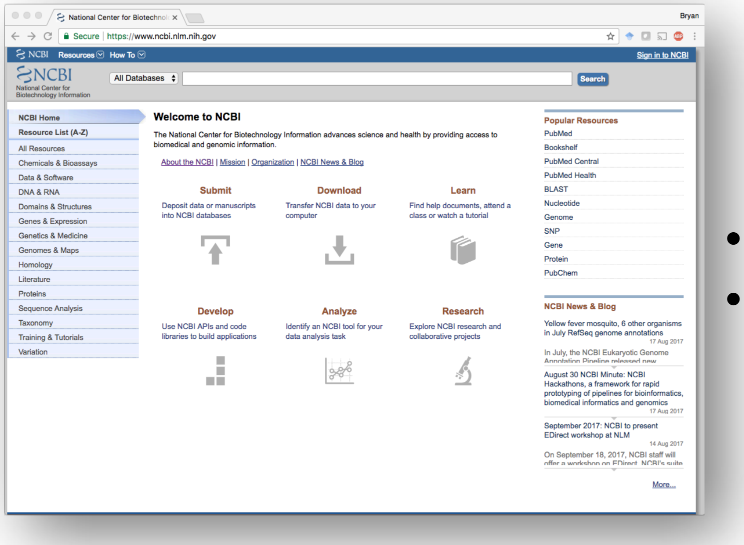
THE
OHIO
STATE
UNIVERSITY

Big data, Bioinformatics, & Biodiversity.



Tara Pelletier, Ariadna Morales, Greg Wheeler, Megan Smith

The biological sciences are data rich…



Welcome to NCBI

The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.

About the NCBI | Mission | Organization | NCBI News & Blog

Submit
Deposit data or manuscripts into NCBI databases

Download
Transfer NCBI data to your computer

Learn
Find help documents, attend a class or watch a tutorial

Develop
Use NCBI APIs and code libraries to build applications

Analyze
Identify an NCBI tool for your data analysis task

Research
Explore NCBI research and collaborative projects

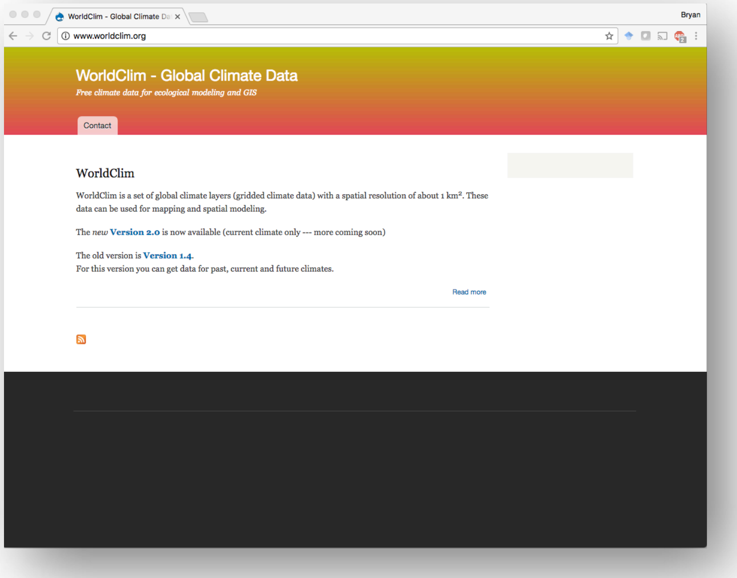
NCBI News & Blog

Yellow fever mosquito, 6 other organisms in July RefSeq genome annotations
17 Aug 2017
In July, the NCBI Eukaryotic Genome Annotation Pipeline released new

August 30 NCBI Minute: NCBI Hackathons, a framework for rapid prototyping of pipelines for bioinformatics, biomedical informatics and genomics
17 Aug 2017

September 2017: NCBI to present EDirect workshop at NLM
14 Aug 2017
On September 18, 2017, NCBI staff will offer a workshop on EDirect, NCBI's suite

- Gene sequences
- Climate data



WorldClim - Global Climate Data

Free climate data for ecological modeling and GIS

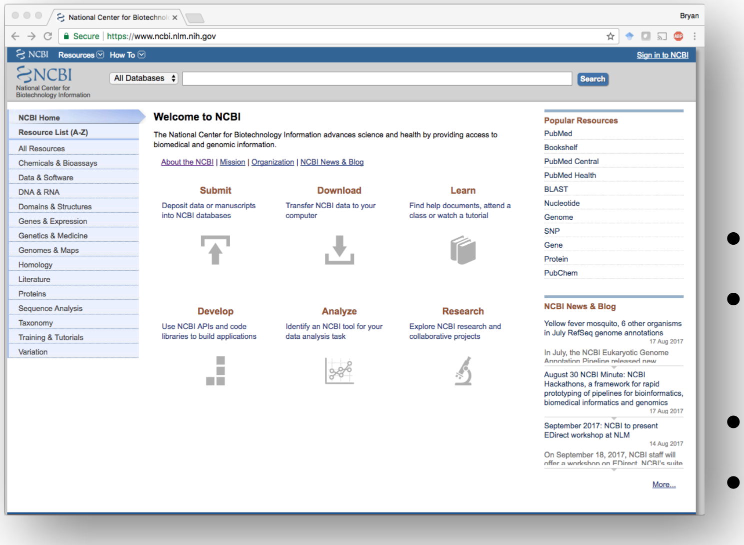
WorldClim

WorldClim is a set of global climate layers (gridded climate data) with a spatial resolution of about 1 km^2. These data can be used for mapping and spatial modeling.

The *new* **Version 2.0** is now available (current climate only --- more coming soon)

The old version is **Version 1.4**.
For this version you can get data for past, current and future climates.

The biological sciences are data rich…



Welcome to NCBI

The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.

About the NCBI | Mission | Organization | NCBI News & Blog

Submit — Deposit data or manuscripts into NCBI databases

Download — Transfer NCBI data to your computer

Learn — Find help documents, attend a class or watch a tutorial

Develop — Use NCBI APIs and code libraries to build applications

Analyze — Identify an NCBI tool for your data analysis task

Research — Explore NCBI research and collaborative projects

NCBI News & Blog

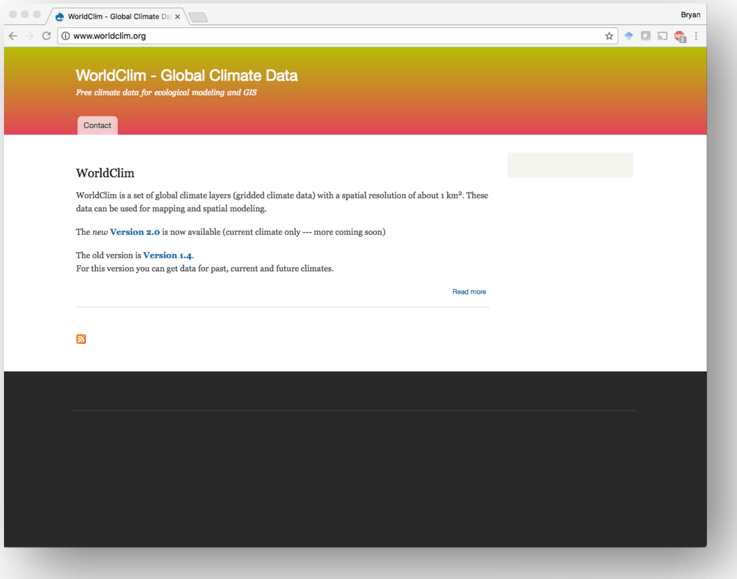
Yellow fever mosquito, 6 other organisms in July RefSeq genome annotations
17 Aug 2017
In July, the NCBI Eukaryotic Genome Annotation Pipeline released new

August 30 NCBI Minute: NCBI Hackathons, a framework for rapid prototyping of pipelines for bioinformatics, biomedical informatics and genomics
17 Aug 2017

September 2017: NCBI to present EDirect workshop at NLM
14 Aug 2017
On September 18, 2017, NCBI staff will offer a workshop on EDirect, NCBI's suite

More...

- Gene sequences
- Climate data
- Collection localities
- Genomes



WorldClim - Global Climate Data

Free climate data for ecological modeling and GIS

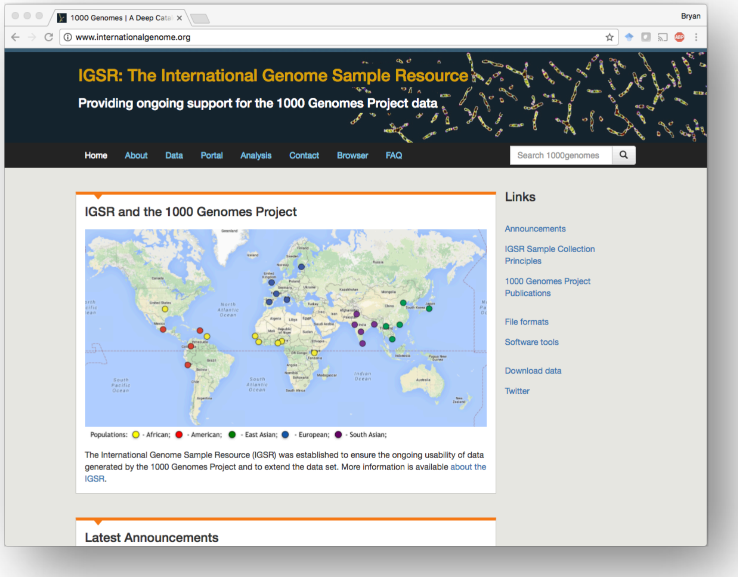
WorldClim

WorldClim is a set of global climate layers (gridded climate data) with a spatial resolution of about 1 km^2. These data can be used for mapping and spatial modeling.

The *new* **Version 2.0** is now available (current climate only --- more coming soon)

The old version is **Version 1.4**.
For this version you can get data for past, current and future climates.

Read more



IGSR: The International Genome Sample Resource

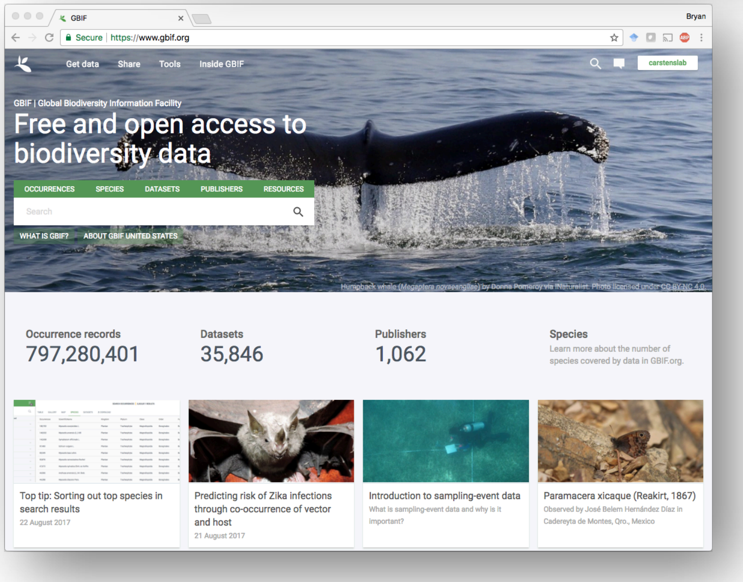
Providing ongoing support for the 1000 Genomes Project data

IGSR and the 1000 Genomes Project

Populations: ● - African; ● - American; ● - East Asian; ● - European; ● - South Asian;

The International Genome Sample Resource (IGSR) was established to ensure the ongoing usability of data generated by the 1000 Genomes Project and to extend the data set. More information is available about the IGSR.

Latest Announcements



GBIF | Global Biodiversity Information Facility

Free and open access to biodiversity data

Humpback whale (*Megaptera novaeangliae*) by Donna Pomeroy via iNaturalist. Photo licensed under CC BY-NC 4.0.

Occurrence records	Datasets	Publishers	Species
797,280,401	35,846	1,062	Learn more about the number of species covered by data in GBIF.org.

Top tip: Sorting out top species in search results
22 August 2017

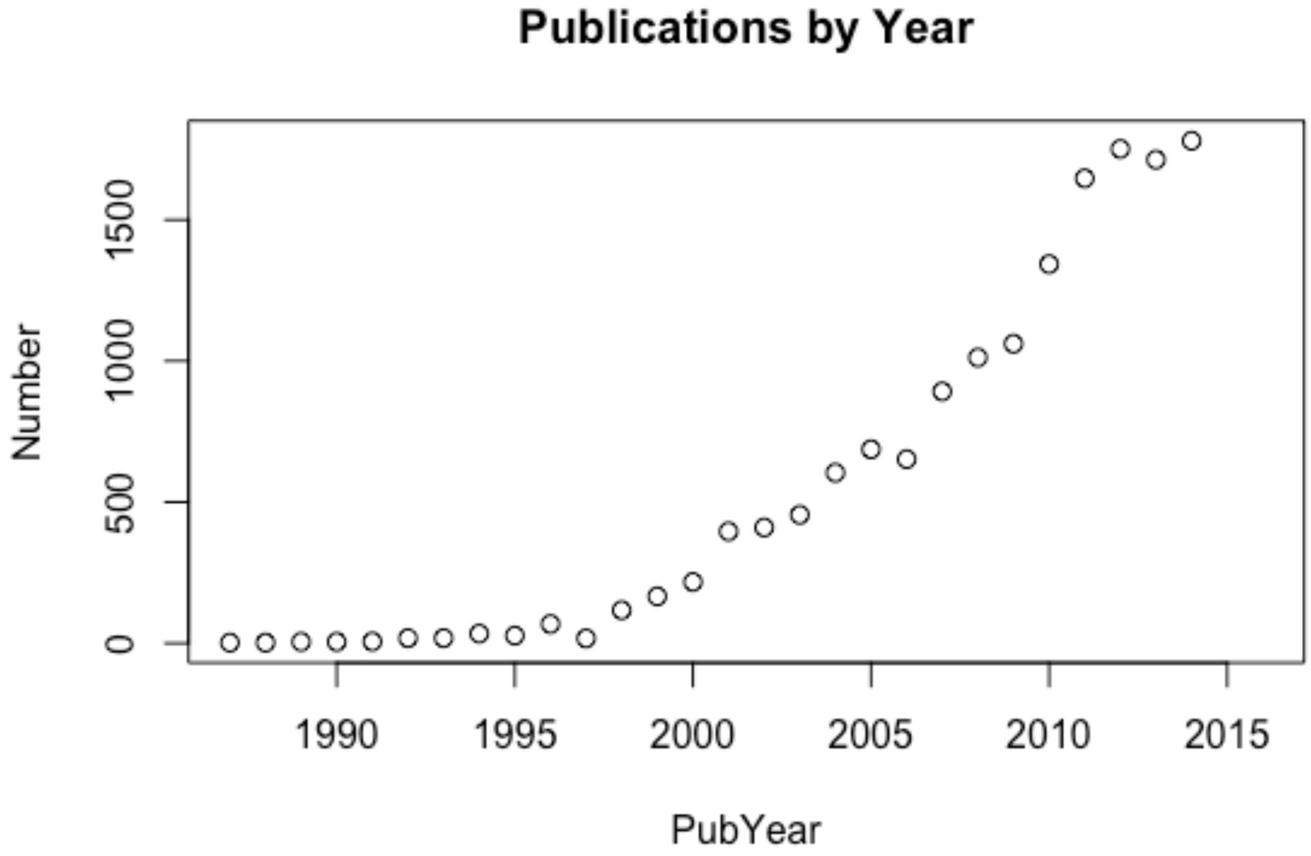
Predicting risk of Zika infections through co-occurrence of vector and host
21 August 2017

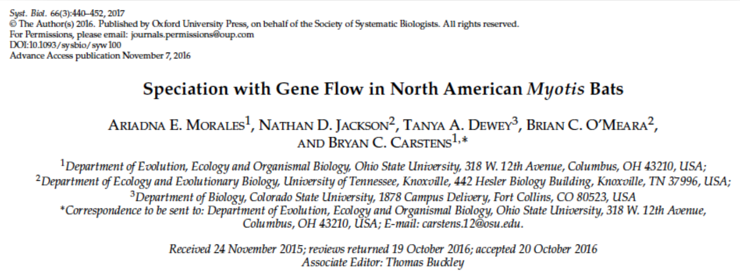
Introduction to sampling-event data
What is sampling-event data and why is it important?

Paramacera xicaque (Reakirt, 1867)
Observed by José Belem Hernández Díaz in Cadereyta de Montes, Qro., Mexico

search = ‘phylogeograph*’ - Web of Science 1987-2015

~40,000 peer-reviewed papers





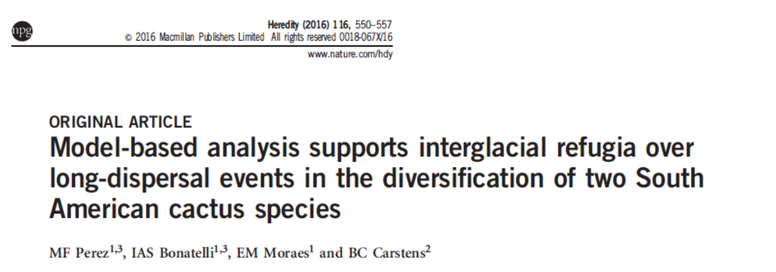
Syst. Biol. 66(3):440–452, 2017
© The Author(s) 2016. Published by Oxford University Press, on behalf of the Society of Systematic Biologists. All rights reserved.
For Permissions, please email: journals.permissions@oup.com
DOI:10.1093/sysbio/syw100
Advance Access publication November 7, 2016

Speciation with Gene Flow in North American *Myotis* Bats

ARIADNA E. MORALES[1], NATHAN D. JACKSON[2], TANYA A. DEWEY[3], BRIAN C. O'MEARA[2],
AND BRYAN C. CARSTENS[1,*]

[1]*Department of Evolution, Ecology and Organismal Biology, Ohio State University, 318 W. 12th Avenue, Columbus, OH 43210, USA;*
[2]*Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, 442 Hesler Biology Building, Knoxville, TN 37996, USA;*
[3]*Department of Biology, Colorado State University, 1878 Campus Delivery, Fort Collins, CO 80523, USA*
**Correspondence to be sent to: Department of Evolution, Ecology and Organismal Biology, Ohio State University, 318 W. 12th Avenue, Columbus, OH 43210, USA; E-mail: carstens.12@osu.edu.*

Received 24 November 2015; reviews returned 19 October 2016; accepted 20 October 2016
Associate Editor: Thomas Buckley

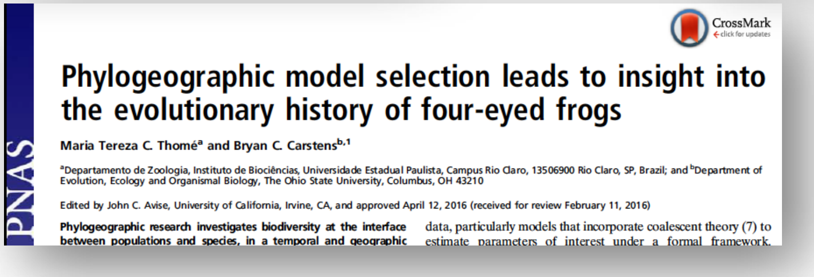


npg
Heredity (2016) 116, 550–557
© 2016 Macmillan Publishers Limited All rights reserved 0018-067X/16
www.nature.com/hdy

ORIGINAL ARTICLE

Model-based analysis supports interglacial refugia over long-dispersal events in the diversification of two South American cactus species

MF Perez[1,3], IAS Bonatelli[1,3], EM Moraes[1] and BC Carstens[2]



PNAS

CrossMark
click for updates

Phylogeographic model selection leads to insight into the evolutionary history of four-eyed frogs

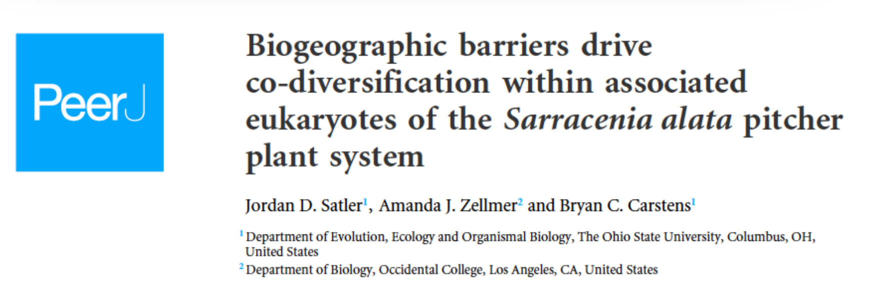
Maria Tereza C. Thomé[a] and Bryan C. Carstens[b,1]

[a]Departamento de Zoologia, Instituto de Biociências, Universidade Estadual Paulista, Campus Rio Claro, 13506900 Rio Claro, SP, Brazil; and [b]Department of Evolution, Ecology and Organismal Biology, The Ohio State University, Columbus, OH 43210

Edited by John C. Avise, University of California, Irvine, CA, and approved April 12, 2016 (received for review February 11, 2016)

Phylogeographic research investigates biodiversity at the interface between populations and species, in a temporal and geographic

data, particularly models that incorporate coalescent theory (7) to estimate parameters of interest under a formal framework.

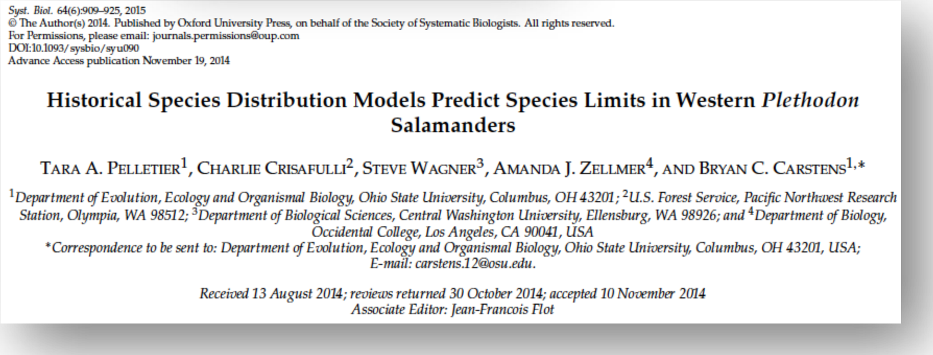


PeerJ

Biogeographic barriers drive co-diversification within associated eukaryotes of the *Sarracenia alata* pitcher plant system

Jordan D. Satler[1], Amanda J. Zellmer[2] and Bryan C. Carstens[1]

[1] Department of Evolution, Ecology and Organismal Biology, The Ohio State University, Columbus, OH, United States
[2] Department of Biology, Occidental College, Los Angeles, CA, United States



Syst. Biol. 64(6):909–925, 2015
© The Author(s) 2014. Published by Oxford University Press, on behalf of the Society of Systematic Biologists. All rights reserved.
For Permissions, please email: journals.permissions@oup.com
DOI:10.1093/sysbio/syu090
Advance Access publication November 19, 2014

Historical Species Distribution Models Predict Species Limits in Western *Plethodon* Salamanders

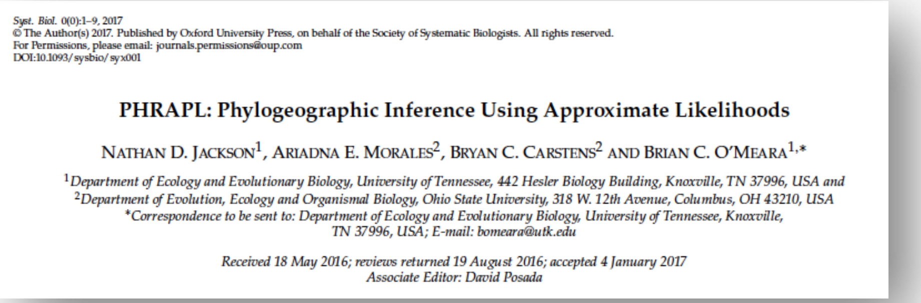
TARA A. PELLETIER[1], CHARLIE CRISAFULLI[2], STEVE WAGNER[3], AMANDA J. ZELLMER[4], AND BRYAN C. CARSTENS[1,*]

[1]*Department of Evolution, Ecology and Organismal Biology, Ohio State University, Columbus, OH 43201;* [2]*U.S. Forest Service, Pacific Northwest Research Station, Olympia, WA 98512;* [3]*Department of Biological Sciences, Central Washington University, Ellensburg, WA 98926; and* [4]*Department of Biology, Occidental College, Los Angeles, CA 90041, USA*
**Correspondence to be sent to: Department of Evolution, Ecology and Organismal Biology, Ohio State University, Columbus, OH 43201, USA; E-mail: carstens.12@osu.edu.*

Received 13 August 2014; reviews returned 30 October 2014; accepted 10 November 2014
Associate Editor: Jean-Francois Flot

Phylogeographic methods development…

- **PHRAPL** (DEB-1257784)



Syst. Biol. 0(0):1–9, 2017

DOI:10.1093/sysbio/syx001

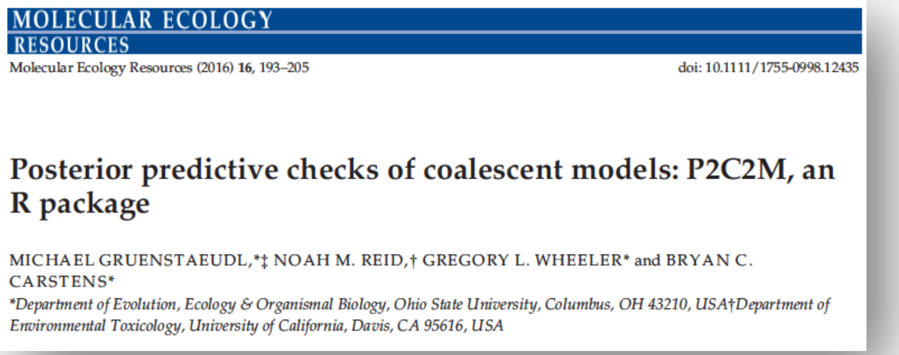
PHRAPL: Phylogeographic Inference Using Approximate Likelihoods

Nathan D. Jackson[1], Ariadna E. Morales[2], Bryan C. Carstens[2] and Brian C. O'Meara[1,*]

[1]*Department of Ecology and Evolutionary Biology, University of Tennessee, 442 Hesler Biology Building, Knoxville, TN 37996, USA and*
[2]*Department of Evolution, Ecology and Organismal Biology, Ohio State University, 318 W. 12th Avenue, Columbus, OH 43210, USA*
**Correspondence to be sent to: Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, TN 37996, USA; E-mail: bomeara@utk.edu*

Received 18 May 2016; reviews returned 19 August 2016; accepted 4 January 2017
Associate Editor: David Posada

- **P2C2M** (DBI-1661029)



MOLECULAR ECOLOGY
RESOURCES

Molecular Ecology Resources (2016) **16**, 193–205 doi: 10.1111/1755-0998.12435

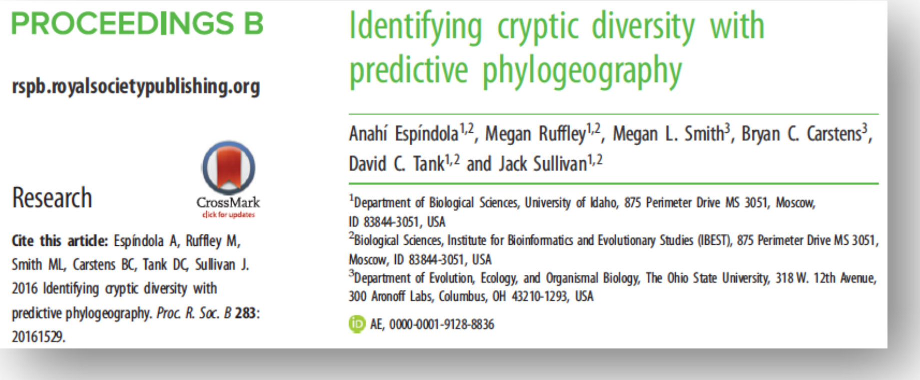
Posterior predictive checks of coalescent models: P2C2M, an R package

MICHAEL GRUENSTAEUDL,*‡ NOAH M. REID,† GREGORY L. WHEELER* and BRYAN C. CARSTENS*

**Department of Evolution, Ecology & Organismal Biology, Ohio State University, Columbus, OH 43210, USA†Department of Environmental Toxicology, University of California, Davis, CA 95616, USA*

Phylogeographic methods development…

- **Predictive phylogeography** (DEB-1457519)



PROCEEDINGS B

rspb.royalsocietypublishing.org

Research

Cite this article: Espíndola A, Ruffley M, Smith ML, Carstens BC, Tank DC, Sullivan J. 2016 Identifying cryptic diversity with predictive phylogeography. *Proc. R. Soc. B* **283**: 20161529.

Identifying cryptic diversity with predictive phylogeography

Anahí Espíndola[1,2], Megan Ruffley[1,2], Megan L. Smith[3], Bryan C. Carstens[3], David C. Tank[1,2] and Jack Sullivan[1,2]

[1]Department of Biological Sciences, University of Idaho, 875 Perimeter Drive MS 3051, Moscow, ID 83844-3051, USA
[2]Biological Sciences, Institute for Bioinformatics and Evolutionary Studies (IBEST), 875 Perimeter Drive MS 3051, Moscow, ID 83844-3051, USA
[3]Department of Evolution, Ecology, and Organismal Biology, The Ohio State University, 318 W. 12th Avenue, 300 Aronoff Labs, Columbus, OH 43210-1293, USA

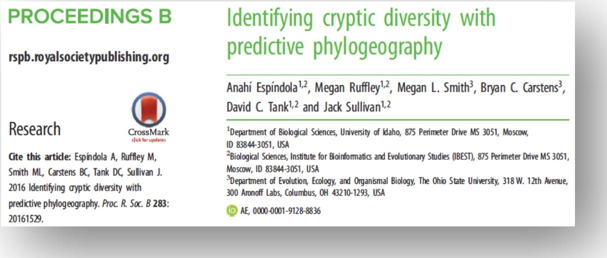
AE, 0000-0001-9128-8836

Phylogeographic methods development…

- **Predictive phylogeography** (DEB-1457519)

A **framework** that seeks to answer key questions relevant to organismal biology.

- **Multiple species** (regional to global, particular clades to broader taxonomic groups)
- **Integrative** – b/c incorporates all sorts of data (environmental, organismal, genetic)
- relies on **machine learning** to identify key variables



PROCEEDINGS B

rspb.royalsocietypublishing.org

Research

Cite this article: Espíndola A, Ruffley M, Smith ML, Carstens BC, Tank DC, Sullivan J. 2016 Identifying cryptic diversity with predictive phylogeography. *Proc. R. Soc. B* **283**: 20161529.

Identifying cryptic diversity with predictive phylogeography

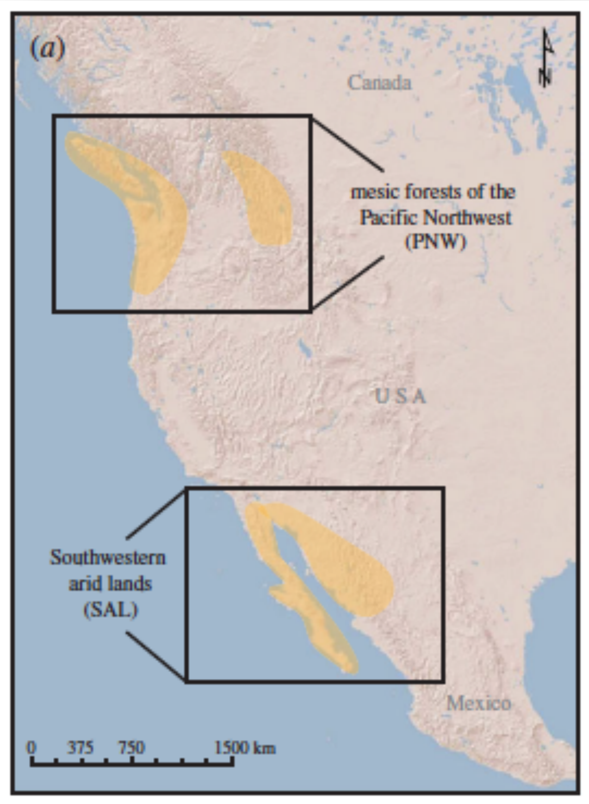
Anahí Espíndola[1,2], Megan Ruffley[1,2], Megan L. Smith[3], Bryan C. Carstens[3], David C. Tank[1,2] and Jack Sullivan[1,2]

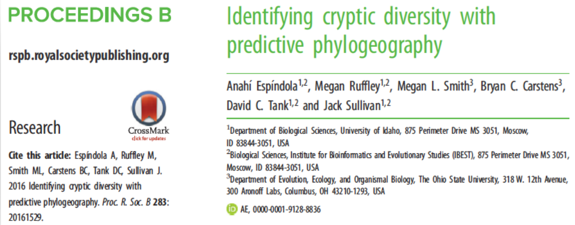
[1]Department of Biological Sciences, University of Idaho, 875 Perimeter Drive MS 3051, Moscow, ID 83844-3051, USA
[2]Biological Sciences, Institute for Bioinformatics and Evolutionary Studies (IBEST), 875 Perimeter Drive MS 3051, Moscow, ID 83844-3051, USA
[3]Department of Evolution, Ecology, and Organismal Biology, The Ohio State University, 318 W. 12th Avenue, 300 Aronoff Labs, Columbus, OH 43210-1293, USA

AE, 0000-0001-9128-8836

Building a predictive framework

Goal: to develop a predictive framework to identify species that are likely to contain cryptic diversity.





PROCEEDINGS B

rspb.royalsocietypublishing.org

Research

CrossMark

click for updates

Cite this article: Espíndola A, Ruffley M, Smith ML, Carstens BC, Tank DC, Sullivan J. 2016 Identifying cryptic diversity with predictive phylogeography. *Proc. R. Soc. B* **283**: 20161529.

Identifying cryptic diversity with predictive phylogeography

Anahí Espíndola[1,2], Megan Ruffley[1,2], Megan L. Smith[3], Bryan C. Carstens[3], David C. Tank[1,2] and Jack Sullivan[1,2]

[1] Department of Biological Sciences, University of Idaho, 875 Perimeter Drive MS 3051, Moscow, ID 83844-3051, USA

[2] Biological Sciences, Institute for Bioinformatics and Evolutionary Studies (IBEST), 875 Perimeter Drive MS 3051, Moscow, ID 83844-3051, USA

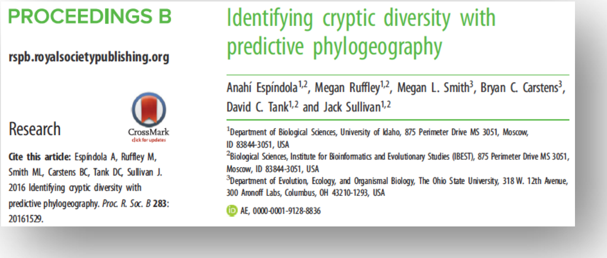
[3] Department of Evolution, Ecology, and Organismal Biology, The Ohio State University, 318 W. 12th Avenue, 300 Aronoff Labs, Columbus, OH 43210-1293, USA

AE, 0000-0001-9128-8836

Building a predictive framework

Goal: to develop a predictive framework to identify species that are likely to contain cryptic diversity.

1. Train the model using existing data.



PROCEEDINGS B

rspb.royalsocietypublishing.org

Research

CrossMark
click for updates

Cite this article: Espíndola A, Ruffley M, Smith ML, Carstens BC, Tank DC, Sullivan J. 2016 Identifying cryptic diversity with predictive phylogeography. *Proc. R. Soc. B* **283**: 20161529.

Identifying cryptic diversity with predictive phylogeography

Anahí Espíndola[1,2], Megan Ruffley[1,2], Megan L. Smith[3], Bryan C. Carstens[3], David C. Tank[1,2] and Jack Sullivan[1,2]

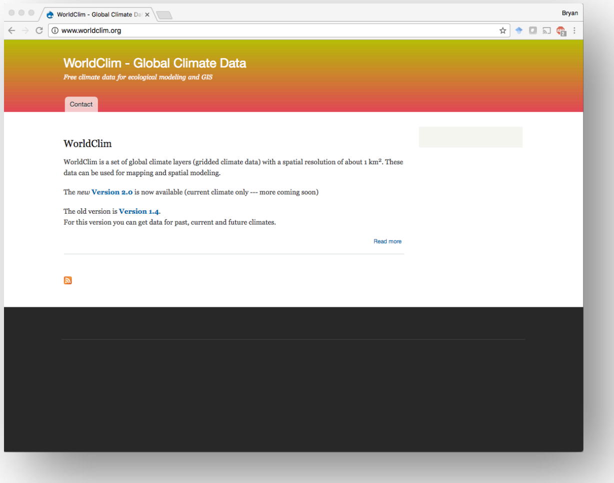
[1] Department of Biological Sciences, University of Idaho, 875 Perimeter Drive MS 3051, Moscow, ID 83844-3051, USA
[2] Biological Sciences, Institute for Bioinformatics and Evolutionary Studies (IBEST), 875 Perimeter Drive MS 3051, Moscow, ID 83844-3051, USA
[3] Department of Evolution, Ecology, and Organismal Biology, The Ohio State University, 318 W. 12th Avenue, 300 Aronoff Labs, Columbus, OH 43210-1293, USA

AE, 0000-0001-9128-8836

Building a predictive framework

Goal: to develop a predictive framework to identify species that are likely to contain cryptic diversity.

1. Train the model using existing data.



WorldClim - Global Climate Data

Free climate data for ecological modeling and GIS

Contact

WorldClim

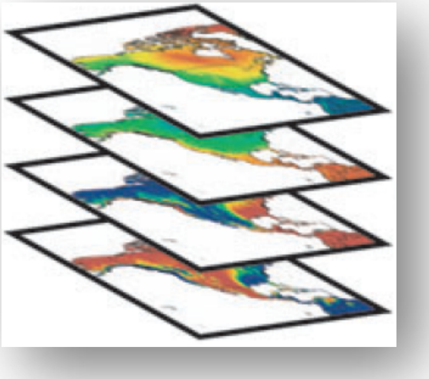
WorldClim is a set of global climate layers (gridded climate data) with a spatial resolution of about 1 km². These data can be used for mapping and spatial modeling.

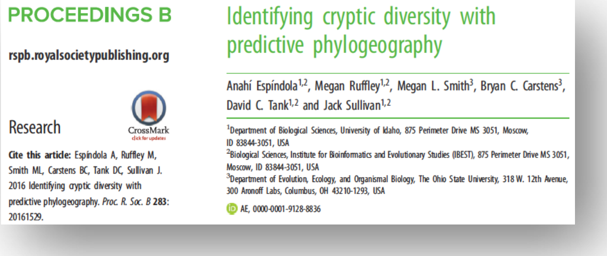
The *new* **Version 2.0** is now available (current climate only --- more coming soon)

The old version is **Version 1.4**.
For this version you can get data for past, current and future climates.

Read more

- climate data from WorldClim
- species distribution models for all taxa





PROCEEDINGS B

rspb.royalsocietypublishing.org

Research

CrossMark
click for updates

Cite this article: Espíndola A, Ruffley M, Smith ML, Carstens BC, Tank DC, Sullivan J. 2016 Identifying cryptic diversity with predictive phylogeography. *Proc. R. Soc. B* **283**: 20161529.

Identifying cryptic diversity with predictive phylogeography

Anahí Espíndola[1,2], Megan Ruffley[1,2], Megan L. Smith[3], Bryan C. Carstens[3], David C. Tank[1,2] and Jack Sullivan[1,2]

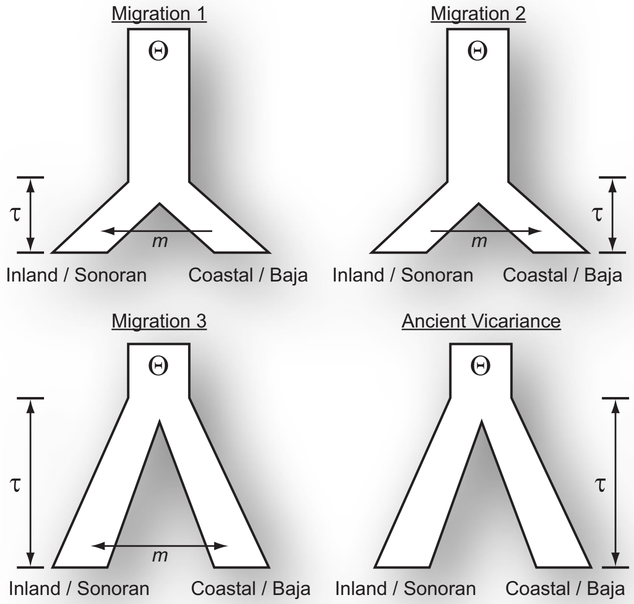
[1]Department of Biological Sciences, University of Idaho, 875 Perimeter Drive MS 3051, Moscow, ID 83844-3051, USA
[2]Biological Sciences, Institute for Bioinformatics and Evolutionary Studies (IBEST), 875 Perimeter Drive MS 3051, Moscow, ID 83844-3051, USA
[3]Department of Evolution, Ecology, and Organismal Biology, The Ohio State University, 318 W. 12th Avenue, 300 Aronoff Labs, Columbus, OH 43210-1293, USA

AE, 0000-0001-9128-8836

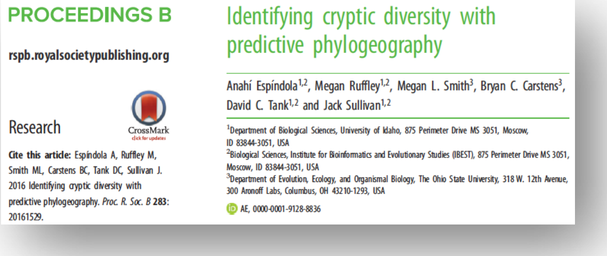
Building a predictive framework

Goal: to develop a predictive framework to identify species that are likely to contain cryptic diversity.

1. Train the model using existing data.



- genetic data (sequence, SNPs, MSATs)
- ABC used to calculate posterior probability of historical demographic models for all taxa



PROCEEDINGS B

rspb.royalsocietypublishing.org

Research

Cite this article: Espíndola A, Ruffley M, Smith ML, Carstens BC, Tank DC, Sullivan J. 2016 Identifying cryptic diversity with predictive phylogeography. *Proc. R. Soc. B* **283**: 20161529.

Identifying cryptic diversity with predictive phylogeography

Anahí Espíndola[1,2], Megan Ruffley[1,2], Megan L. Smith[3], Bryan C. Carstens[3], David C. Tank[1,2] and Jack Sullivan[1,2]

[1] Department of Biological Sciences, University of Idaho, 875 Perimeter Drive MS 3051, Moscow, ID 83844-3051, USA
[2] Biological Sciences, Institute for Bioinformatics and Evolutionary Studies (IBEST), 875 Perimeter Drive MS 3051, Moscow, ID 83844-3051, USA
[3] Department of Evolution, Ecology, and Organismal Biology, The Ohio State University, 318 W. 12th Avenue, 300 Aronoff Labs, Columbus, OH 43210-1293, USA

AE, 0000-0001-9128-8836

Building a predictive framework

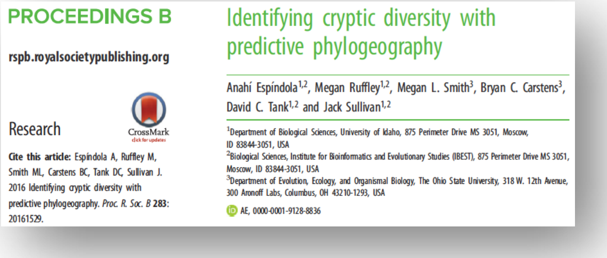
Goal: to develop a predictive framework to identify species that are likely to contain cryptic diversity.

2. Build data table

For each species in focal ecosystem:

1. species distribution model (climate layers)

2. evolutionary model (posterior probability)



PROCEEDINGS B

rspb.royalsocietypublishing.org

Research

CrossMark
click for updates

Cite this article: Espíndola A, Ruffley M, Smith ML, Carstens BC, Tank DC, Sullivan J. 2016 Identifying cryptic diversity with predictive phylogeography. *Proc. R. Soc. B* **283**: 20161529.

Identifying cryptic diversity with predictive phylogeography

Anahí Espíndola[1,2], Megan Ruffley[1,2], Megan L. Smith[3], Bryan C. Carstens[3], David C. Tank[1,2] and Jack Sullivan[1,2]

[1] Department of Biological Sciences, University of Idaho, 875 Perimeter Drive MS 3051, Moscow, ID 83844-3051, USA
[2] Biological Sciences, Institute for Bioinformatics and Evolutionary Studies (IBEST), 875 Perimeter Drive MS 3051, Moscow, ID 83844-3051, USA
[3] Department of Evolution, Ecology, and Organismal Biology, The Ohio State University, 318 W. 12th Avenue, 300 Aronoff Labs, Columbus, OH 43210-1293, USA

AE, 0000-0001-9128-8836

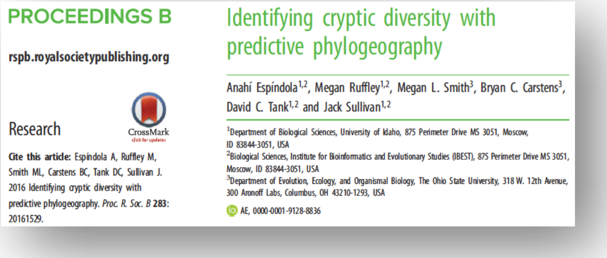
Building a predictive framework

Goal: to develop a predictive framework to identify species that are likely to contain cryptic diversity.

2. Build data table

For each species in focal ecosystem:

1. species distribution model (climate layers)
2. evolutionary model (posterior probability)
3. life history traits (categorical)
4. taxonomic traits (categorical)



PROCEEDINGS B

rspb.royalsocietypublishing.org

Research

CrossMark
click for updates

Cite this article: Espíndola A, Ruffley M, Smith ML, Carstens BC, Tank DC, Sullivan J. 2016 Identifying cryptic diversity with predictive phylogeography. *Proc. R. Soc. B* **283**: 20161529.

Identifying cryptic diversity with predictive phylogeography

Anahí Espíndola[1,2], Megan Ruffley[1,2], Megan L. Smith[3], Bryan C. Carstens[3], David C. Tank[1,2] and Jack Sullivan[1,2]

[1] Department of Biological Sciences, University of Idaho, 875 Perimeter Drive MS 3051, Moscow, ID 83844-3051, USA
[2] Biological Sciences, Institute for Bioinformatics and Evolutionary Studies (IBEST), 875 Perimeter Drive MS 3051, Moscow, ID 83844-3051, USA
[3] Department of Evolution, Ecology, and Organismal Biology, The Ohio State University, 318 W. 12th Avenue, 300 Aronoff Labs, Columbus, OH 43210-1293, USA

iD AE, 0000-0001-9128-8836

Building a predictive framework

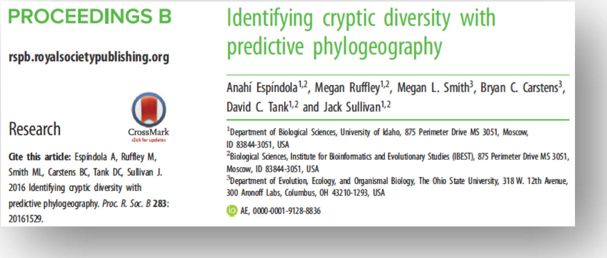
Goal: to develop a predictive framework to identify species that are likely to contain cryptic diversity.

2. Build data table

For each species in focal ecosystem:

1. species distribution model (climate layers)
2. evolutionary model (posterior probability)
3. life history traits (categorical)
4. taxonomic traits (categorical)

How do we analyze these disparate data?



PROCEEDINGS B

rspb.royalsocietypublishing.org

Research

Cite this article: Espíndola A, Ruffley M, Smith ML, Carstens BC, Tank DC, Sullivan J. 2016 Identifying cryptic diversity with predictive phylogeography. *Proc. R. Soc. B* **283**: 20161529.

Identifying cryptic diversity with predictive phylogeography

Anahí Espíndola[1,2], Megan Ruffley[1,2], Megan L. Smith[3], Bryan C. Carstens[3], David C. Tank[1,2] and Jack Sullivan[1,2]

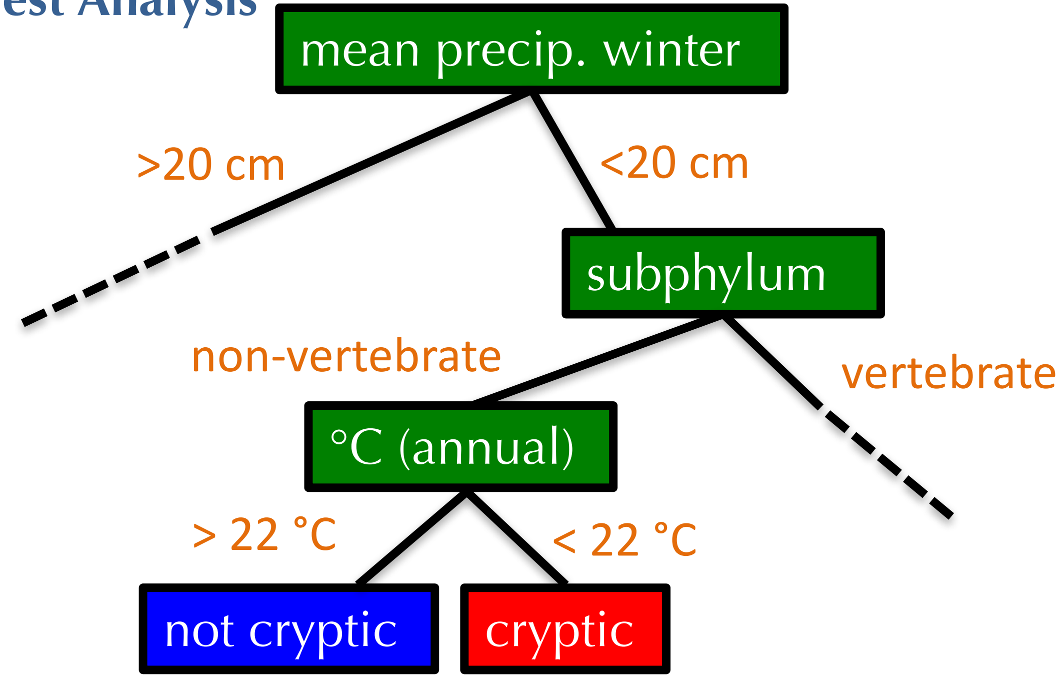
[1] Department of Biological Sciences, University of Idaho, 875 Perimeter Drive MS 3051, Moscow, ID 83844-3051, USA
[2] Biological Sciences, Institute for Bioinformatics and Evolutionary Studies (IBEST), 875 Perimeter Drive MS 3051, Moscow, ID 83844-3051, USA
[3] Department of Evolution, Ecology, and Organismal Biology, The Ohio State University, 318 W. 12th Avenue, 300 Aronoff Labs, Columbus, OH 43210-1293, USA

AE, 0000-0001-9128-8836

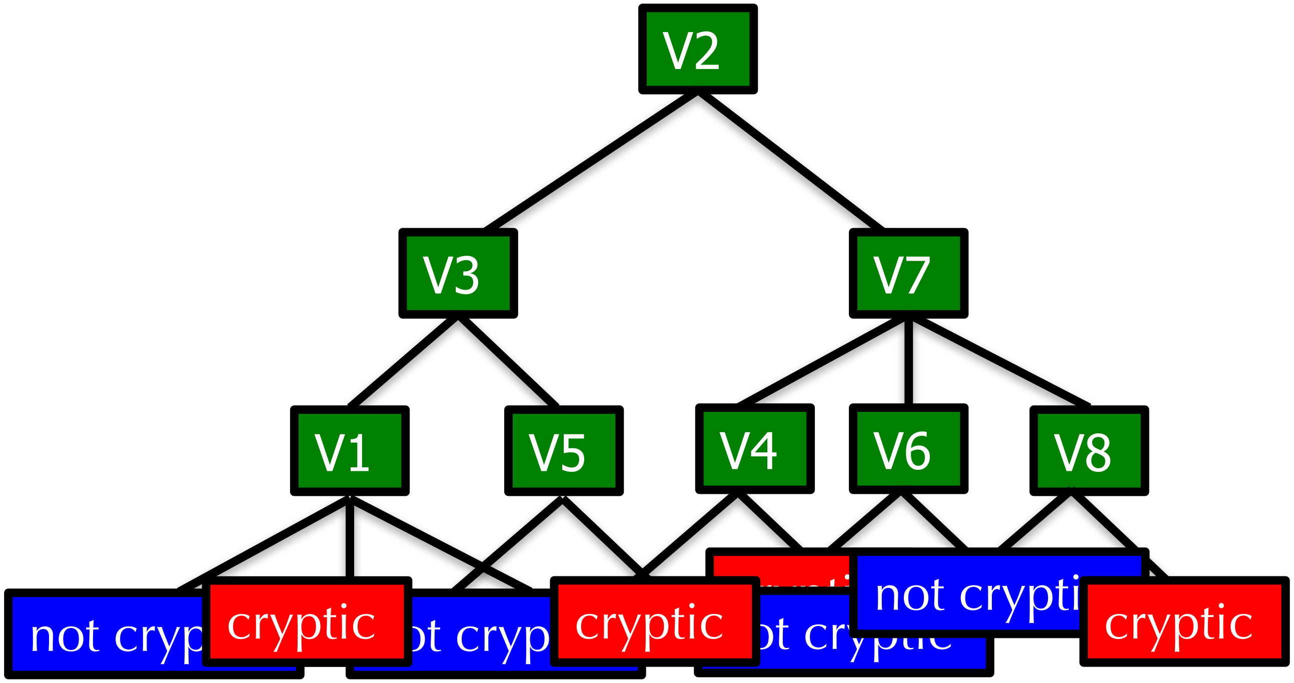
Random Forest Analysis

Goal: to develop a predictive framework to identify species that are likely to contain cryptic diversity.

3. Random Forest Analysis



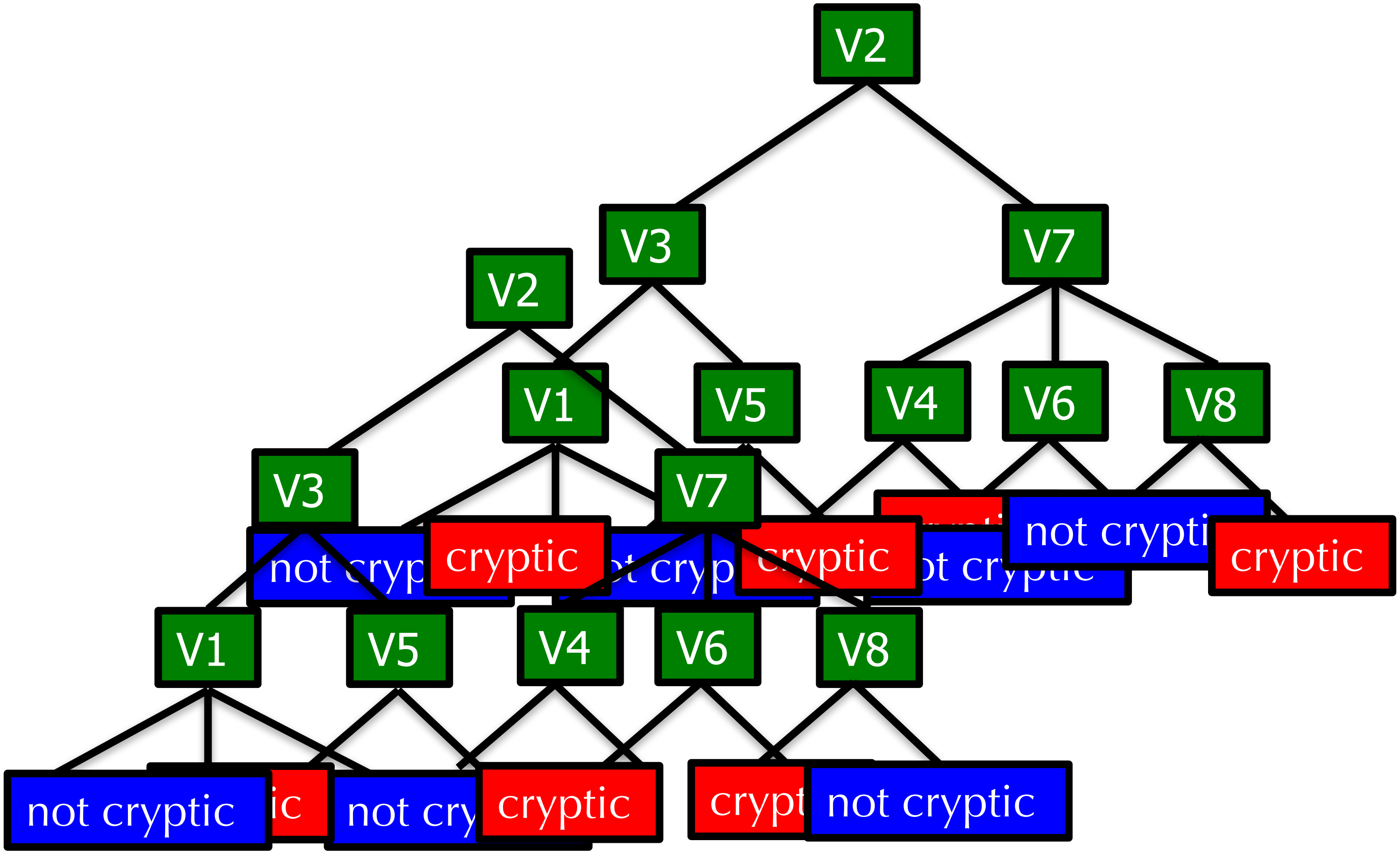
Random Forest Analysis



Build decision tree by choosing predictor variables at random from data table.

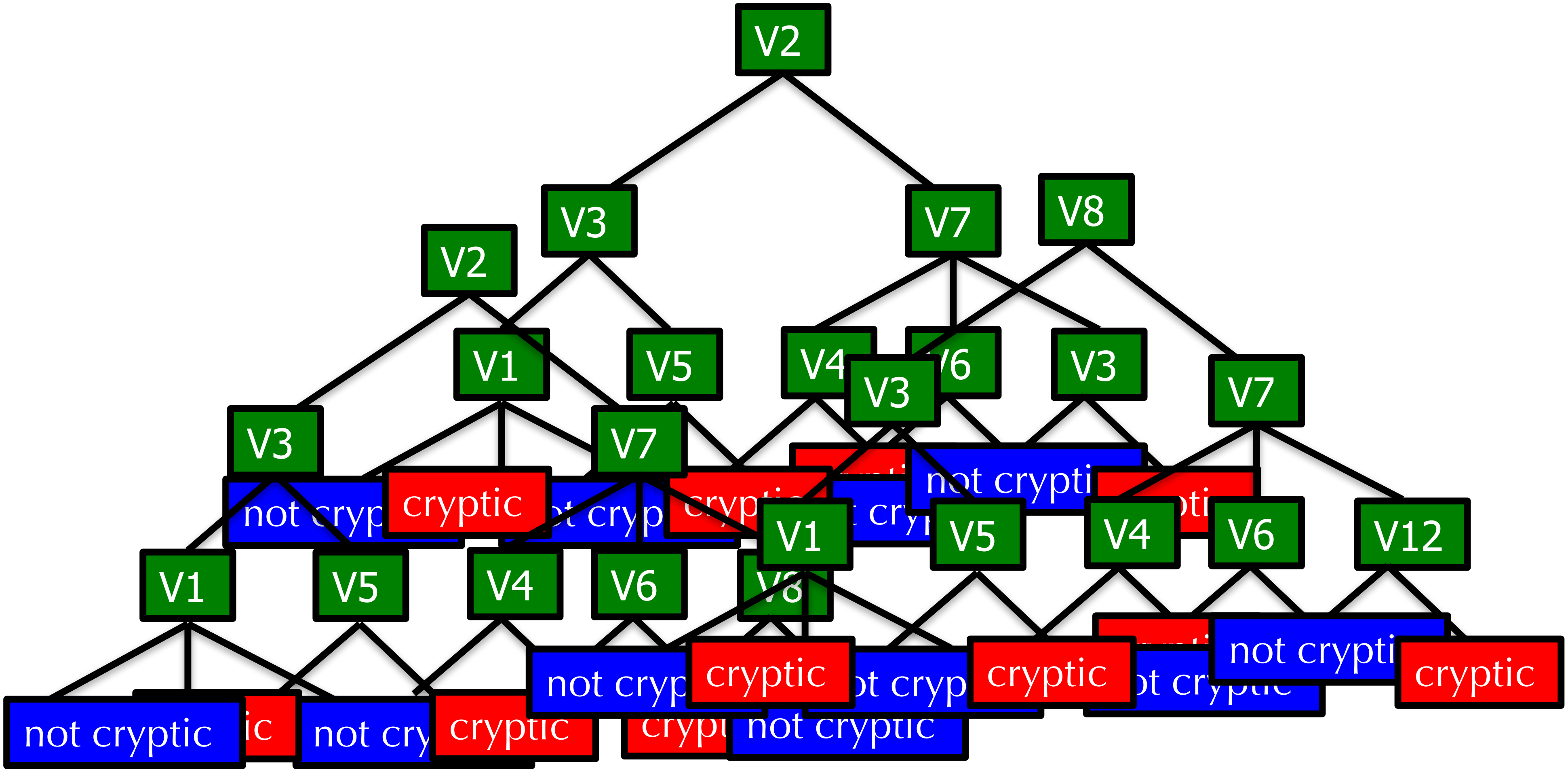
One tree is likely to be a bad predictor of the response variables (cryptic, noncryptic)

Random Forest Analysis

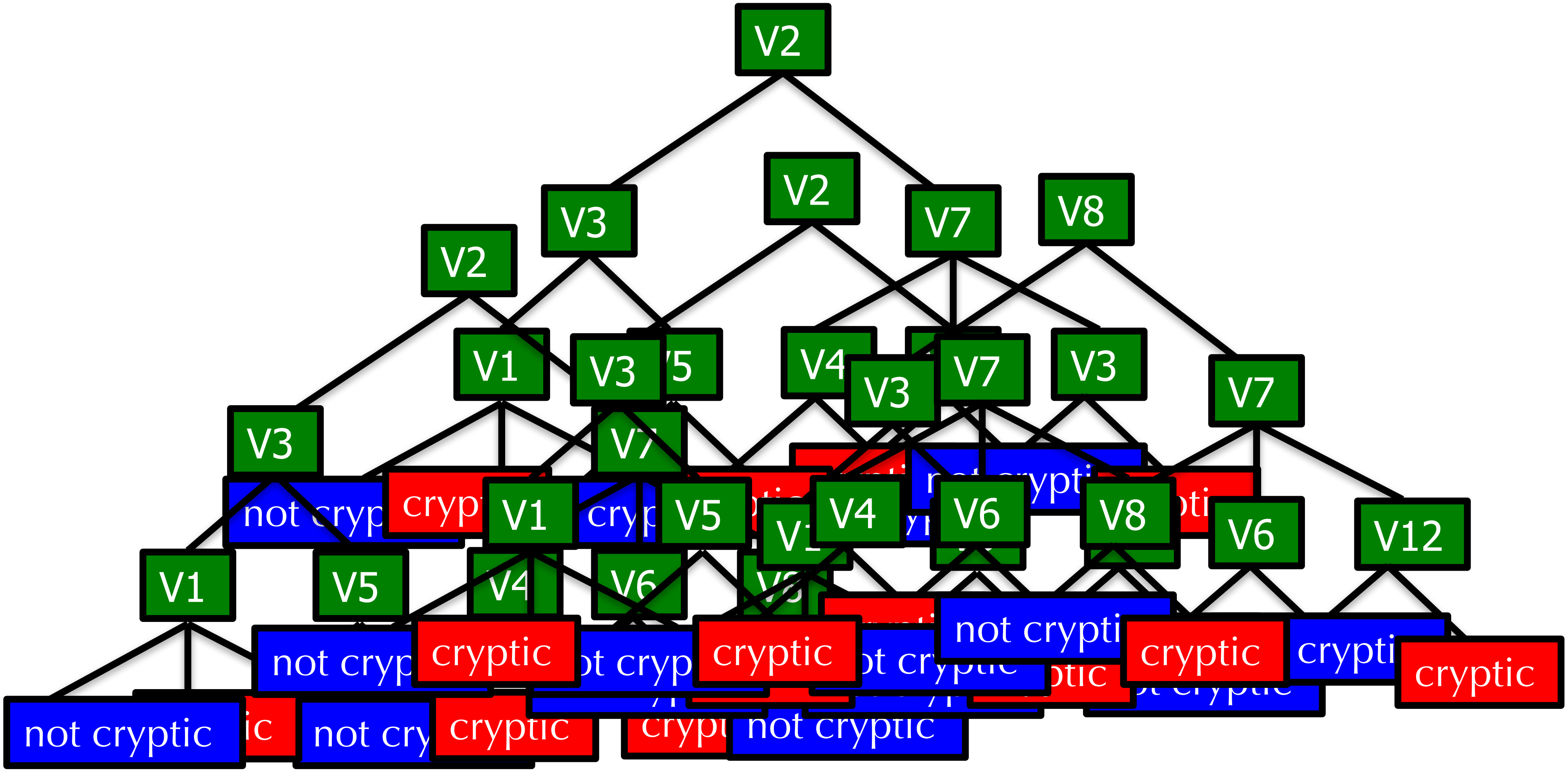


Repeating this process once may produce a better decision tree…

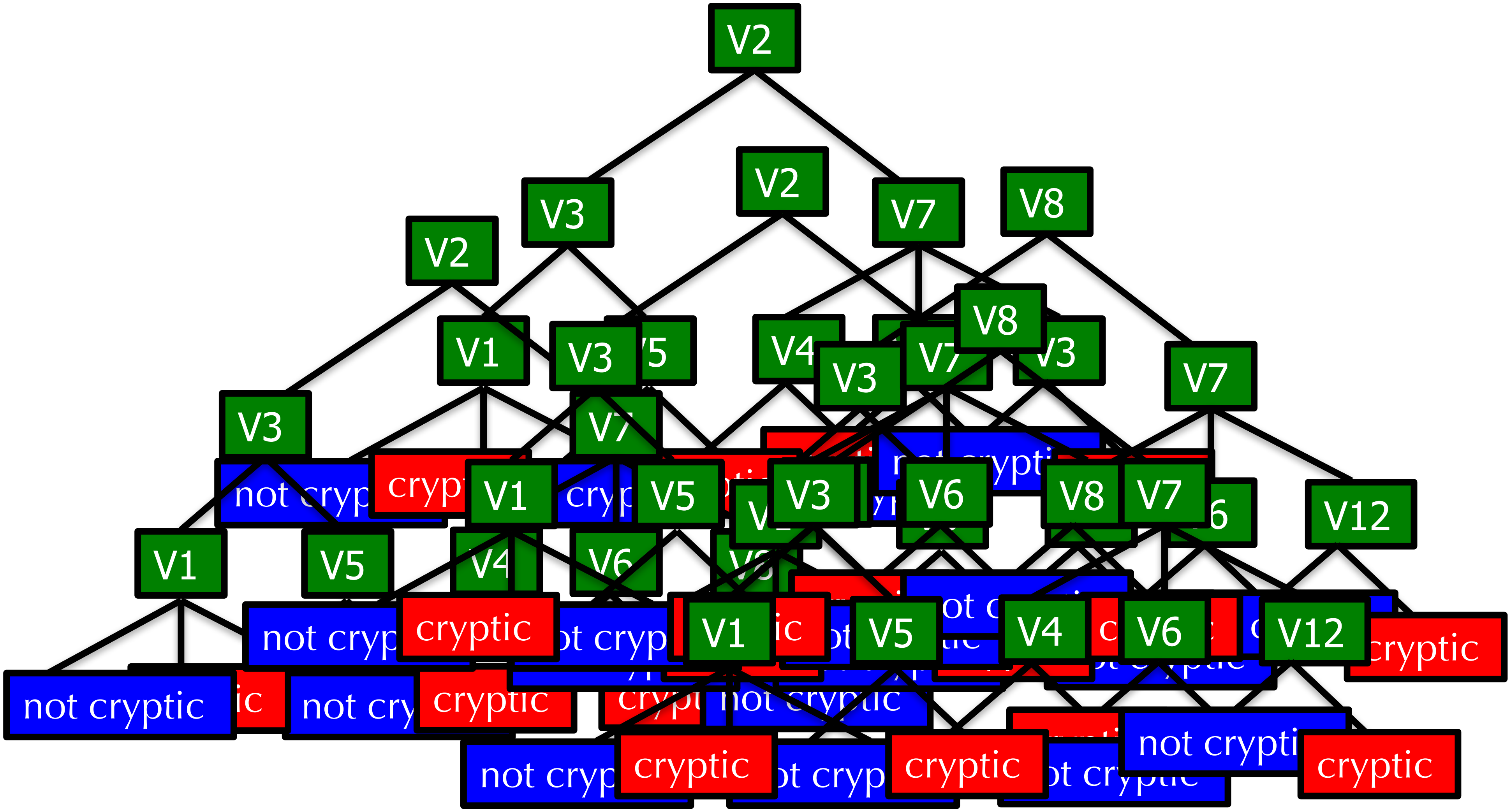
Random Forest Analysis



Random Forest Analysis



Random Forest Analysis



Repeating this process many times and using a consensus tree produces the best decision tree.

Building a predictive framework

Goal: to develop a predictive framework to identify species that are likely to contain cryptic diversity.

4. Evaluation

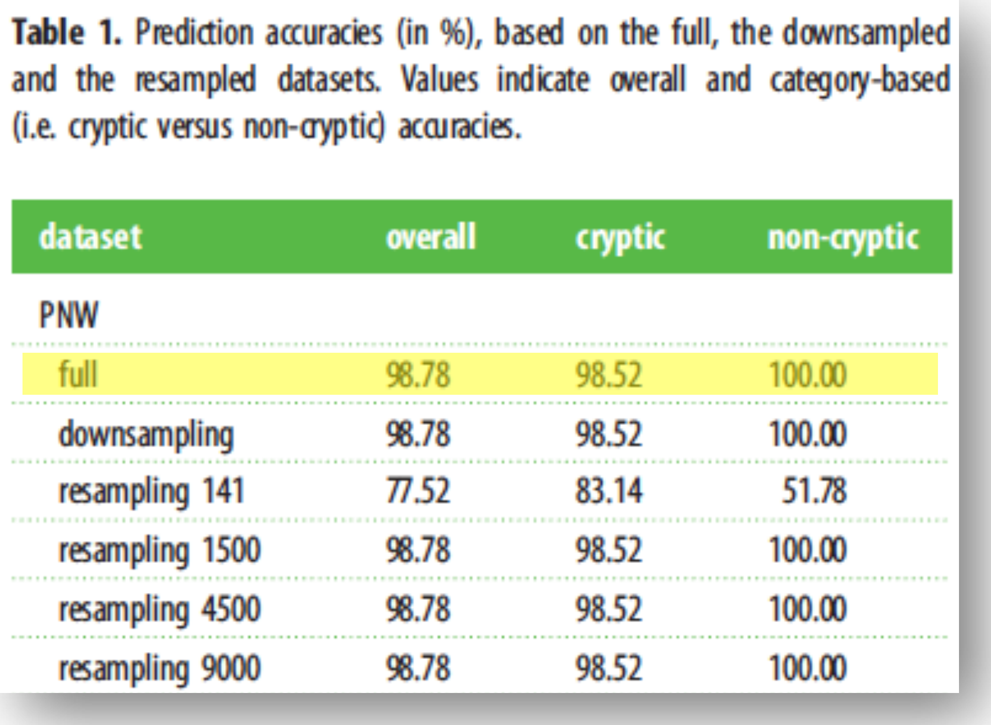
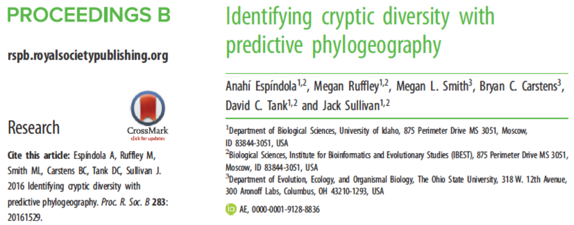


Table 1. Prediction accuracies (in %), based on the full, the downsampled and the resampled datasets. Values indicate overall and category-based (i.e. cryptic versus non-cryptic) accuracies.

dataset	overall	cryptic	non-cryptic
PNW			
full	98.78	98.52	100.00
downsampling	98.78	98.52	100.00
resampling 141	77.52	83.14	51.78
resampling 1500	98.78	98.52	100.00
resampling 4500	98.78	98.52	100.00
resampling 9000	98.78	98.52	100.00



PROCEEDINGS B

rspb.royalsocietypublishing.org

Research

Cite this article: Espíndola A, Ruffley M, Smith ML, Carstens BC, Tank DC, Sullivan J. 2016 Identifying cryptic diversity with predictive phylogeography. *Proc. R. Soc. B* **283**: 20161529.

Identifying cryptic diversity with predictive phylogeography

Anahí Espíndola[1,2], Megan Ruffley[1,2], Megan L. Smith[3], Bryan C. Carstens[3], David C. Tank[1,2] and Jack Sullivan[1,2]

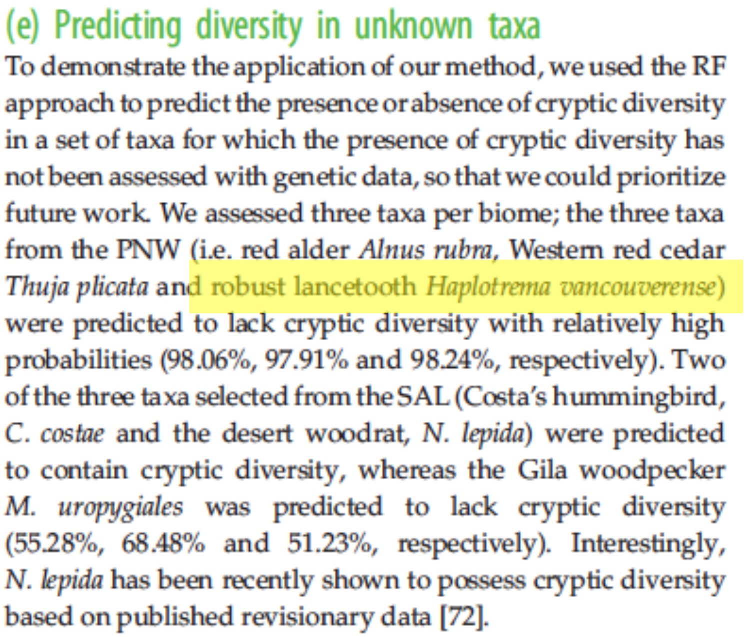
[1] Department of Biological Sciences, University of Idaho, 875 Perimeter Drive MS 3051, Moscow, ID 83844-3051, USA
[2] Biological Sciences, Institute for Bioinformatics and Evolutionary Studies (IBEST), 875 Perimeter Drive MS 3051, Moscow, ID 83844-3051, USA
[3] Department of Evolution, Ecology, and Organismal Biology, The Ohio State University, 318 W. 12th Avenue, 300 Aronoff Labs, Columbus, OH 43210-1293, USA

AE, 0000-0001-9128-8836

Building a predictive framework

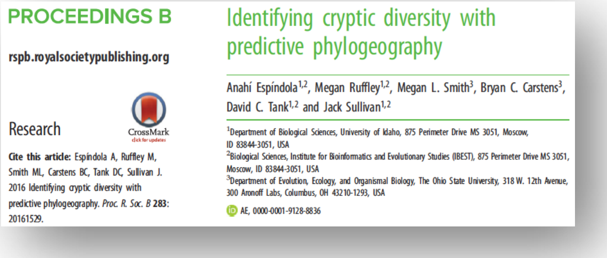
Goal: to develop a predictive framework to identify species that are likely to contain cryptic diversity.

5. Prediction



(e) Predicting diversity in unknown taxa

To demonstrate the application of our method, we used the RF approach to predict the presence or absence of cryptic diversity in a set of taxa for which the presence of cryptic diversity has not been assessed with genetic data, so that we could prioritize future work. We assessed three taxa per biome; the three taxa from the PNW (i.e. red alder *Alnus rubra*, Western red cedar *Thuja plicata* and robust lancetooth *Haplotrema vancouverense*) were predicted to lack cryptic diversity with relatively high probabilities (98.06%, 97.91% and 98.24%, respectively). Two of the three taxa selected from the SAL (Costa's hummingbird, *C. costae* and the desert woodrat, *N. lepida*) were predicted to contain cryptic diversity, whereas the Gila woodpecker *M. uropygiales* was predicted to lack cryptic diversity (55.28%, 68.48% and 51.23%, respectively). Interestingly, *N. lepida* has been recently shown to possess cryptic diversity based on published revisionary data [72].



PROCEEDINGS B

rspb.royalsocietypublishing.org

Research

CrossMark
click for updates

Cite this article: Espíndola A, Ruffley M, Smith ML, Carstens BC, Tank DC, Sullivan J. 2016 Identifying cryptic diversity with predictive phylogeography. *Proc. R. Soc. B* **283**: 20161529.

Identifying cryptic diversity with predictive phylogeography

Anahí Espíndola[1,2], Megan Ruffley[1,2], Megan L. Smith[3], Bryan C. Carstens[3], David C. Tank[1,2] and Jack Sullivan[1,2]

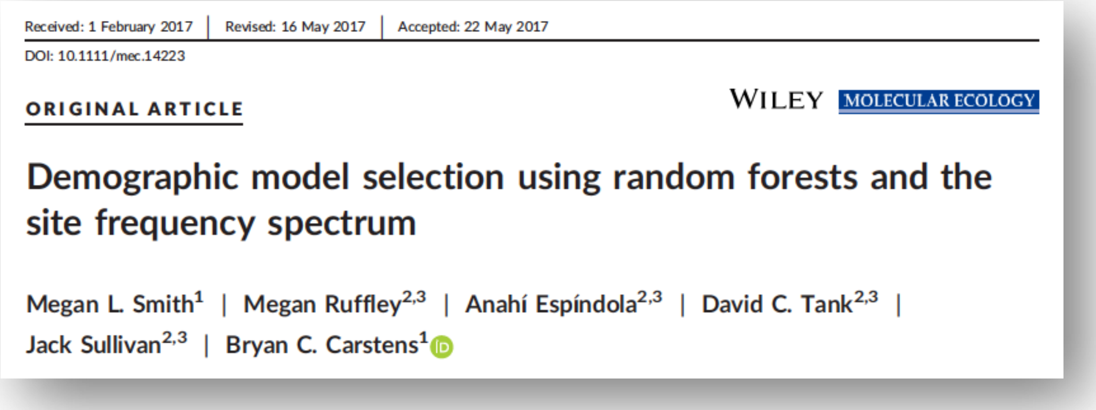
[1]Department of Biological Sciences, University of Idaho, 875 Perimeter Drive MS 3051, Moscow, ID 83844-3051, USA
[2]Biological Sciences, Institute for Bioinformatics and Evolutionary Studies (IBEST), 875 Perimeter Drive MS 3051, Moscow, ID 83844-3051, USA
[3]Department of Evolution, Ecology, and Organismal Biology, The Ohio State University, 318 W. 12th Avenue, 300 Aronoff Labs, Columbus, OH 43210-1293, USA

iD AE, 0000-0001-9128-8836

Building a predictive framework

Haplotrema vancouverense was predicted to be non-cryptic absent genetic data…

Megan Smith used SNPs from RADseq, developed a novel approach to analyzing these data, and confirmed that the snail was non-cryptic.



Received: 1 February 2017 | Revised: 16 May 2017 | Accepted: 22 May 2017

DOI: 10.1111/mec.14223

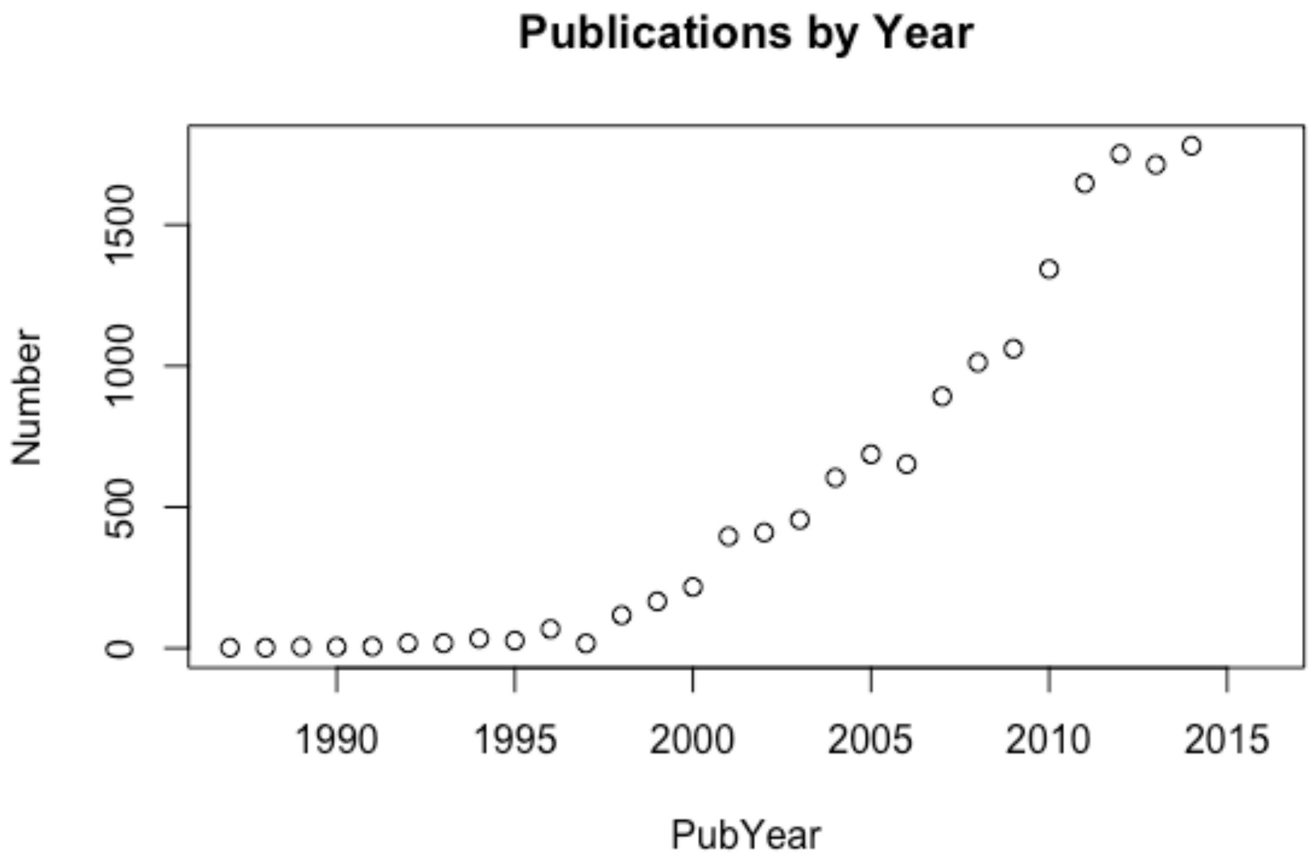
ORIGINAL ARTICLE

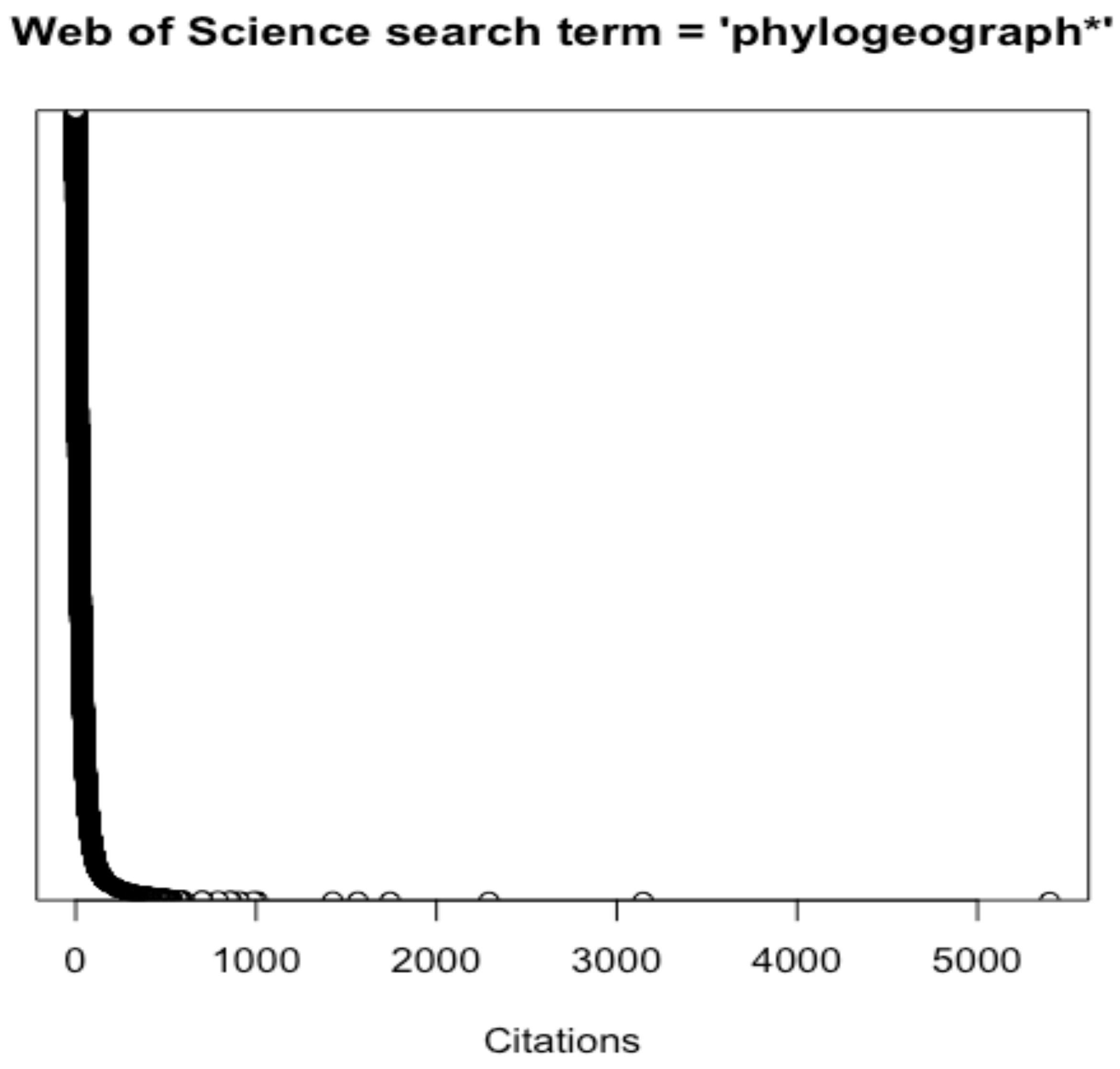
WILEY MOLECULAR ECOLOGY

Demographic model selection using random forests and the site frequency spectrum

Megan L. Smith[1] | Megan Ruffley[2,3] | Anahí Espíndola[2,3] | David C. Tank[2,3] | Jack Sullivan[2,3] | Bryan C. Carstens[1]

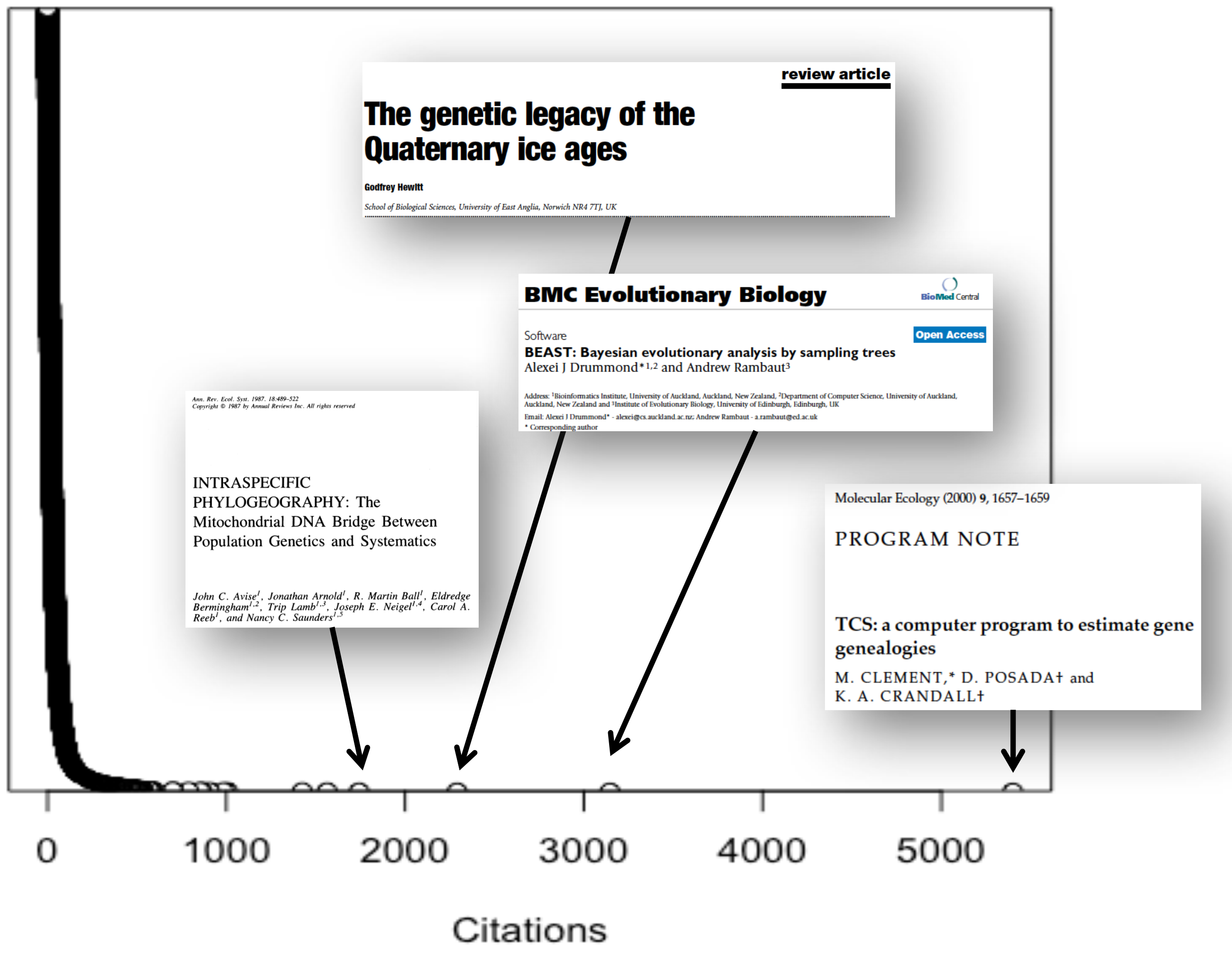
Predictive frameworks **recycle** and **repurpose** existing data.




Web of Science search term = 'phylogeograph*'
0
1000
2000
3000
4000
5000
Citations

Program notes & review papers…

Web of Science search term = 'phylogeograph*'



review article

The genetic legacy of the Quaternary ice ages

Godfrey Hewitt

School of Biological Sciences, University of East Anglia, Norwich NR4 7TJ, UK

BMC Evolutionary Biology

BioMed Central

Software

Open Access

BEAST: Bayesian evolutionary analysis by sampling trees

Alexei J Drummond*[1,2] and Andrew Rambaut[3]

Address: [1]Bioinformatics Institute, University of Auckland, Auckland, New Zealand, [2]Department of Computer Science, University of Auckland, Auckland, New Zealand and [3]Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, UK

Email: Alexei J Drummond* - alexei@cs.auckland.ac.nz; Andrew Rambaut - a.rambaut@ed.ac.uk

* Corresponding author

Ann. Rev. Ecol. Syst. 1987. 18:489–522
Copyright © 1987 by Annual Reviews Inc. All rights reserved

INTRASPECIFIC PHYLOGEOGRAPHY: The Mitochondrial DNA Bridge Between Population Genetics and Systematics

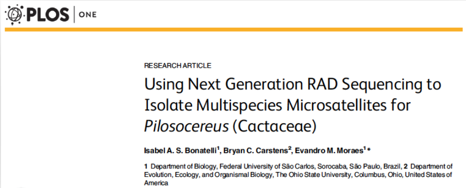
John C. Avise[1], Jonathan Arnold[1], R. Martin Ball[1], Eldredge Bermingham[1,2], Trip Lamb[1,3], Joseph E. Neigel[1,4], Carol A. Reeb[1], and Nancy C. Saunders[1,5]

Molecular Ecology (2000) **9**, 1657–1659

PROGRAM NOTE

TCS: a computer program to estimate gene genealogies

M. CLEMENT,* D. POSADA† and K. A. CRANDALL†



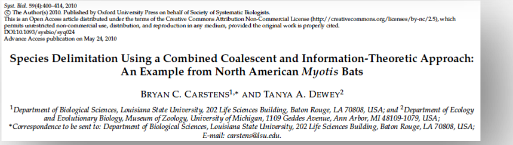
PLOS | ONE

RESEARCH ARTICLE

Using Next Generation RAD Sequencing to Isolate Multispecies Microsatellites for *Pilosocereus* (Cactaceae)

Isabel A. S. Bonatelli[1], Bryan C. Carstens[2], Evandro M. Moraes[1]*

1 Department of Biology, Federal University of São Carlos, Sorocaba, São Paulo, Brazil, **2** Department of Evolution, Ecology, and Organismal Biology, The Ohio State University, Columbus, Ohio, United States of America

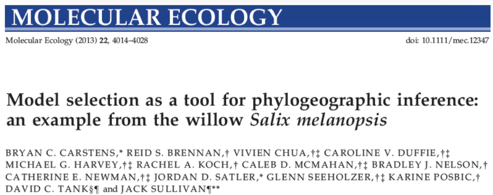


Syst. Biol. 59(4):400–414, 2010
© The Author(s) 2010. Published by Oxford University Press on behalf of Society of Systematic Biologists.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/2.5), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
DOI:10.1093/sysbio/syq024
Advance Access publication on May 24, 2010

Species Delimitation Using a Combined Coalescent and Information-Theoretic Approach: An Example from North American *Myotis* Bats

BRYAN C. CARSTENS[1,*] AND TANYA A. DEWEY[2]

[1] *Department of Biological Sciences, Louisiana State University, 202 Life Sciences Building, Baton Rouge, LA 70808, USA; and* [2] *Department of Ecology and Evolutionary Biology, Museum of Zoology, University of Michigan, 1109 Geddes Avenue, Ann Arbor, MI 48109-1079, USA;*
**Correspondence to be sent to: Department of Biological Sciences, Louisiana State University, 202 Life Sciences Building, Baton Rouge, LA 70808, USA; E-mail: carstens@lsu.edu.*

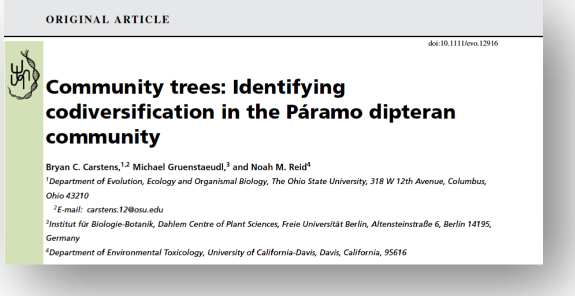


MOLECULAR ECOLOGY

Molecular Ecology (2013) **22**, 4014–4028 doi: 10.1111/mec.12347

Model selection as a tool for phylogeographic inference: an example from the willow *Salix melanopsis*

BRYAN C. CARSTENS,* REID S. BRENNAN,† VIVIEN CHUA,†‡ CAROLINE V. DUFFIE,†‡ MICHAEL G. HARVEY,†‡ RACHEL A. KOCH,† CALEB D. MCMAHAN,†‡ BRADLEY J. NELSON,† CATHERINE E. NEWMAN,†‡ JORDAN D. SATLER,* GLENN SEEHOLZER,†‡ KARINE POSBIC,† DAVID C. TANK§¶ and JACK SULLIVAN¶**



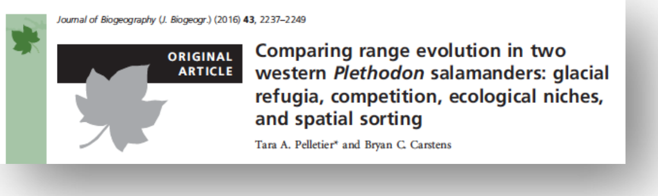
ORIGINAL ARTICLE

doi:10.1111/evo.12916

Community trees: Identifying codiversification in the Páramo dipteran community

Bryan C. Carstens,[1,2] Michael Gruenstaeudl,[3] and Noah M. Reid[4]

[1] *Department of Evolution, Ecology and Organismal Biology, The Ohio State University, 318 W 12th Avenue, Columbus, Ohio 43210*
[2] *E-mail: carstens.12@osu.edu*
[3] *Institut für Biologie-Botanik, Dahlem Centre of Plant Sciences, Freie Universität Berlin, Altensteinstraße 6, Berlin 14195, Germany*
[4] *Department of Environmental Toxicology, University of California-Davis, Davis, California, 95616*



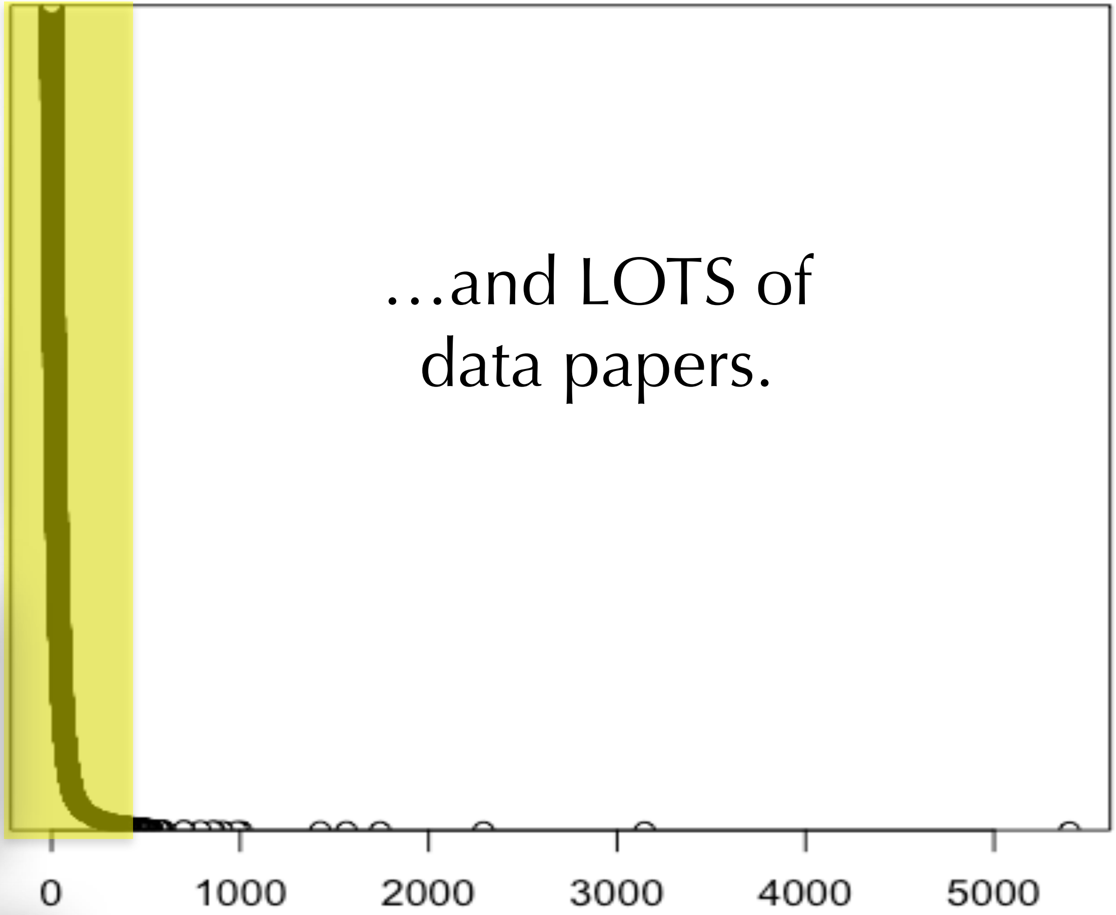
Journal of Biogeography (*J. Biogeogr.*) (2016) **43**, 2237–2249

ORIGINAL ARTICLE

Comparing range evolution in two western *Plethodon* salamanders: glacial refugia, competition, ecological niches, and spatial sorting

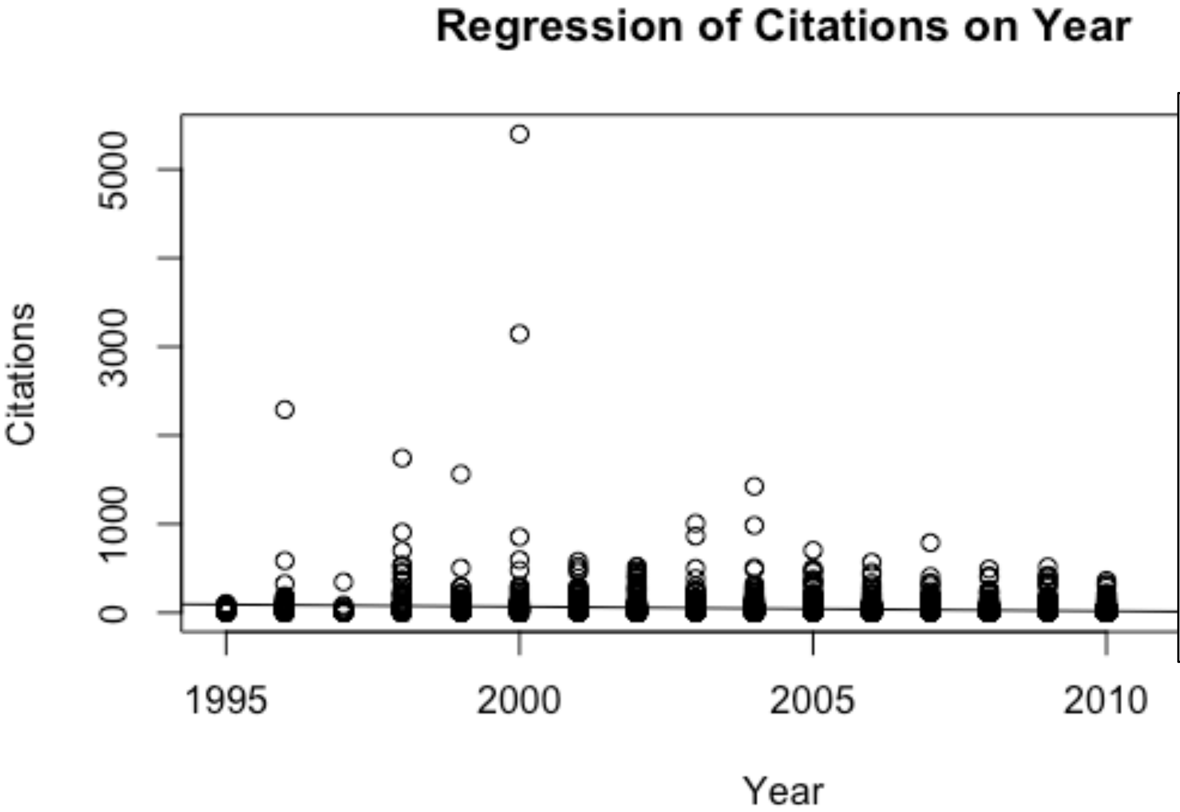
Tara A. Pelletier* and Bryan C. Carstens

Web of Science search term = 'phylogeograph*'



…and LOTS of data papers.

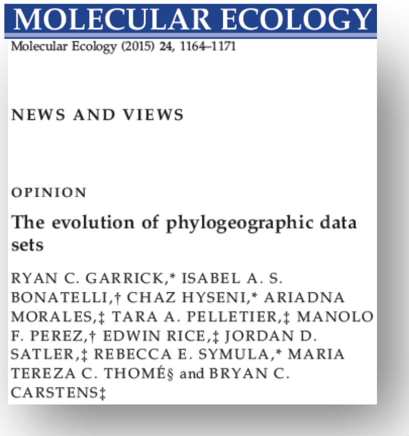
Citations



r^2 ~0.99

mode
= 3

median
= 8



MOLECULAR ECOLOGY

Molecular Ecology (2015) 24, 1164–1171

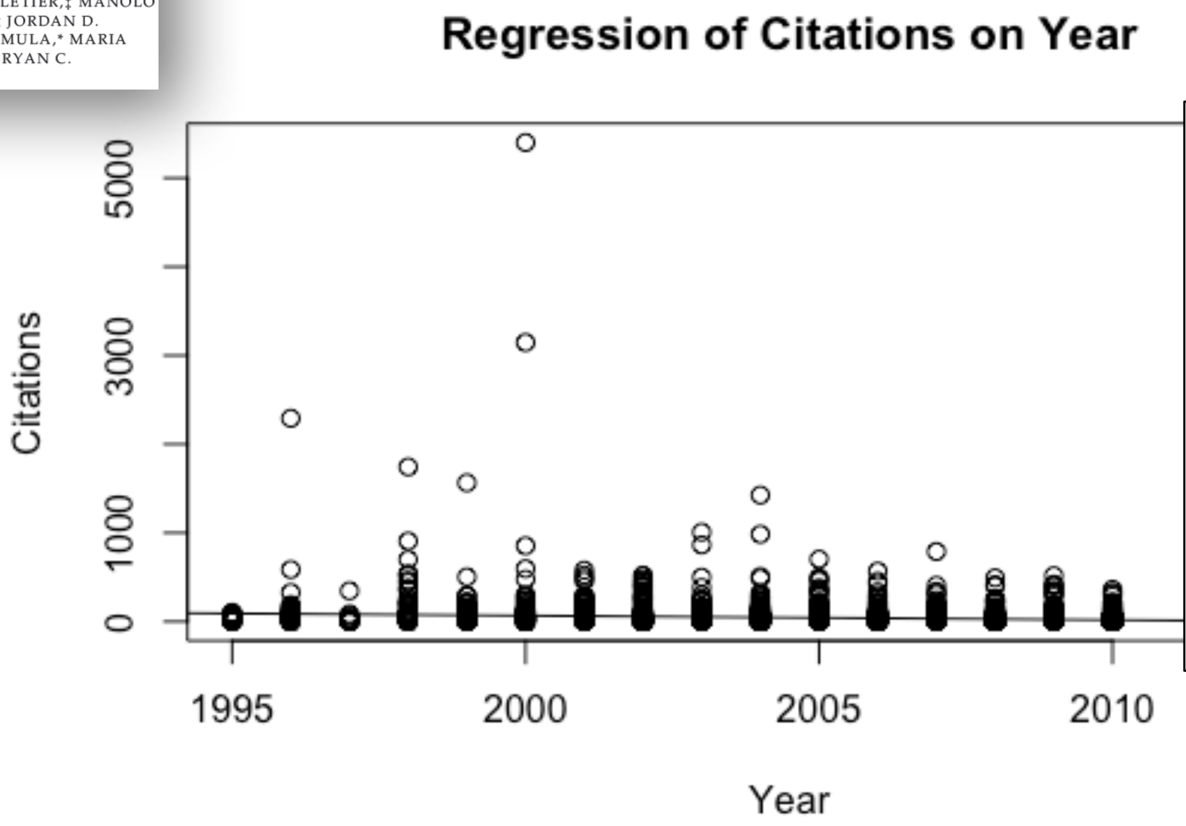
NEWS AND VIEWS

OPINION

The evolution of phylogeographic data sets

RYAN C. GARRICK,* ISABEL A. S. BONATELLI,† CHAZ HYSENI,* ARIADNA MORALES,‡ TARA A. PELLETIER,‡ MANOLO F. PEREZ,† EDWIN RICE,‡ JORDAN D. SATLER,‡ REBECCA E. SYMULA,* MARIA TEREZA C. THOMÉ§ and BRYAN C. CARSTENS‡

- ~200 alleles per investigation (Garrick et al. 2015)
- ~200*30,000 data points



r^2 ~0.99

mode = 3

median = 8

Ecologists and Evolutionary Biologists…

- …want to learn about interesting ecosystems and species

- … hope to understand how biodiversity evolves

…should use big data!

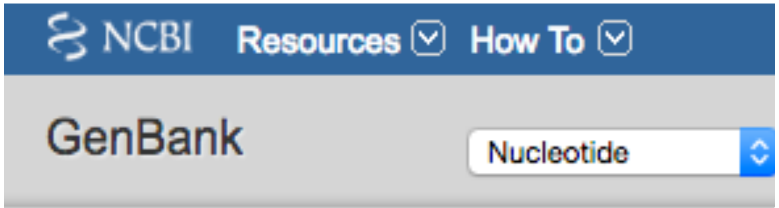
1. **aggregate** available data on a global scale

2. **analyze** these data using predictive modeling

WorldClim - Global Climate Data

Free climate data for ecological modeling and GIS

19 bioclimatic variables at ~1km global resolution

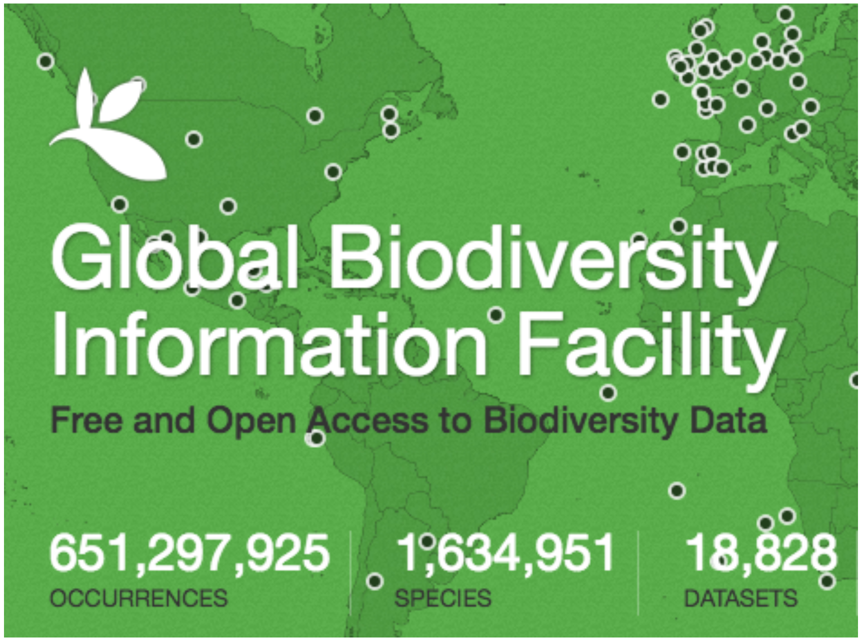


NCBI Resources How To

GenBank

Nucleotide

>200,000,000 sequences



Global Biodiversity Information Facility

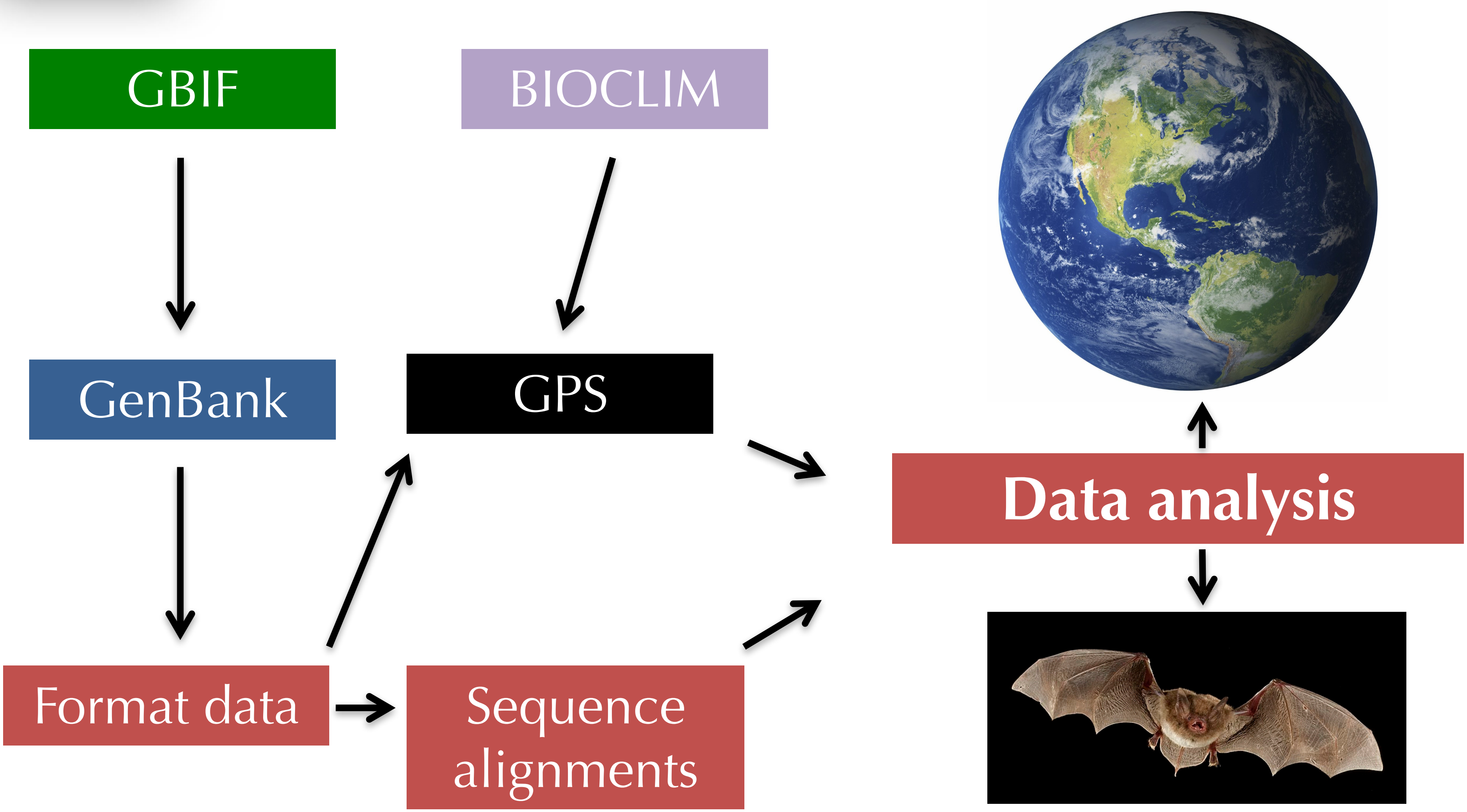
Free and Open Access to Biodiversity Data

651,297,925	1,634,951	18,828
OCCURRENCES	SPECIES	DATASETS

Greg Wheeler helped develop the initial versions of the scripts to aggregate available data.

Phylogatr

Tara Pelletier developed the complete set of Python & R scripts to aggregate available data.



GBIF

BIOCLIM

GenBank

GPS

Data analysis

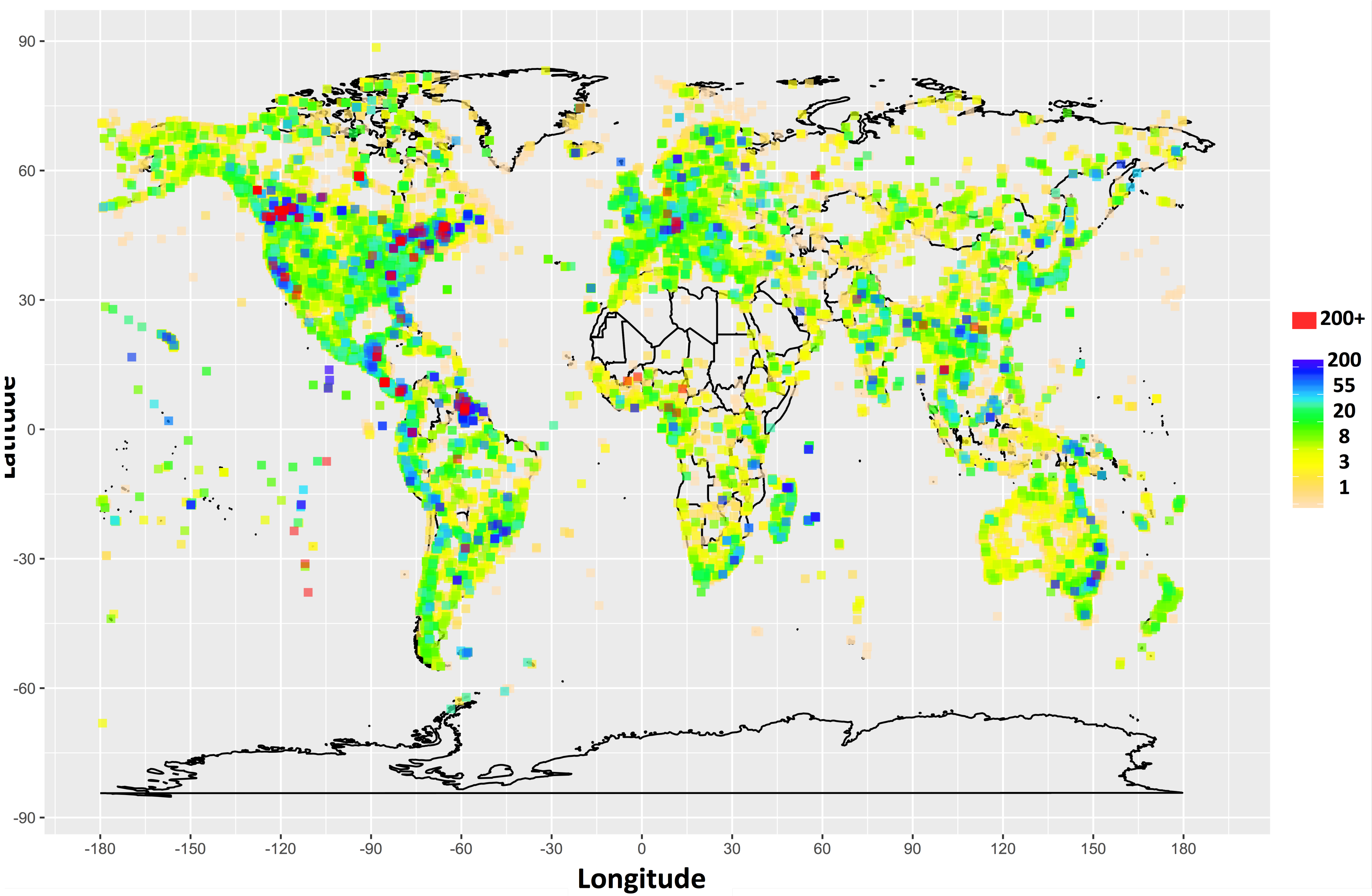
Format data

Sequence alignments

Big Data!

Phylogatr (Tara Pelletier)

- 561,534 –georeferenced sequences
- 42,206 – species w/ georeferenced sequence data
- 12,266 – sequence alignments
- 10,991– species with alignments

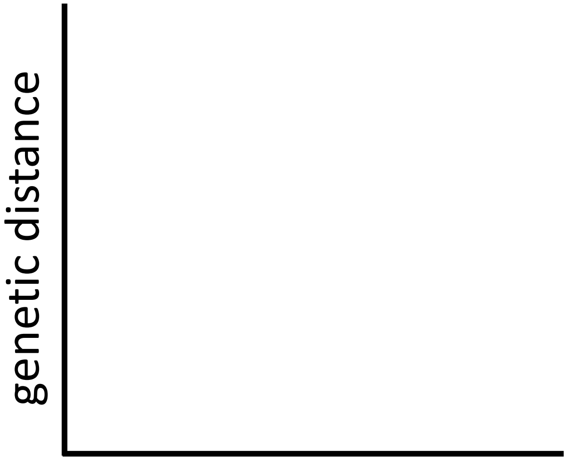
90
60
30
0
-30
-60
-90
Latitude
-180
-150
-120
-90
-60
-30
0
30
60
90
120
150
180
Longitude
🟥 200+
200
55
20
8
3
1

Big Data!

Phylogatr (w/ Tara Pelletier)

- **Global processes** (structure of genetic diversity)
- **Classic questions on global scales** (response to climate change)
- **Quantifying biodiversity** (species limits in major clades)

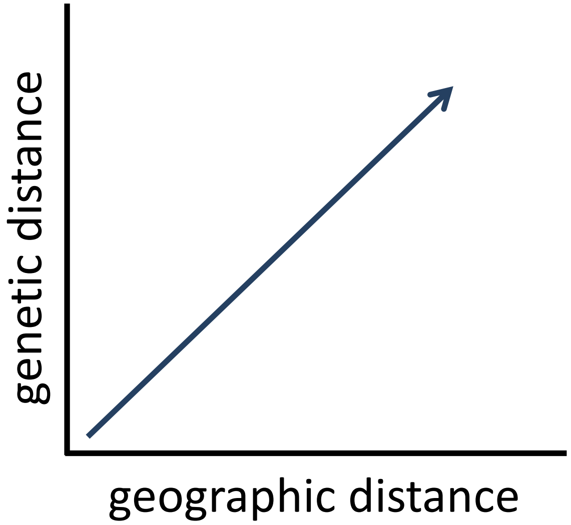
What factors promote intraspecific genetic structure?



What factors promote intraspecific genetic structure?

Wright 1943

- **Isolation by distance (IBD)**

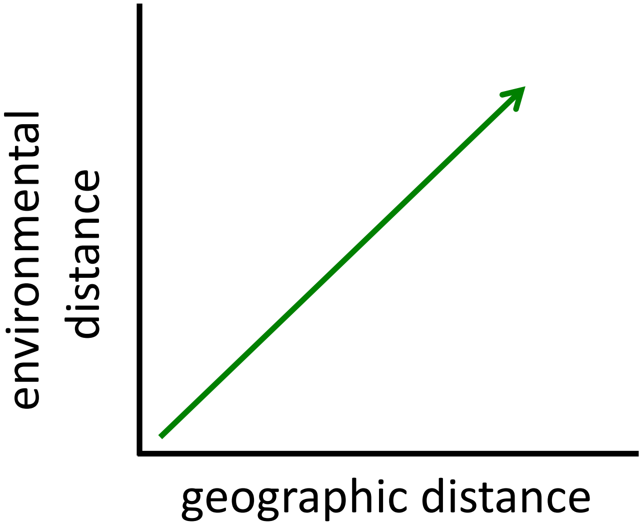


What factors promote intraspecific genetic structure?

Wright 1943

- **Isolation by distance (IBD)**

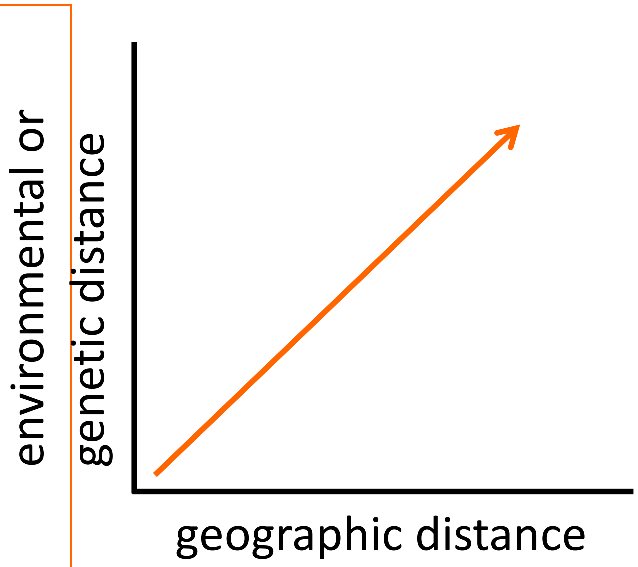
- **Isolation by environment (IBE)**



What factors promote intraspecific genetic structure?

Wright 1943

- **Isolation by distance (IBD)**
- **Isolation by environment (IBE)**
- correlation within species between environment and geography on a global scale (*r* = 0.77)
- **IBD/E**: multiple matrix regression with randomization (Wang 2013)



What factors promote intraspecific genetic structure?

Group	# datasets	prop.sig Geo.	P-value Geo.	prop.sig Env.	P-value Env.
Fungi	23	0.04			
Mosses					
Ferns					
Gymnosperms					
Angiosperms					
Arthropods					
Vertebrates					
Annelida					
Cnidaria					
Echinodermata					
Mollusca					
Nematoda					
Platyhelminthes					
Total					

each species tested at *P* = 0.05

What factors promote intraspecific genetic structure?

Group	# datasets	prop.sig Geo.	**P-value Geo.**	prop.sig Env.	P-value Env.
Fungi	23	0.04	**0.69**		
Mosses					
Ferns					
Gymnosperms					
Angiosperms					
Arthropods					
Vertebrates					
Annelida					
Cnidaria					
Echinodermata					
Mollusca					
Nematoda					
Platyhelminthes					
Total					

Exact Binomial test: Is the proportion of species that are isolated by distance higher than expected by chance?

What factors promote intraspecific genetic structure?

Group	# datasets	prop.sig Geo.	P-value Geo.	**prop.sig Env.**	**P-value Env.**
Fungi	23	0.04	0.69	**0.04**	**0.69**
Mosses					
Ferns					
Gymnosperms					
Angiosperms					
Arthropods					
Vertebrates					
Annelida					
Cnidaria					
Echinodermata					
Mollusca					
Nematoda					
Platyhelminthes					
Total					

Because we're using a **multiple matrix regression with randomization** both IBD and IBE are considered.

What factors promote intraspecific genetic structure?

Group	# datasets	prop.sig Geo.	P-value Geo.	prop.sig Env.	P-value Env.
Fungi	23				
Mosses	10				
Ferns	7				
Gymnosperms	111				
Angiosperms	870				
Arthropods	6015				
Vertebrates	2723				
Annelida	33				
Cnidaria	6				
Echinodermata	14				
Mollusca	44				
Nematoda	6				
Platyhelminthes	15				
Total	**9877**				

What factors promote intraspecific genetic structure?

Group	# datasets	prop.sig Geo.	P-value Geo.	prop.sig Env.	P-value Env.
Fungi	23	0.04		0.04	
Mosses	10	0		0	
Ferns	7	0		0	
Gymnosperms	111	0.07		0.06	
Angiosperms	870	0.1		0.1	
Arthropods	6015	0.15		0.13	
Vertebrates	2723	0.29		0.21	
Annelida	33	0.21		0.15	
Cnidaria	6	0.5		0	
Echinodermata	14	0.21		0.21	
Mollusca	44	0.16		0.16	
Nematoda	6	0.33		0.33	
Platyhelminthes	15	0		0.2	
Total	**9877**	**0.19**		**0.15**	

What factors promote intraspecific genetic structure?

Group	# datasets	prop.sig Geo.	P-value Geo.	prop.sig Env.	P-value Env.
Fungi	23	0.04	0.69	0.04	0.69
Mosses	10	0	1	0	1
Ferns	7	0	1	0	1
Gymnosperms	111	0.07	0.19	0.06	0.32
Angiosperms	870	0.1	<0.01	0.1	<0.01
Arthropods	6015	0.15	<0.01	0.13	<0.01
Vertebrates	2723	0.29	<0.01	0.21	<0.01
Annelida	33	0.21	<0.01	0.15	0.02
Cnidaria	6	0.5	<0.01	0	1
Echinodermata	14	0.21	0.03	0.21	0.03
Mollusca	44	0.16	0.01	0.16	0.01
Nematoda	6	0.33	0.03	0.33	0.03
Platyhelminthes	15	0	1	0.2	0.0362
Total	**9877**	**0.19**	**<0.01**	**0.15**	**<0.01**

What explains this variation in IBD/E across biological groups?

Building predictive frameworks for big data analysis

Isolation by distance / environment analyses grouped categorically by result:

- not significant / significant (at the species level)

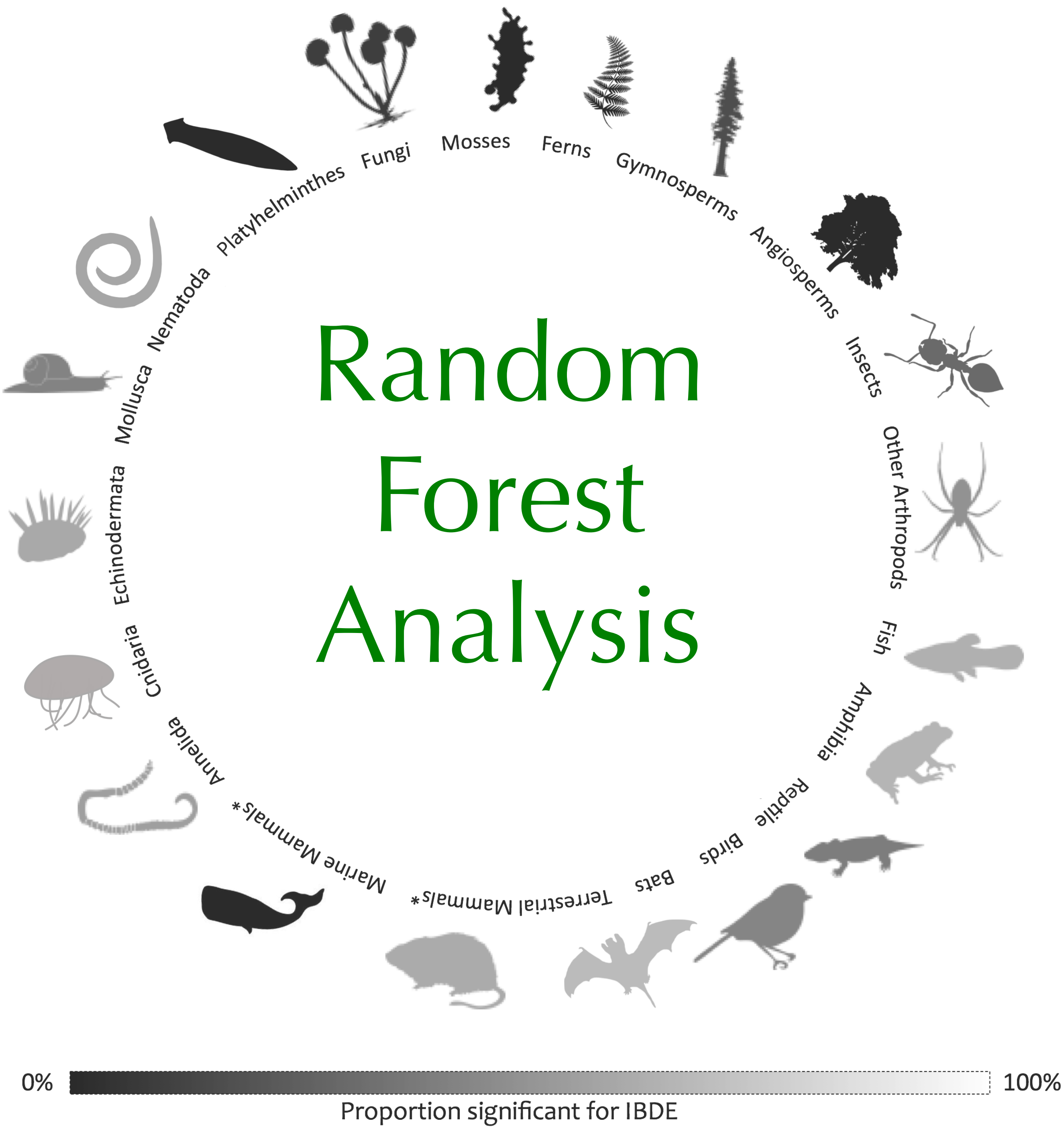
- data table containing general information about organisms and environment

Building predictive frameworks for big data analysis

Data table: 33 variables used in machine learning analysis

- environmental characteristics (canopy cover, wetlands, habitat type)
- organismal traits (metabolism, taxonomy, type of gene)
- geographic (max. latitude, range area, mid-point latitude)

Random Forest Analysis



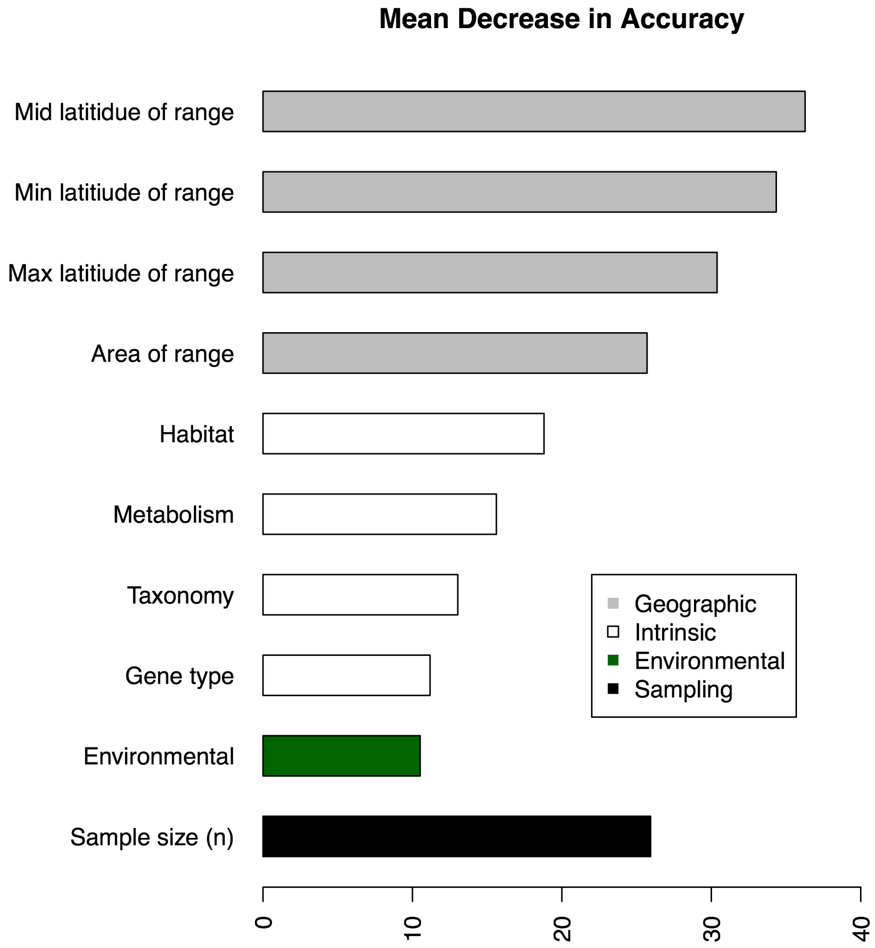
3/27/18

What factors promote intraspecific genetic structure?

Variable performance quantified by measuring the mean decrease in accuracy (MDA) of the predictive function that occurs when that variable is omitted from analysis.

What factors promote intraspecific genetic structure?

Variable performance quantified by measuring the mean decrease in accuracy (MDA) of the predictive function that occurs when that variable is omitted from analysis.



What factors promote intraspecific genetic structure?

Variable	mean with IBD	mean without IBD	t-test p-value
Area (km^{2})	6.57x10^{6}	2.98 x10^{6}	3.05x10^{-11}
Minimum distance from equator	30.04982	31.88308	2.92E-06
Mid-point latitude of range	31.86497	33.42718	0.0009313
Length of latitude°	14.66759	9.74694	2.20E-16

Geography!

What factors promote intraspecific genetic structure?

- If IBD/E is a precursor to local adaptation, organismal traits represent evolved responses to aspects of the environment

- more precise organismal traits are needed…

What factors promote intraspecific genetic structure?

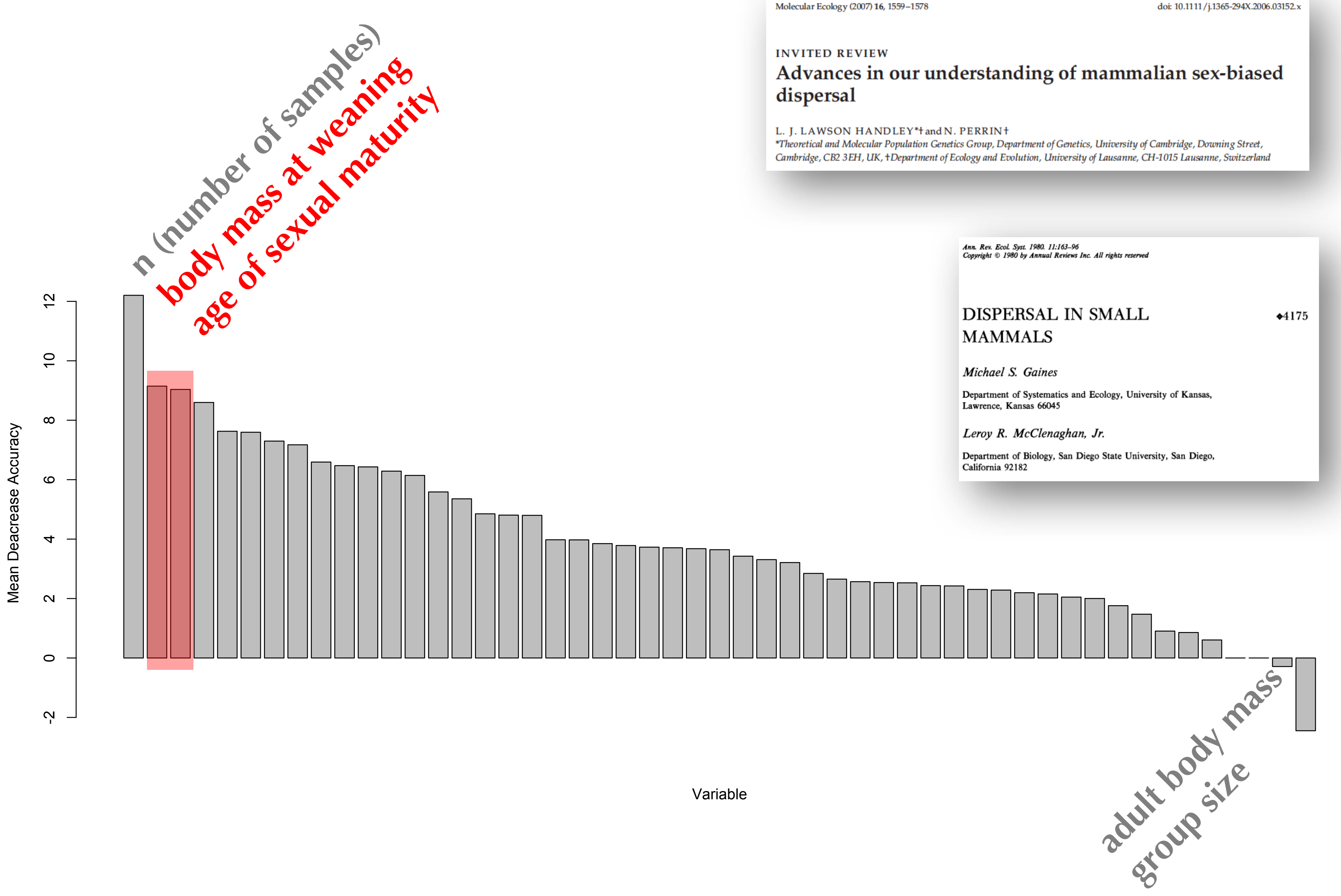
- If IBD/E is a precursor to local adaptation, organismal traits represent evolved responses to aspects of the environment

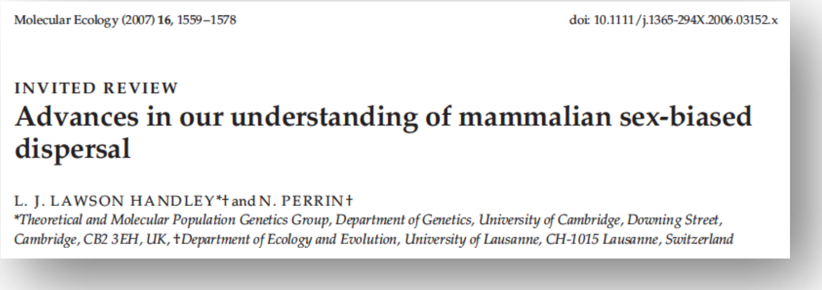
- more precise organismal traits are needed…

panTHERIA database of Class Mammalia (Jones et al. 2009)

- 55 organismal traits by 4630 species.

- repeat the RF analysis using only mammals (954 species)



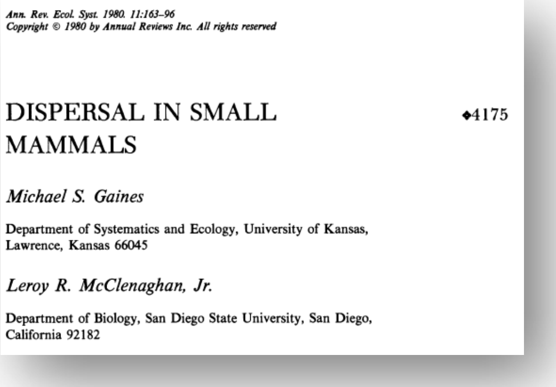

Molecular Ecology (2007) 16, 1559–1578 doi: 10.1111/j.1365-294X.2006.03152.x

INVITED REVIEW

Advances in our understanding of mammalian sex-biased dispersal

L. J. LAWSON HANDLEY*† and N. PERRIN†

Theoretical and Molecular Population Genetics Group, Department of Genetics, University of Cambridge, Downing Street, Cambridge, CB2 3EH, UK, †Department of Ecology and Evolution, University of Lausanne, CH-1015 Lausanne, Switzerland


Ann. Rev. Ecol. Syst. 1980. 11:163–96
Copyright © 1980 by Annual Reviews Inc. All rights reserved

DISPERSAL IN SMALL MAMMALS ◆4175

Michael S. Gaines

Department of Systematics and Ecology, University of Kansas, Lawrence, Kansas 66045

Leroy R. McClenaghan, Jr.

Department of Biology, San Diego State University, San Diego, California 92182

What factors promote intraspecific genetic structure?

1. on a global scale, environmental and geographic distance are broadly correlated within species

2. geographic attributes such as maximum latitude and range size are the best predictors of which species are likely to exhibit IBD/E

3. organismal traits may be difficult to compare across the Tree of Life

Classic questions…

…on global scales

Traditional comparative phylogeography to particular regions:

- SE US (*Avise 2000*)
- Europe (*Hewitt 2000*)
- Pacific Northwest of NA (*Carstens et al. 2005*)

One traditional goal of phylogeographic work has been to understand how particular species respond to large scale climatic shifts (e.g., such as that of the end Pleistocene).

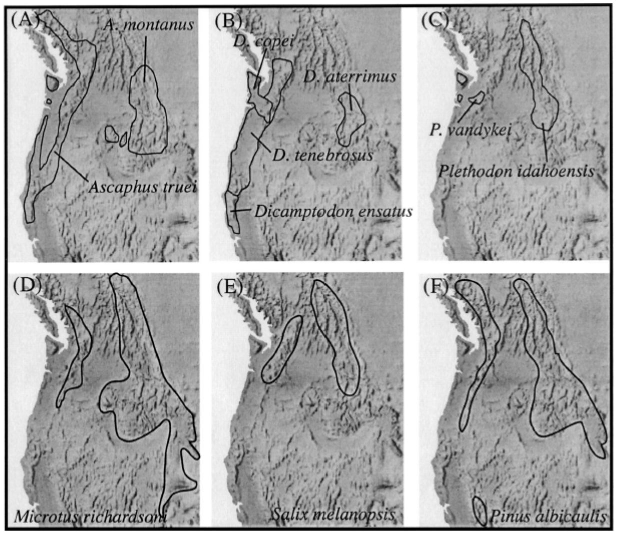
Classic questions…

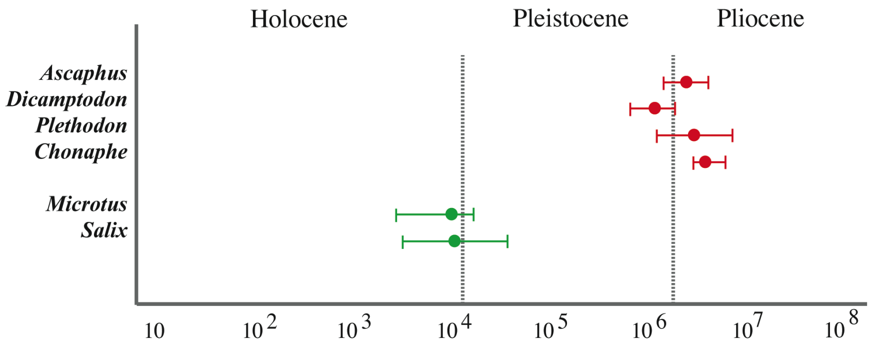
…on global scales

Traditional comparative phylogeography to particular regions:

- SE US (*Avise 2000*)
- Europe (*Hewitt 2000*)
- Pacific Northwest of NA (*Carstens et al. 2005*)

One traditional goal of phylogeographic work has been to understand how particular species respond to large scale climatic shifts (e.g., such as that of the end Pleistocene).





Classic questions…

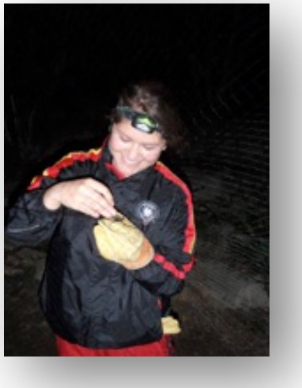
…on global scales

How did bats respond to climate change….
…on a global scale

- downloaded >30,000 sequences from 123 species with greater than 15 georeferenced samples

- **A**pproximate **B**ayesian **C**omputation used to calculate the probability of two models (expansion, bottleneck) in all species

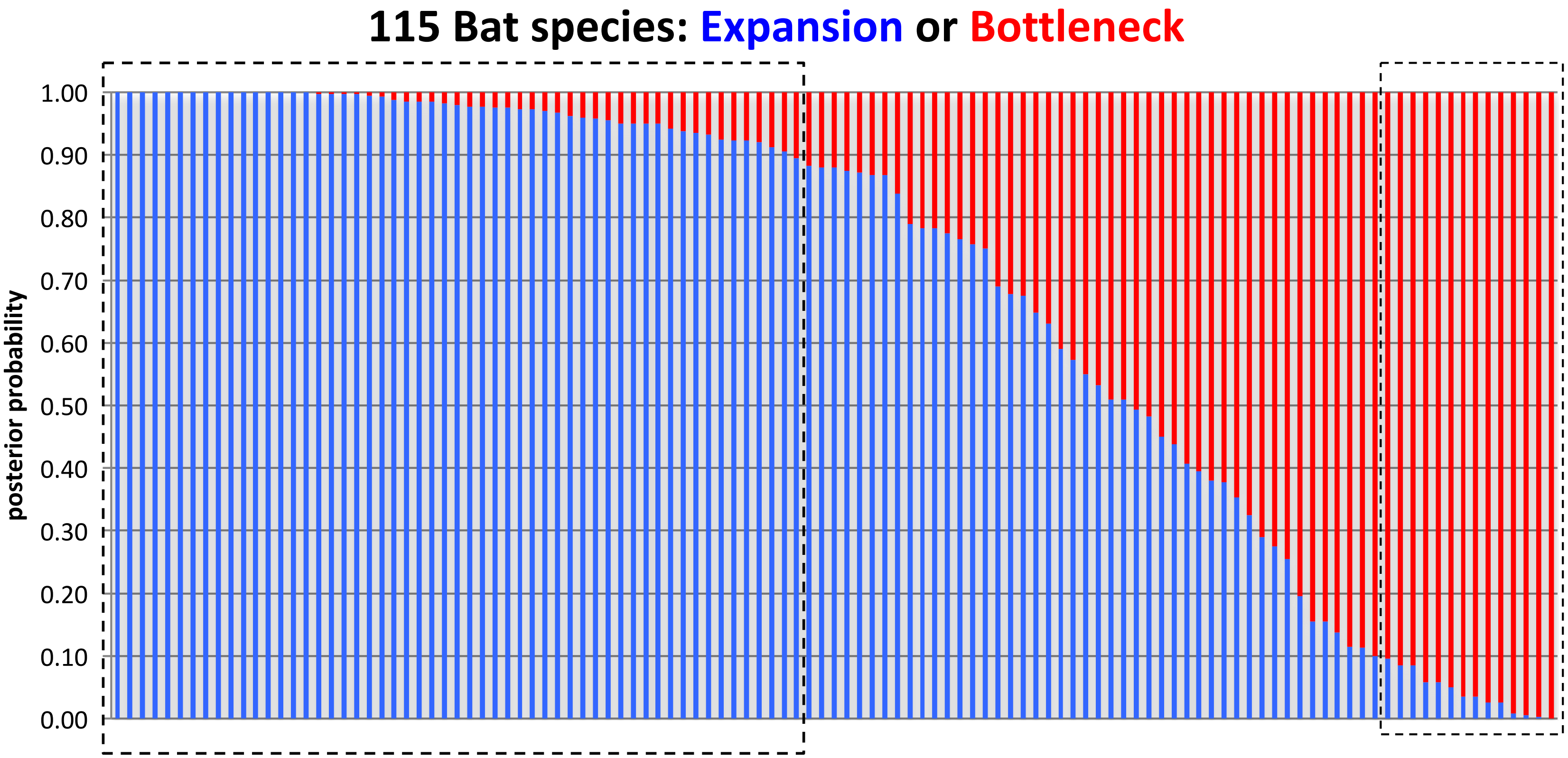
- species distribution modeling to compare predicted size of current range to predicted range size at end Pleistocene (thanks to Ariadna Morales!)



Classic questions…

…on global scales

How did bats respond to climate change on a global scale?



Classic questions…

…on global scales

How did bats respond to climate change on a global scale?

- Adult body mass is significantly correlated with $\mathrm{PP}_{\mathrm{expansion}}$.
- Expansion species are nearly twice the size as bottleneck species (19.9 : 10.8)
- Related to dietary niche, also significantly correlated.

Classic questions…

…on global scales

How did bats respond to climate change on a global scale?

Organismal Traits

- body size
- wing shape
- breeding strategy
- roosting location
- dietary niche

Environmental

- predicted size of current range
- predicted size of range at LGM
- maximum latitude
- mid point latitude
- average temperature in observed range

Identify factors (intrinsic, environmental) that predict the observed response:

Classic questions…

…on global scales

How did bats respond to climate change on a global scale?

RF P=0.9	OOB error	Expansion error	Bottleneck error	n	n Expansion	n Bottleneck
all spatial variables	0.2115	0.089	1.0	52	45	7
change spatial variables	0.1923	0.067	1.0	52	45	7
all variables	0.1522	0.000	1.0	46	39	7
other variables	0.1379	0.000	1.0	58	50	8

RF P=0.7	OOB error	Expansion error	Bottleneck error	n	n Expansion	n Bottleneck
all spatial variables	0.292	0.121	1.0	72	58	14
change spatial variables	0.278	0.121	0.9	72	58	14
all variables	0.194	0.020	1.0	62	51	11
other variables	0.190	0.015	1.0	79	65	14

Random forest prediction error rates unacceptably high due to disparity in response variables…

Quantifying Biodiversity: Bat cryptic species.

Most available DNA sequence data lack georeferencing.

(previous analysis based on 13% of total mtDNA data)

1. downloaded all mtDNA from bats
2. aligned by gene
   - ~ 20,000 barcoding loci (10,421 *cyt b* seqs, 9552 *COI* seqs)
   - 842 nominal species (75% of described)
   - 1116 total species in Chiroptera (Wilson & Reeder, 3rd Ed)
3. estimated distributions of gene trees (by family)
4. used GMYC model (Pons et al. 2006) to estimate number of cryptic bat species

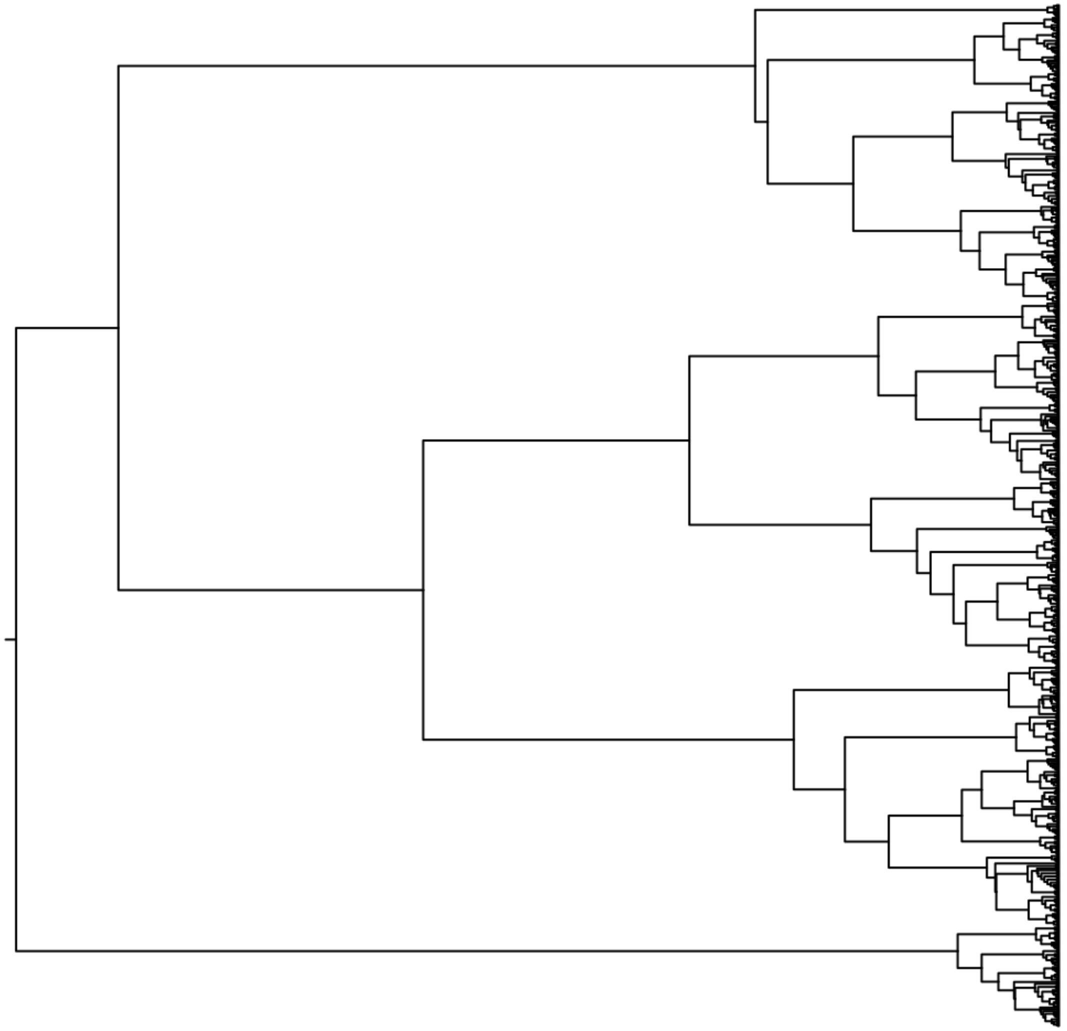
Quantifying Biodiversity: Bat cryptic species.

General **M**ixed **Y**ule – **C**oalescent model (Pons et al. 2006)

Quantifying Biodiversity: Bat cryptic species.

General **M**ixed **Y**ule – **C**oalescent model (Pons et al. 2006)

- similar to *lineage through time plots* in that it considers rates of branching in rooted ultrametric trees

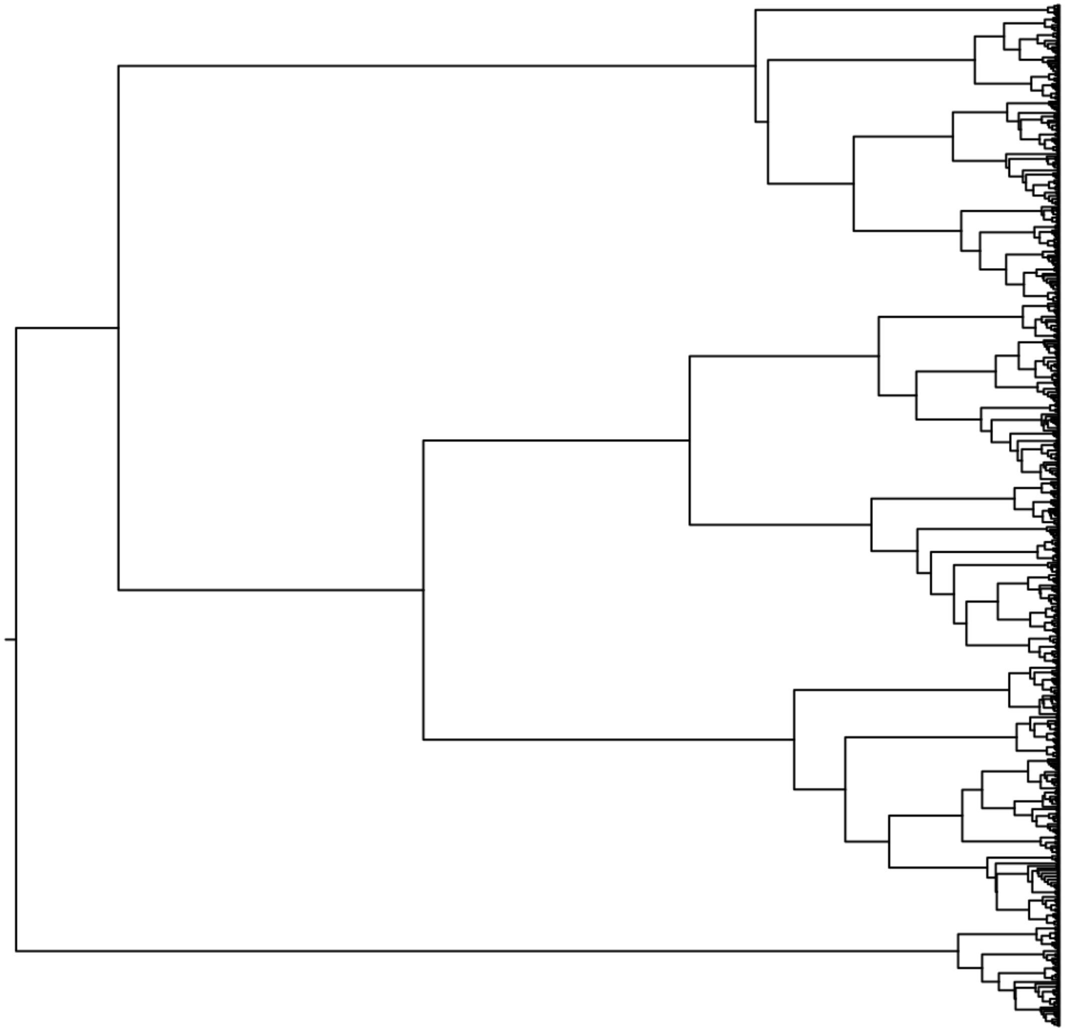


Quantifying Biodiversity: Bat cryptic species.

General **M**ixed **Y**ule – **C**oalescent model (Pons et al. 2006)

- similar to *lineage through time plots* in that it considers rates of branching in rooted ultrametric trees

- attempts to find point where rate of branching transitions from slow to fast



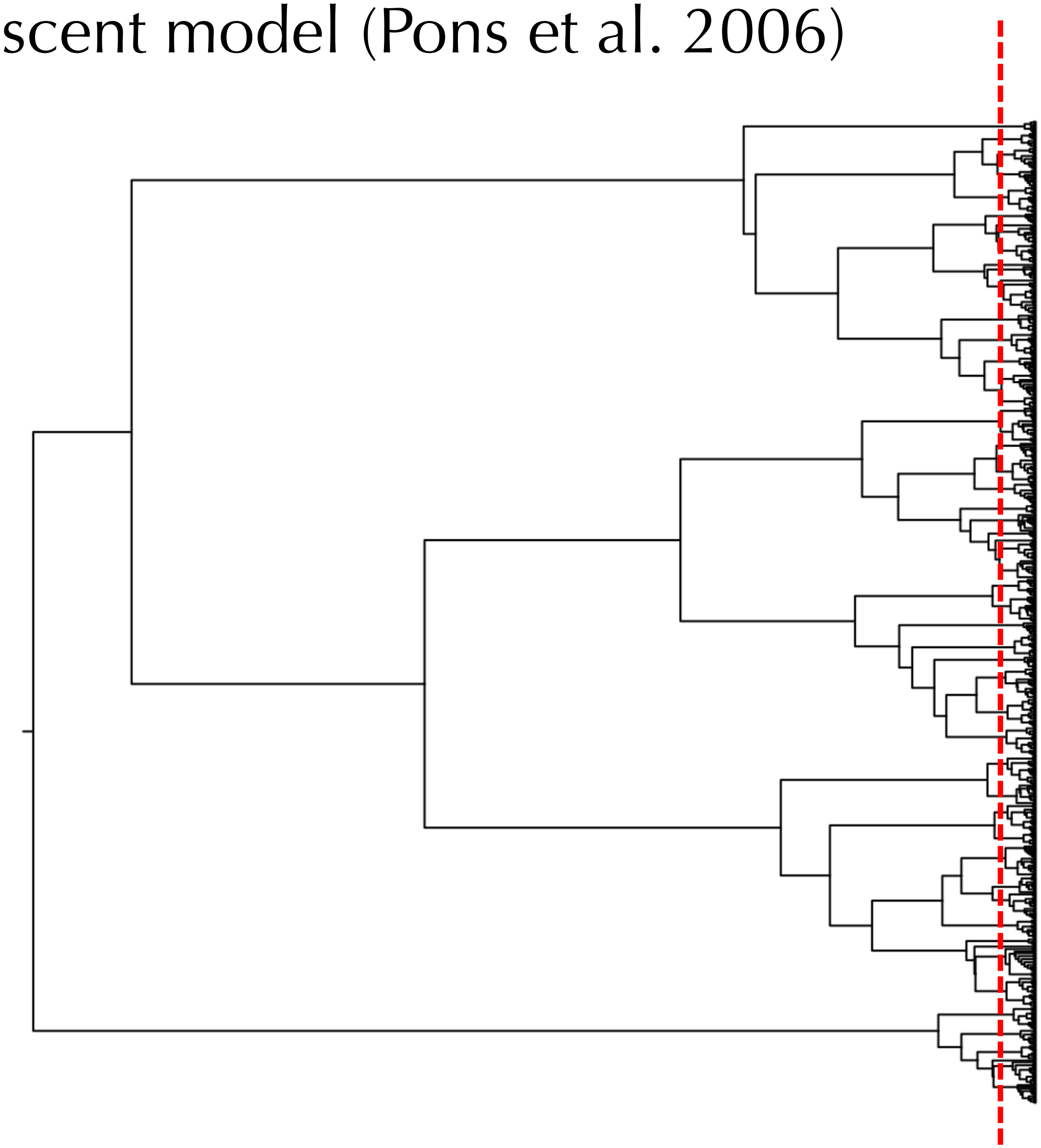
Quantifying Biodiversity: Bat cryptic species.

General **M**ixed **Y**ule – **C**oalescent model (Pons et al. 2006)

- similar to *lineage through time plots* in that it considers rates of branching in rooted ultrametric trees

- attempts to find point where rate of branching transitions from slow to fast

- assumes rate of speciation is slow compared to rate of allele coalescence



Quantifying Biodiversity: Bat cryptic species.

Species Diversity - Chiroptera

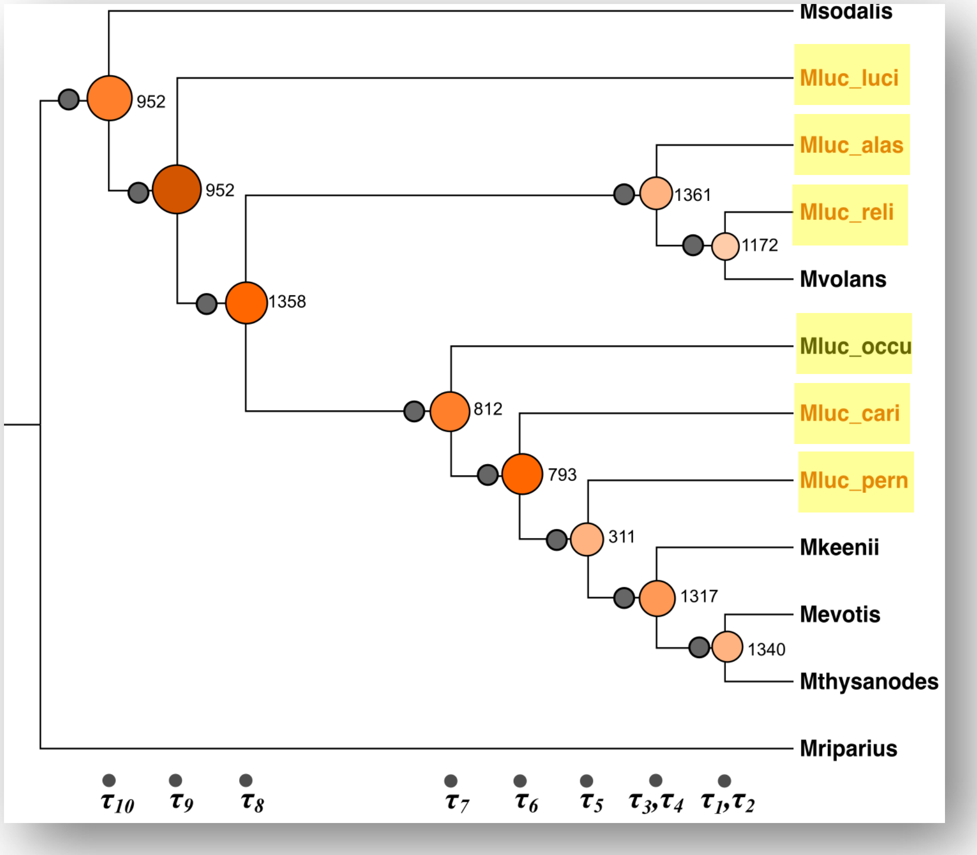
2161 discrete GMYC entities were detected.

GMYC has been criticized as being biased towards overestimation (e.g., Esselstyn et al. 2012), but…

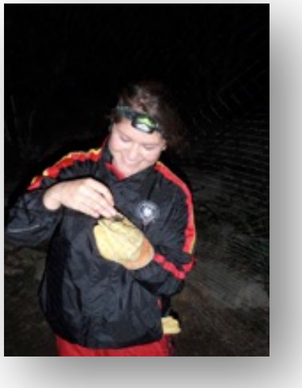
…follow up investigations often confirm GMYC results.

Quantifying Biodiversity: Bat cryptic species.

Species Diversity - Chiroptera



- *Myotis lucifugus* contains multiple described subspecies.
- Our results delimit 4 GMYC entities in *M. lucifugus*
- Morales et al. (in rev) collected ~800 UCE loci from *Myotis* bats.
- *M. lucifugus* subspecies are not monophyletic…



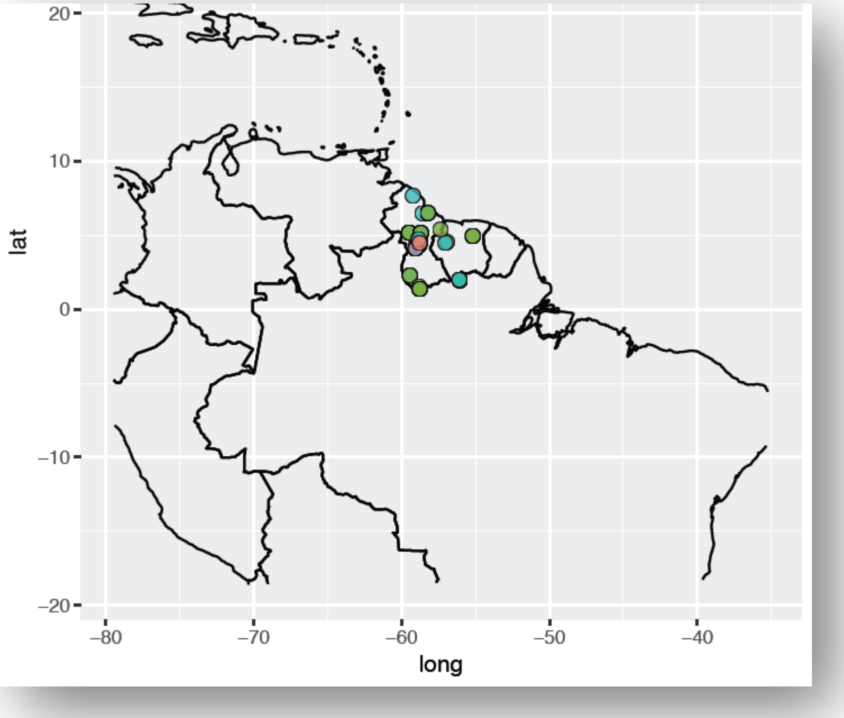
data from Ariadna's DDIG….

Quantifying Biodiversity: Bat cryptic species.

Species Diversity - Chiroptera

Geographic data are available from 332 nominal species, and 134 contained data for >1 of the GMYC entities.

- GMYC entities contradicted by geographic distribution of samples (0.13).



contradicts GMYC: *Artibeus concolor*

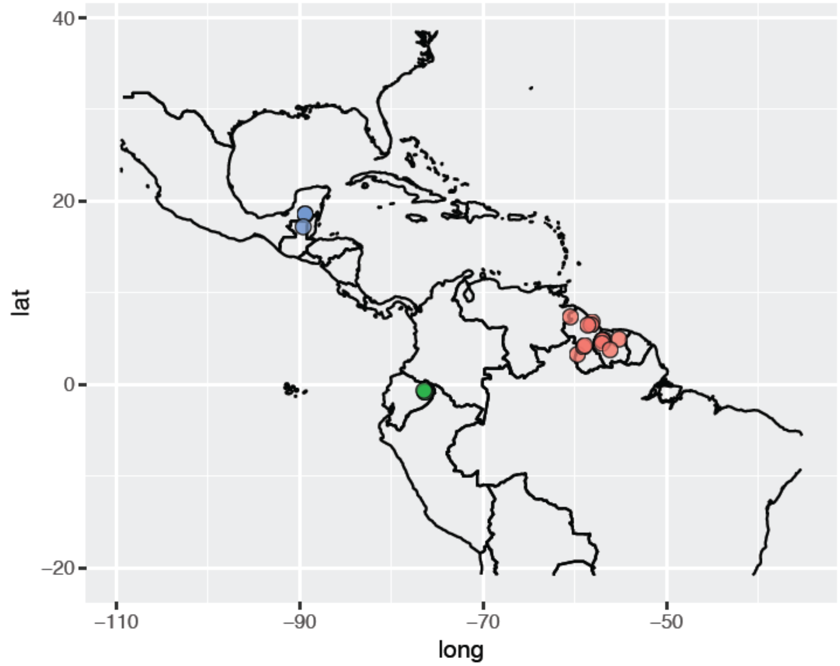
Quantifying Biodiversity: Bat cryptic species.

Species Diversity - Chiroptera

Geographic data are available from 332 nominal species, and 134 contained data for >1 of the GMYC entities.

- GMYC entities contradicted by geographic distribution of samples (0.13).

- GMYC entities corresponded to discrete geographic clusters for all (0.29) or some (0.31) of the delimited groups



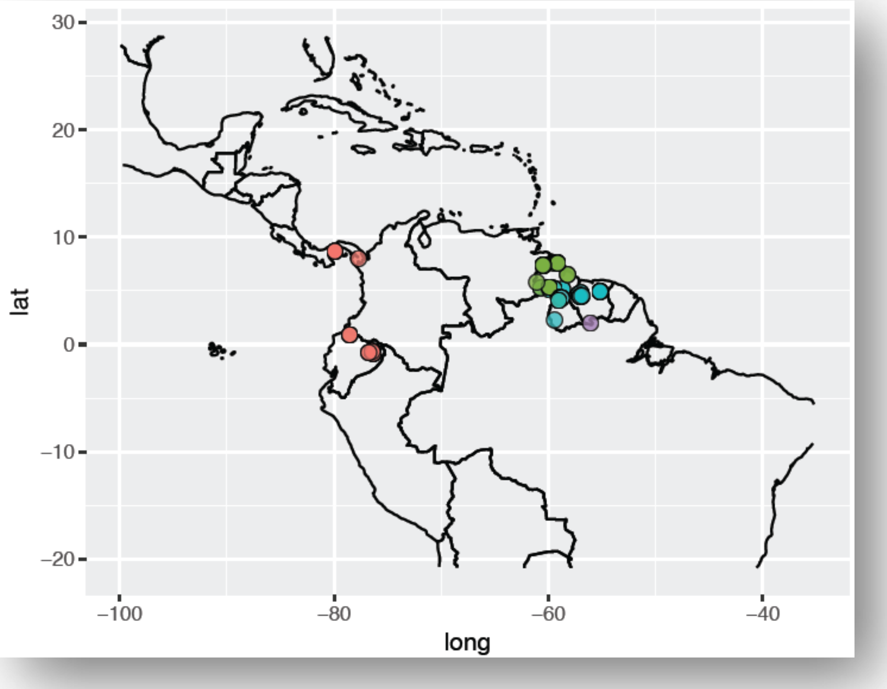
all discrete: *Eptesicus furinalis*

Quantifying Biodiversity: Bat cryptic species.

Species Diversity - Chiroptera

Geographic data are available from 332 nominal species, and 134 contained data for >1 of the GMYC entities.

- GMYC entities contradicted by geographic distribution of samples.
- GMYC entities corresponded to discrete geographic clusters for all (0.29) or some (0.31) of the delimited groups



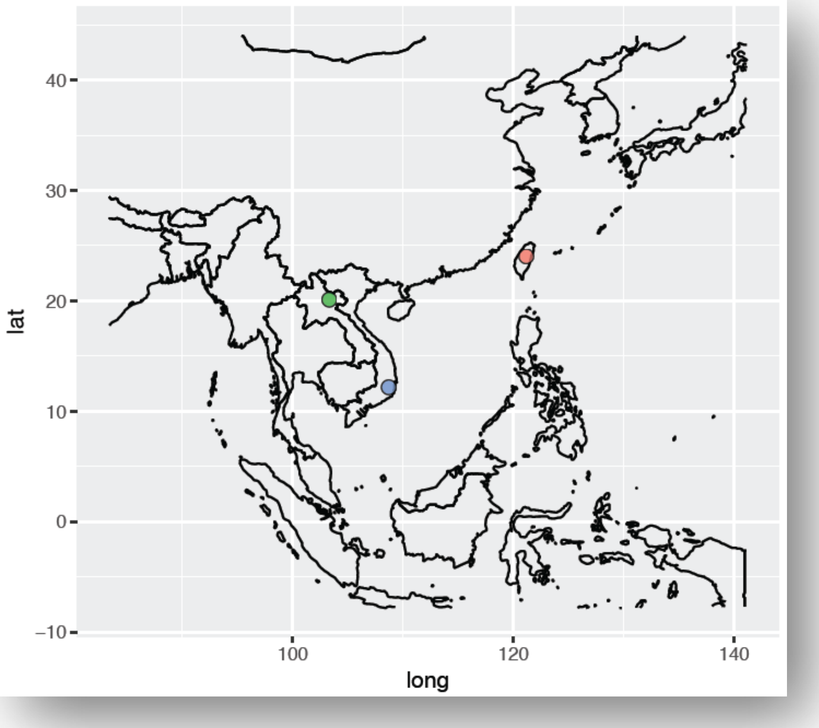
some discrete: *Carollia brevicada*

Quantifying Biodiversity: Bat cryptic species.

Species Diversity - Chiroptera

Geographic data are available from 332 nominal species, and 134 contained data for >1 of the GMYC entities.

- GMYC entities contradicted by geographic distribution of samples.
- GMYC entities corresponded to discrete geographic clusters for all (0.29) or some (0.31) of the delimited groups
- Sampling inadequate to draw conclusions (0.27)



Inadequate sampling: *Coelops frithii*

Quantifying Biodiversity: Bat cryptic species.

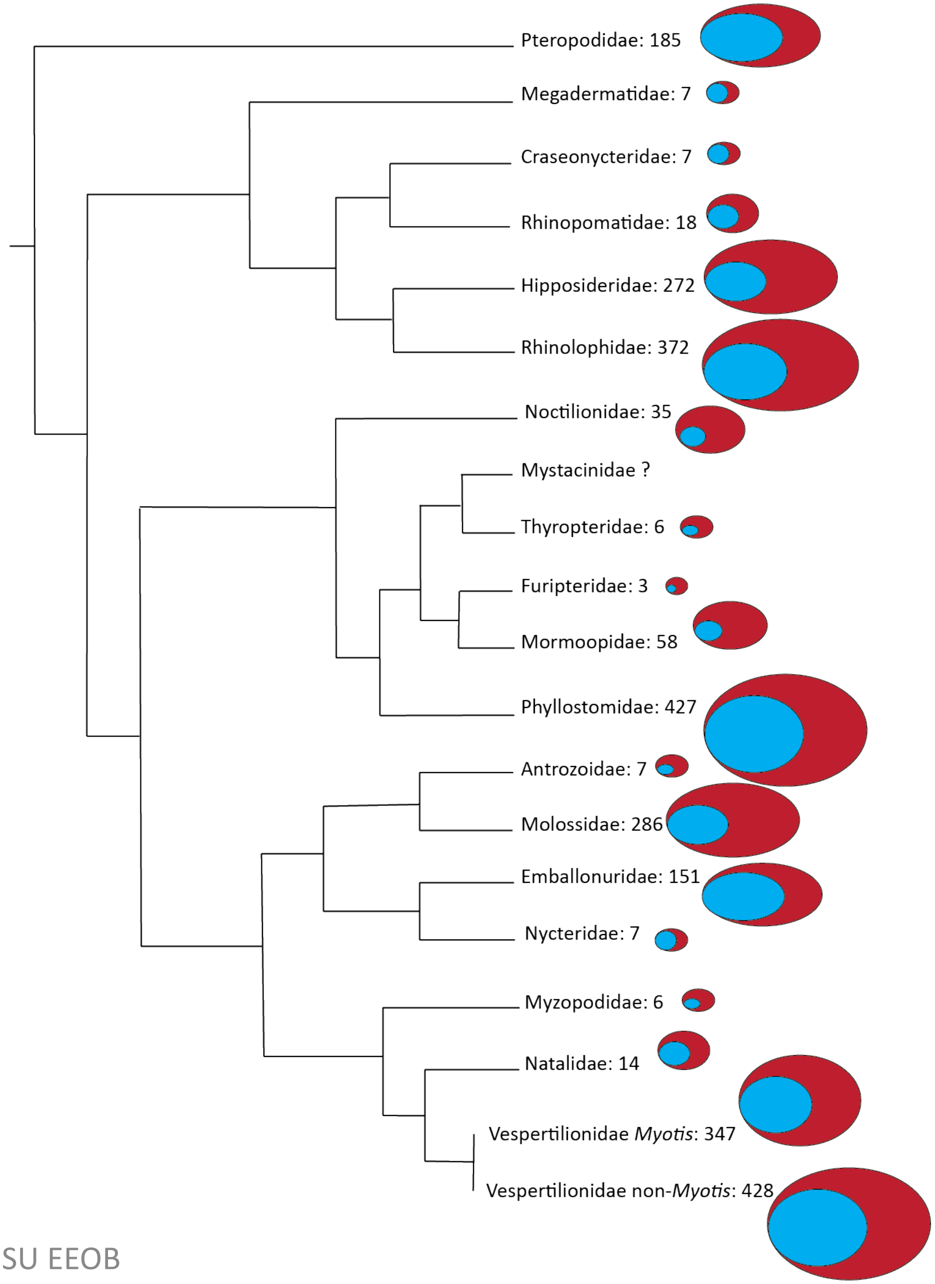
Species Diversity - Chiroptera

Nominal / **Predicted**

- 842 nominal species in analysis
- 1116 described in Chiroptera
- 75% of species represented

- **2073 GMYC entities**

- **~2700 bat species!?!**



Big data!

Conclusions

~300 years of taxonomic work
~ 6 million genetic data points
~600 million occurrence records on GBIF

- NSF Proposal to fund **phylogatr**: database aggregator w/ R pipelines to facilitate meta-analysis of phylogeographic data.

DEPARTMENT OF
EVOLUTION, ECOLOGY AND ORGANISMAL BIOLOGY
eeob.osu.edu

Ohio Supercomputer Center
An OH·TECH Consortium Member

THE OHIO STATE UNIVERSITY

Postdocs
Margaret Koopman
Yi-Hsin Erica Tsai
Amanda Zellmer
Tereza Thomé
Michael Gruenstaeudl
Tara Pelletier

Grad Students
Sarah Hird
Noah Reid
John McVay
Tara Pelletier
Jordan Satler
Ariadna Morales
Greg Wheeler
Megan Smith
Cole Thompson

NSF Funding
DBI - 1661029
DEB - 1701810
DBI - 1560116
DEB - 1501474
DEB - 1457519
DEB - 1403034
DEB - 1500774
DEB - 1257784
OISE -1118408
DEB - 0918212
DEB - 0956069

Undergrads
Danielle Fuselier
Holly Stoute
Dan Ence
Matt Demarest
Maxim Kim
Edwin Rice
Ivy Larkin
Katy Field