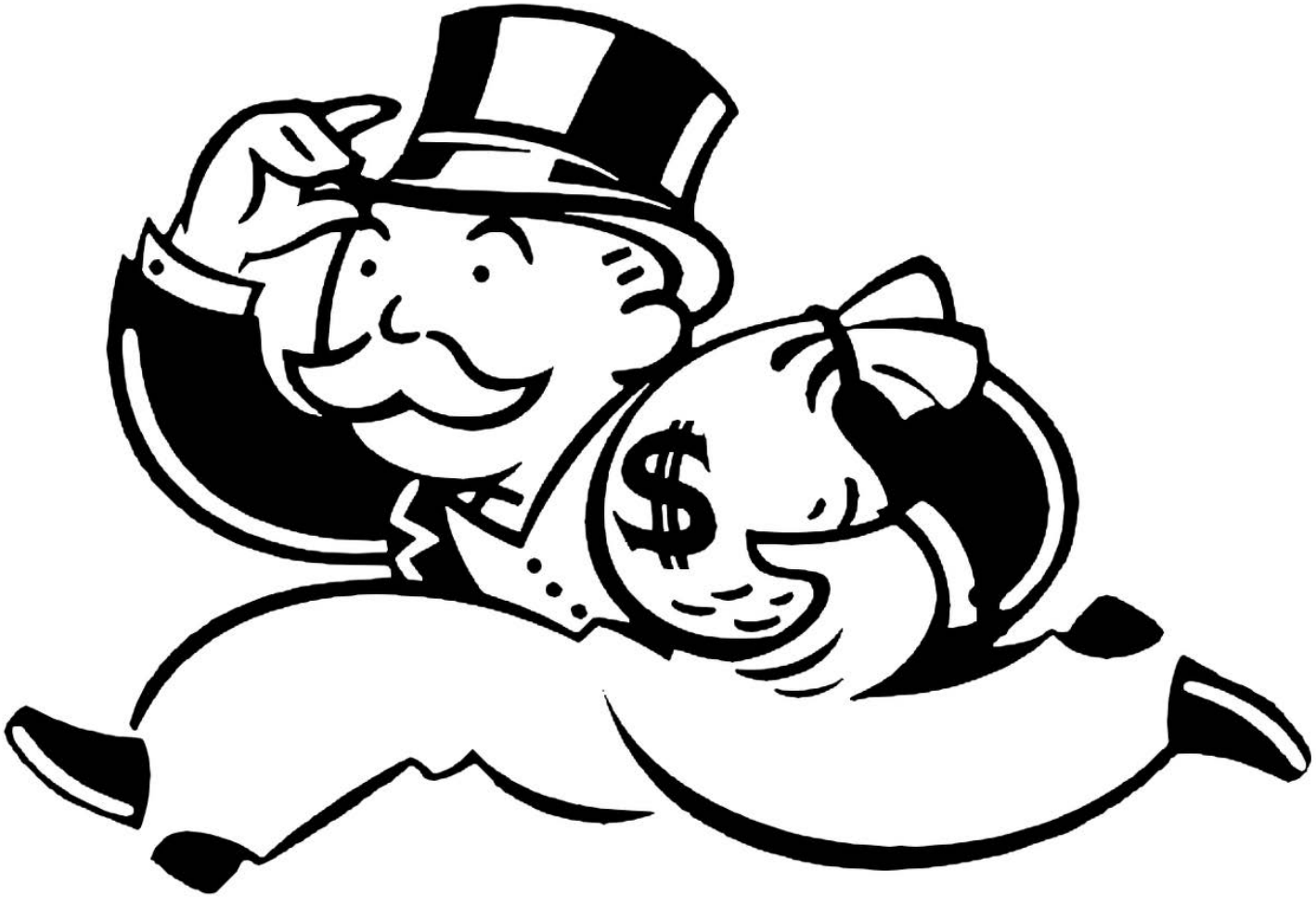
Target Capture: Where next?

Sonal Singhal
University of Michigan

In a non-zero-sum world, the dream would be one genome for every individual.

http://www.pbs.org/newshour

But, unless you are



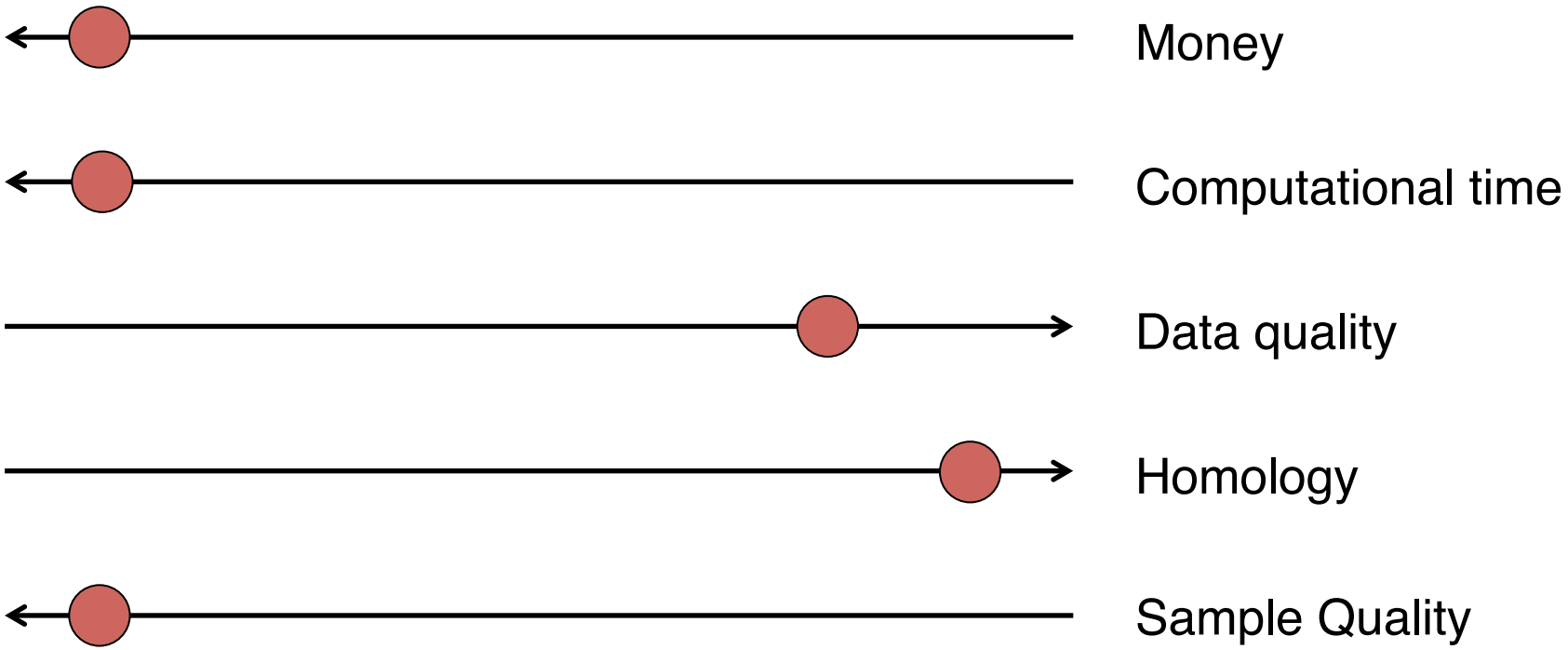
http://marxiststudent.com/

We need to subset the genome.

- Transcriptomes

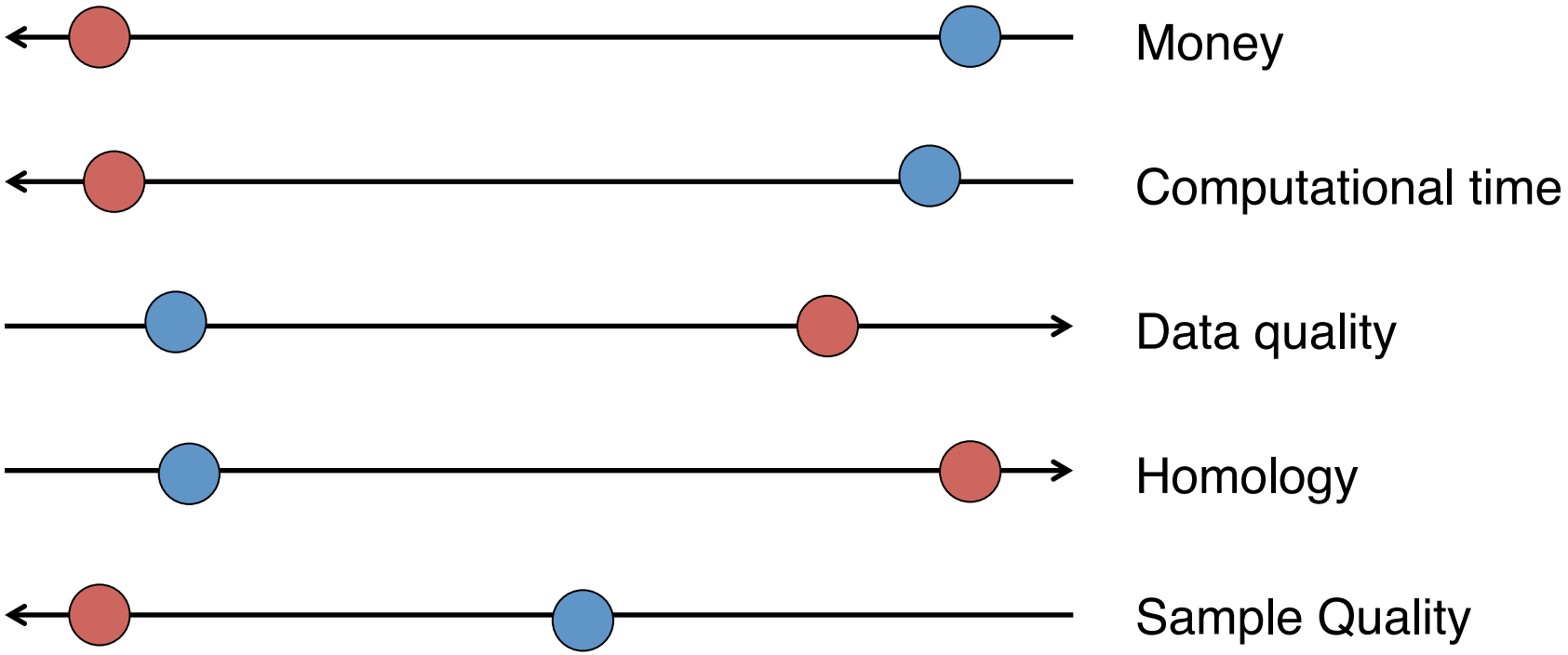
We need to subset the genome.

- Transcriptomes



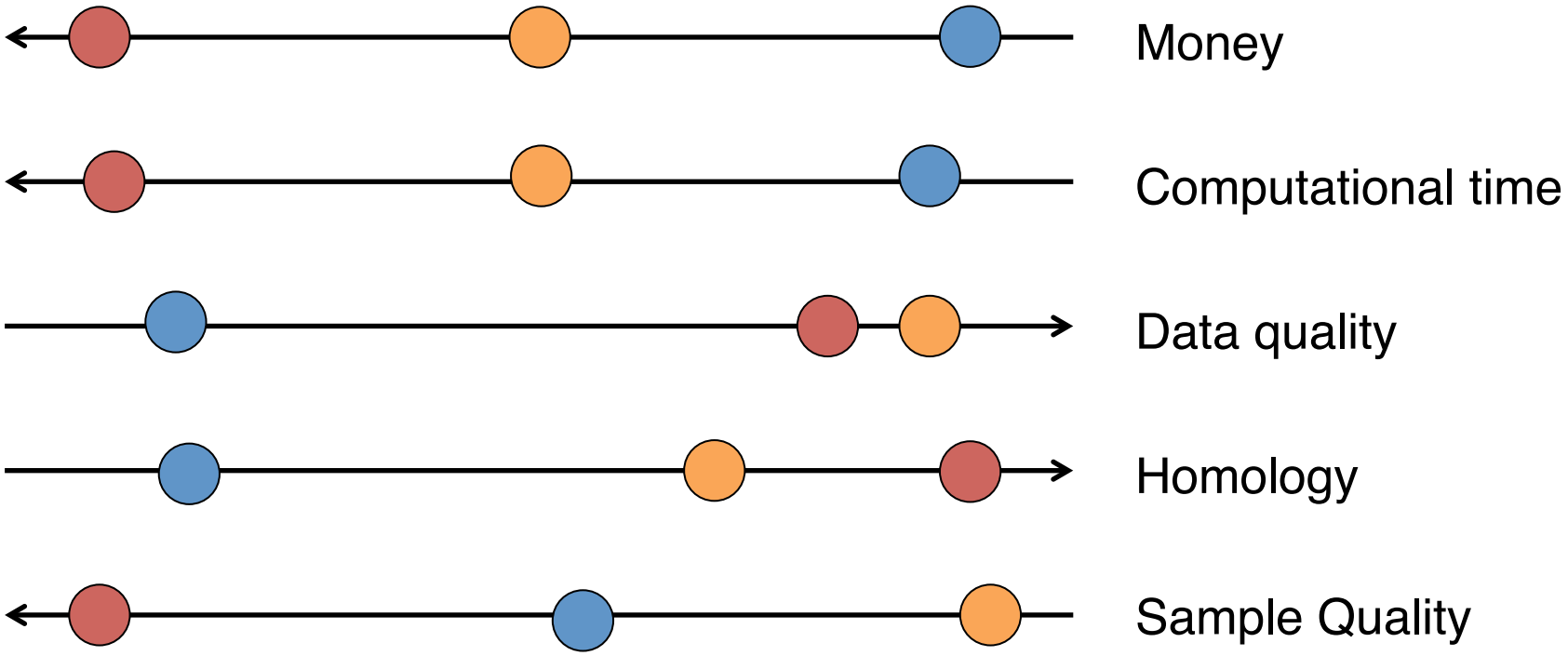
We need to subset the genome.

- Transcriptomes
- Restriction digest based approaches



We need to subset the genome.

- Transcriptomes
- Restriction digest based approaches
- Target capture



The best solution?

http://thisiswhyimbroke.com/

But, what to capture?

Multiplexed DNA Sequence Capture of Mitochondrial Genomes Using PCR Products

Tomislav Maricic, Mark Whitten, Svante Pääbo

Published: November 16, 2010 • http://dx.doi.org/10.1371/journal.pone.0014004

But, what to capture?

Multiplexed DNA Sequence Capture of Mitochondrial Genomes Using PCR Products

Tomislav Maricic, Mark Whitten, Svante Pääbo

Published: November 16, 2010 • http://dx.doi.org/10.1371/journal.pone.0014004

MOLECULAR ECOLOGY RESOURCES

Explore this journal >

Special Issue: Sequence Capture

Exon capture phylogenomics: efficacy across scales of divergence

Jason G. Bragg, Sally Potter, Ke Bi, Craig Moritz

But, what to capture?

Multiplexed DNA Sequence Capture of Mitochondrial Genomes Using PCR Products

Tomislav Maricic, Mark Whitten, Svante Pääbo

Published: November 16, 2010 • http://dx.doi.org/10.1371/journal.pone.0014004

MOLECULAR ECOLOGY RESOURCES

Explore this journal >

Special Issue: Sequence Capture

Exon capture phylogenomics: efficacy across scales of divergence

Jason G. Bragg, Sally Potter, Ke Bi, Craig Moritz

Hybridization Capture Using RAD Probes (hyRAD), a New Tool for Performing Genomic Analyses on Collection Specimens

Tomasz Suchan, Camille Pitteloud, Nadezhda S. Gerasimova, Anna Kostikova, Sarah Schmid, Nils Arrigo, Mila Pajkovic, Michał Ronikier, Nadir Alvarez

Published: March 21, 2016 • http://dx.doi.org/10.1371/journal.pone.0151651

But, what to capture?

Multiplexed DNA Sequence Capture of Mitochondrial Genomes Using PCR Products

Tomislav Maricic, Mark Whitten, Svante Pääbo

Published: November 16, 2010 • http://dx.doi.org/10.1371/journal.pone.0014004

MOLECULAR ECOLOGY RESOURCES

Explore this journal >

Special Issue: Sequence Capture

Exon capture phylogenomics: efficacy across scales of divergence

Jason G. Bragg, Sally Potter, Ke Bi, Craig Moritz

Hybridization Capture Using RAD Probes (hyRAD), a New Tool for Performing Genomic Analyses on Collection Specimens

Tomasz Suchan, Camille Pitteloud, Nadezhda S. Gerasimova, Anna Kostikova, Sarah Schmid, Nils Arrigo, Mila Pajkovic, Michał Ronikier, Nadir Alvarez

Published: March 21, 2016 • http://dx.doi.org/10.1371/journal.pone.0151651

Ultraconserved Elements Anchor Thousands of Genetic Markers Spanning Multiple Evolutionary Timescales

BRANT C. FAIRCLOTH[1,*], JOHN E. MCCORMACK[2], NICHOLAS G. CRAWFORD[3], MICHAEL G. HARVEY[2,4], ROBB T. BRUMFIELD[2,4], AND TRAVIS C. GLENN[5]

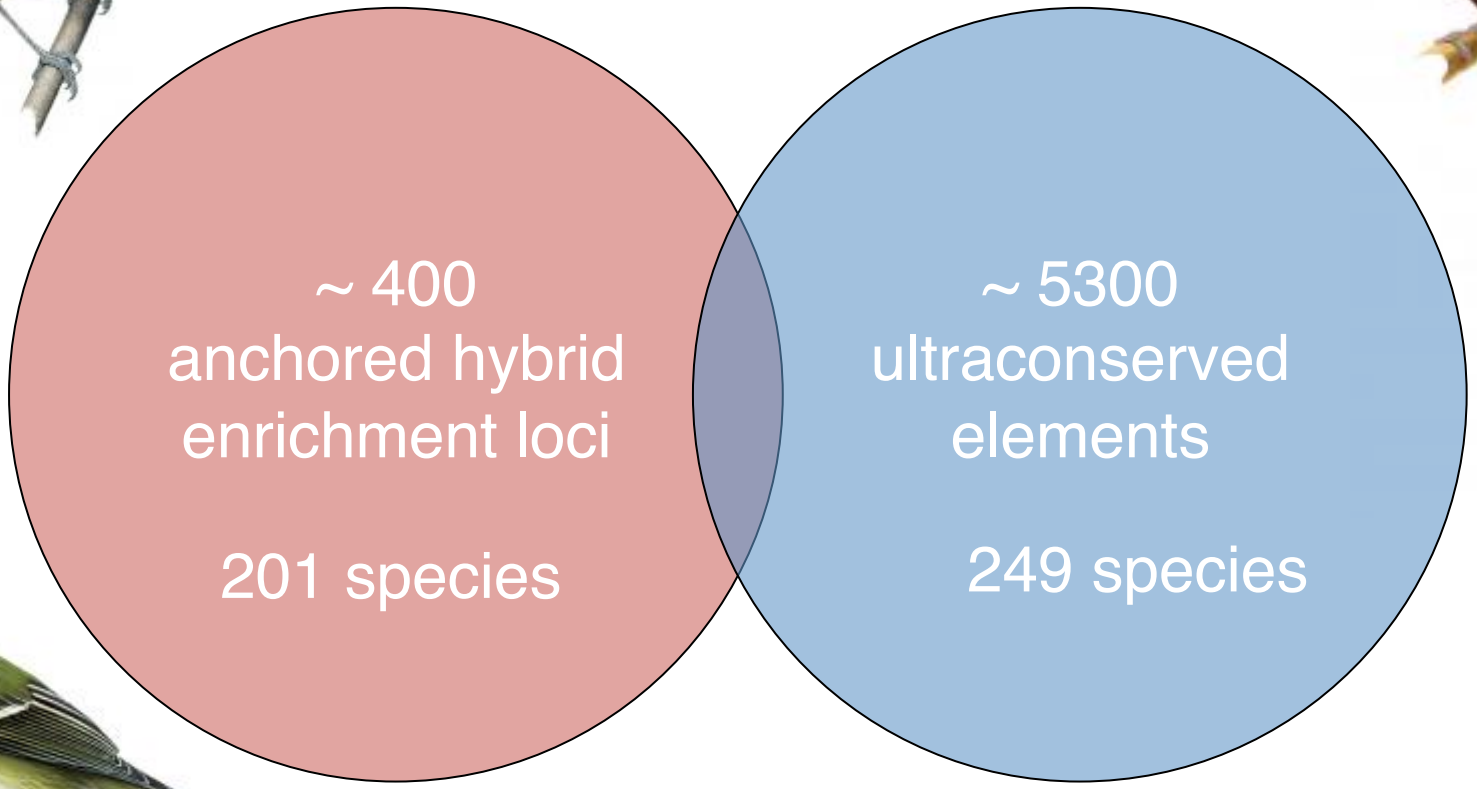
As a reviewer wisely pointed out, “everyone has their favorite probe set”.

What happens if we cannot decide?

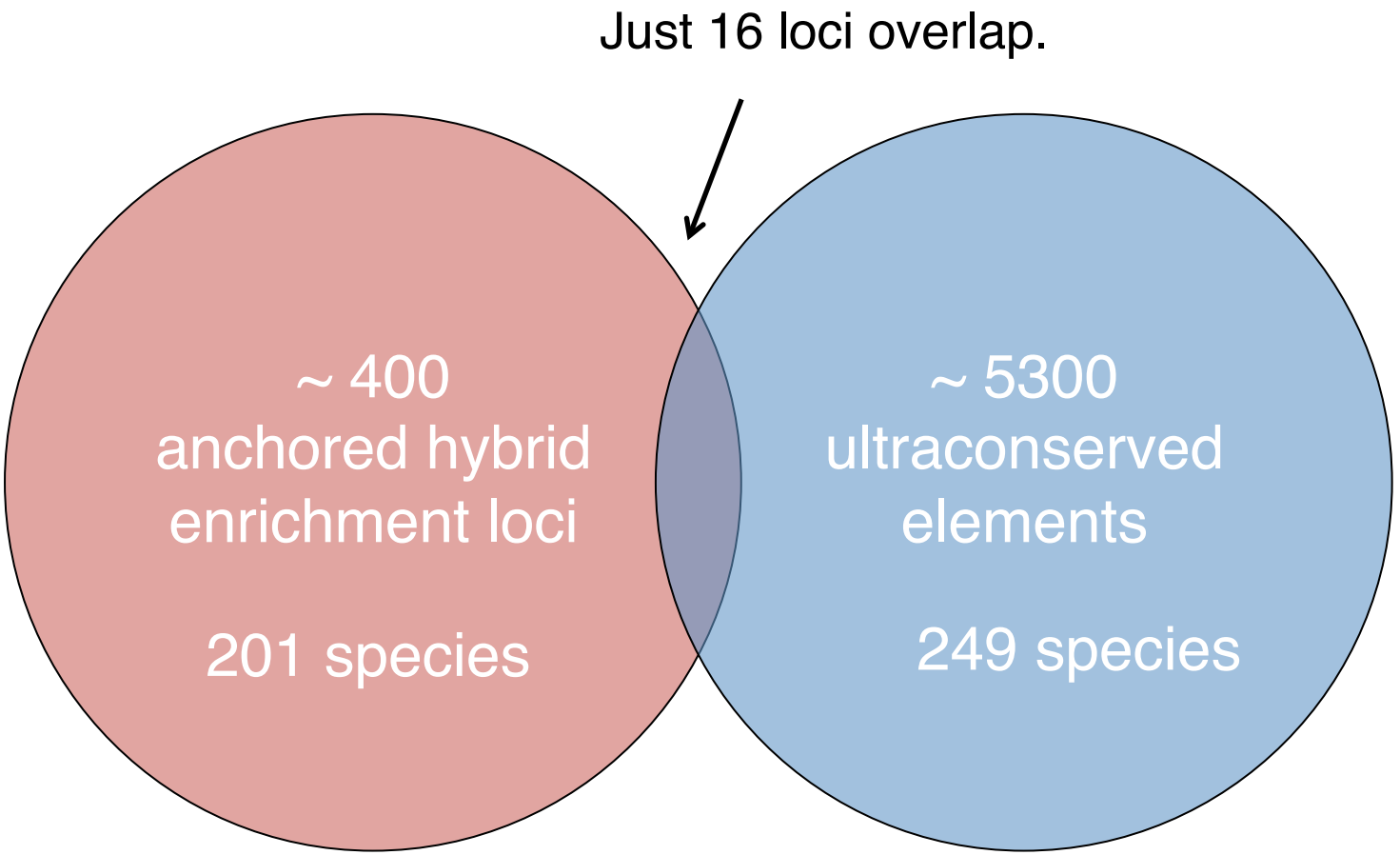
~ 400
anchored hybrid
enrichment loci

201 species

What happens if we cannot decide?



What happens if we cannot decide?

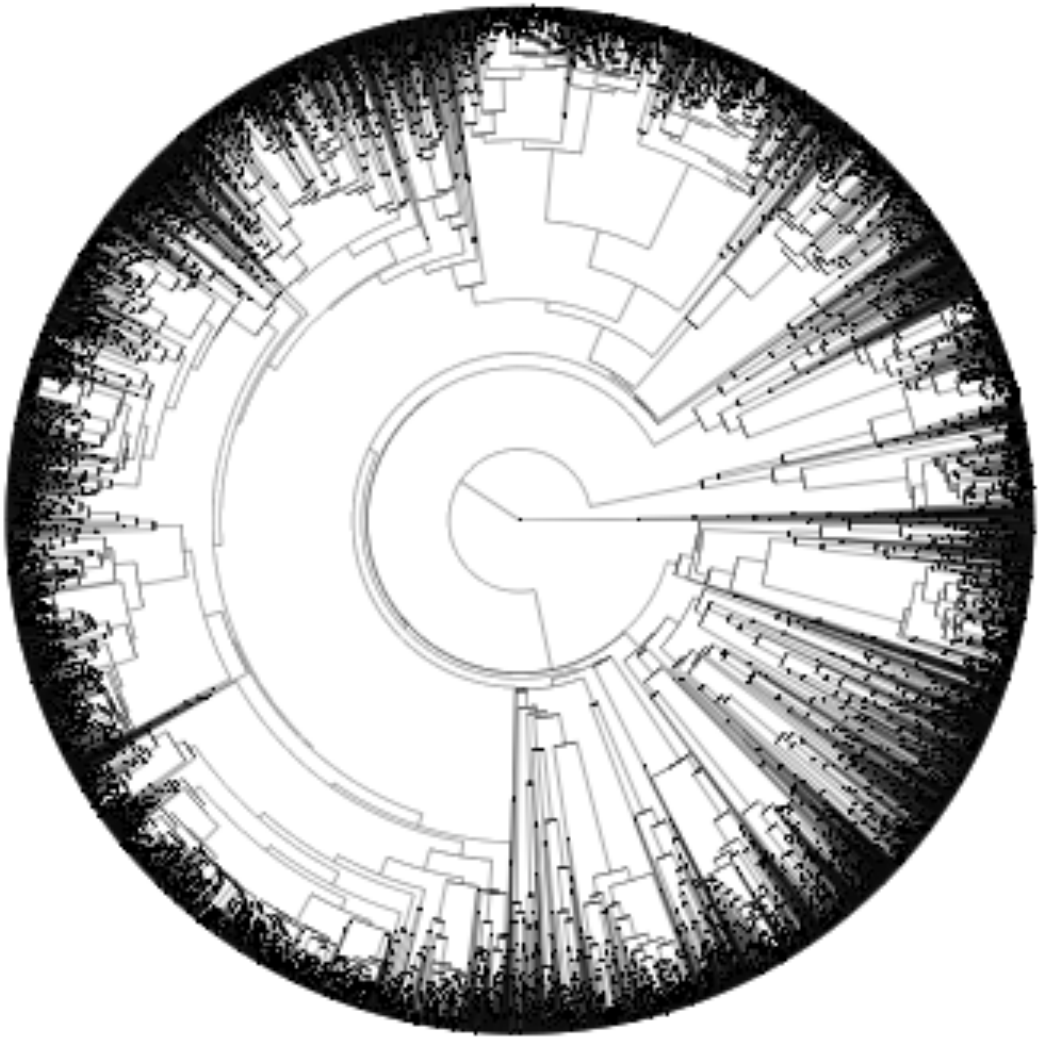


Early population genetics & phylogenetics necessarily targeted homologous markers across studies.

495K cytochrome b
424K cytochrome oxidase
356K 16S ribosomal RNA
164K 18S ribosomal RNA
110K NADH 4

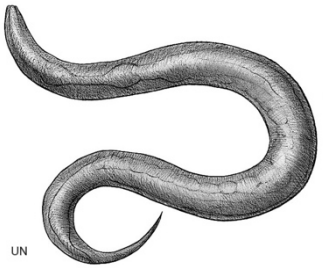
data from NCBI

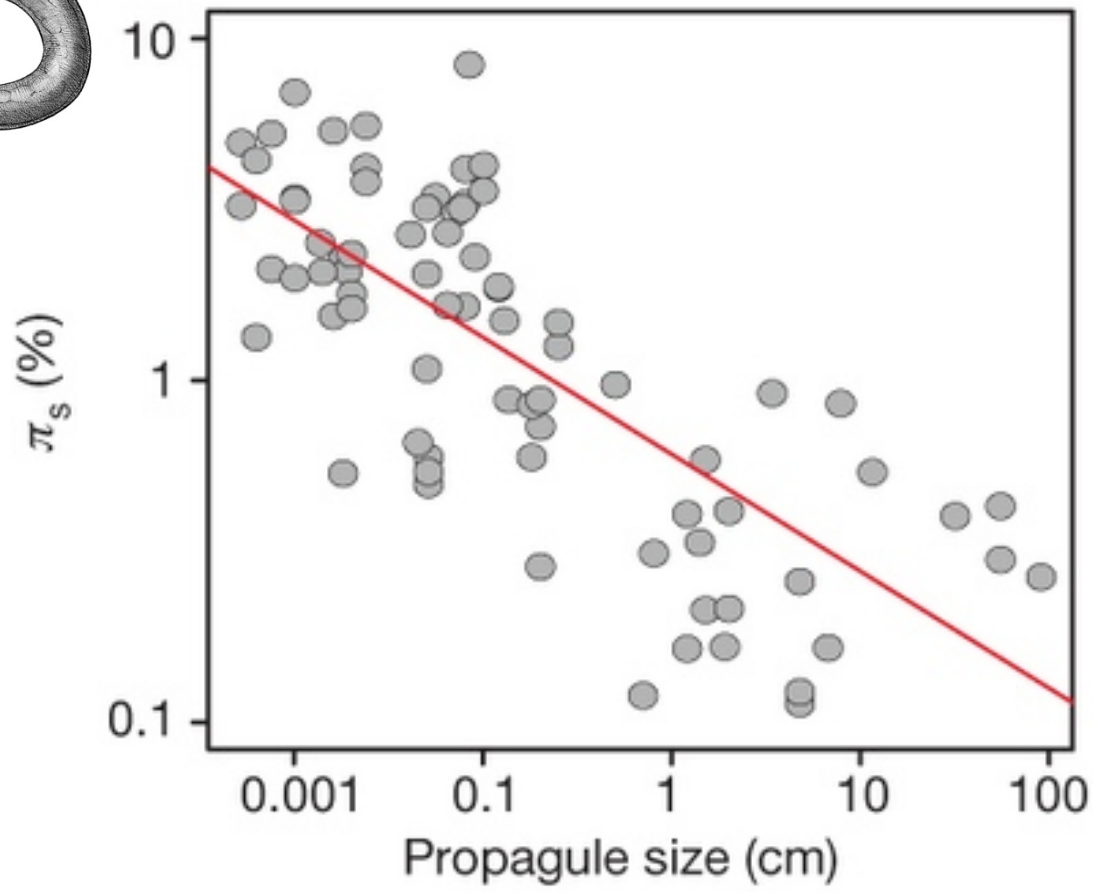
Our failure to do so now hinders comparative phylogenetics.



Jetz et al 2012
(see also
Fritz et al 2009,
Pyron et al 2013,
Zanne et al 2014)

And, it hinders comparative population genomics.





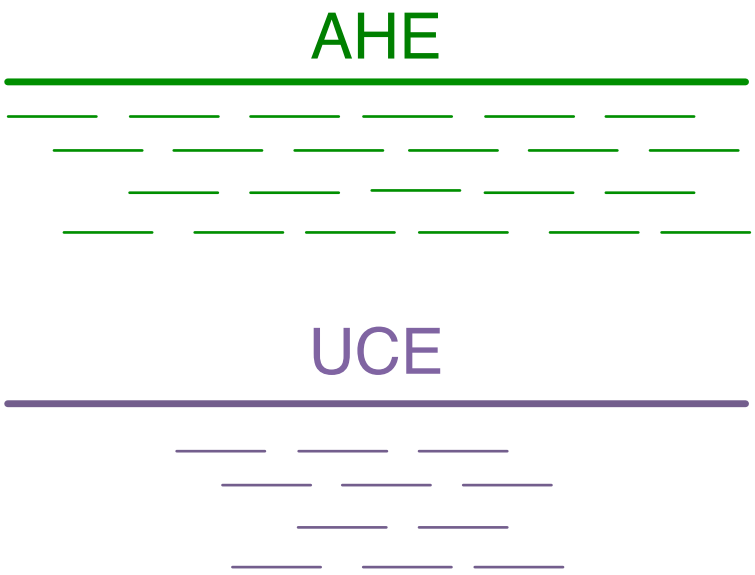
Romiguier et al 2014

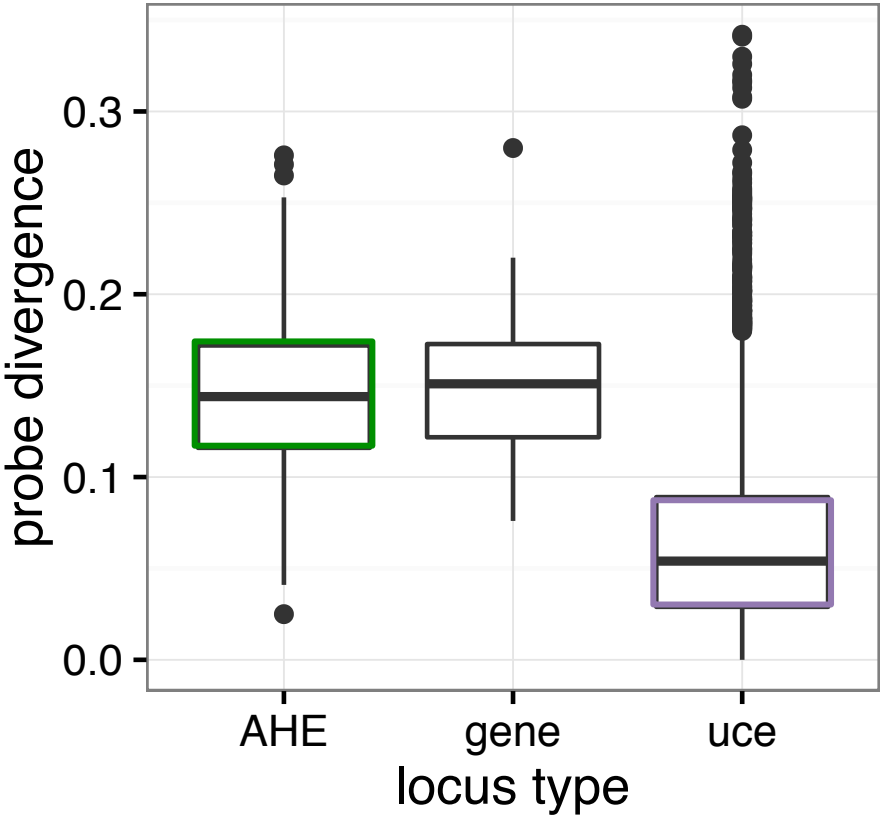
A simple way forward: create integrated probe sets.

anchored hybrid enrichment loci
+
ultraconserved elements
+
genes traditionally used in multiple-locus phylogenetics
+
fishing out the mitochondrial by-catch
=
~5400 markers that can be maximally integrated with most phylogenomic studies

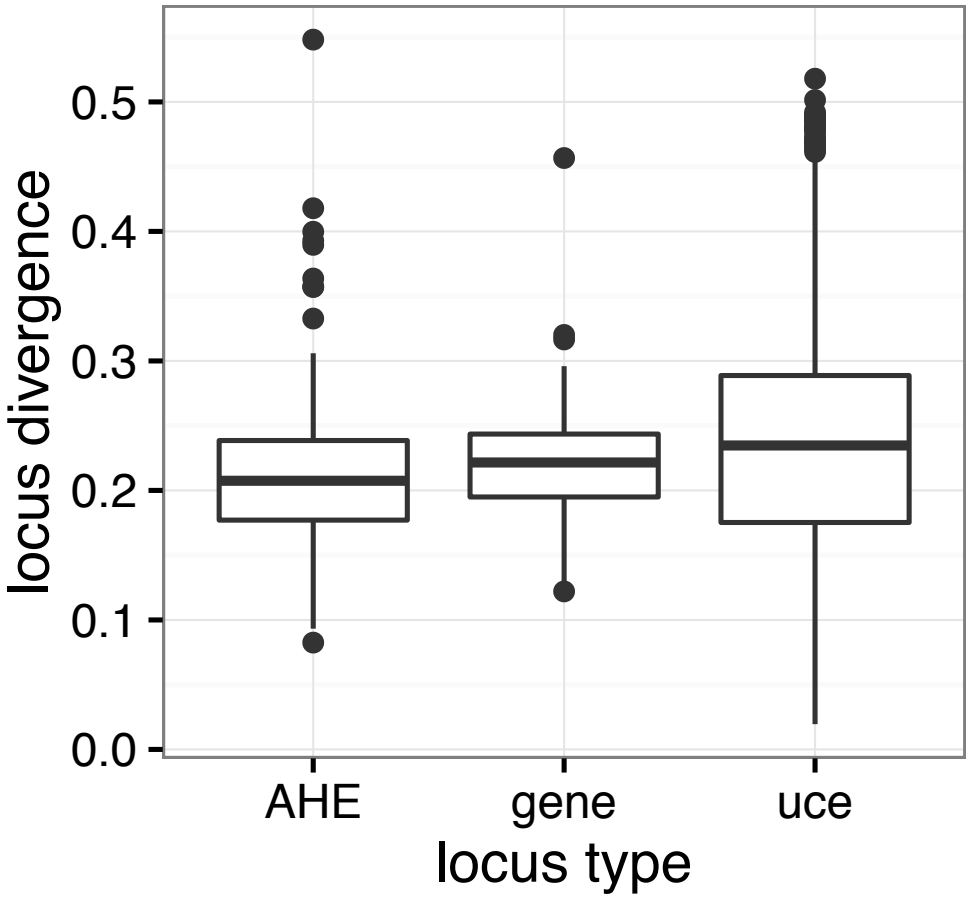
Wiens et al 2012,
Lemmon et al 2013,
Faircloth et al 2012

The markers have different logics.

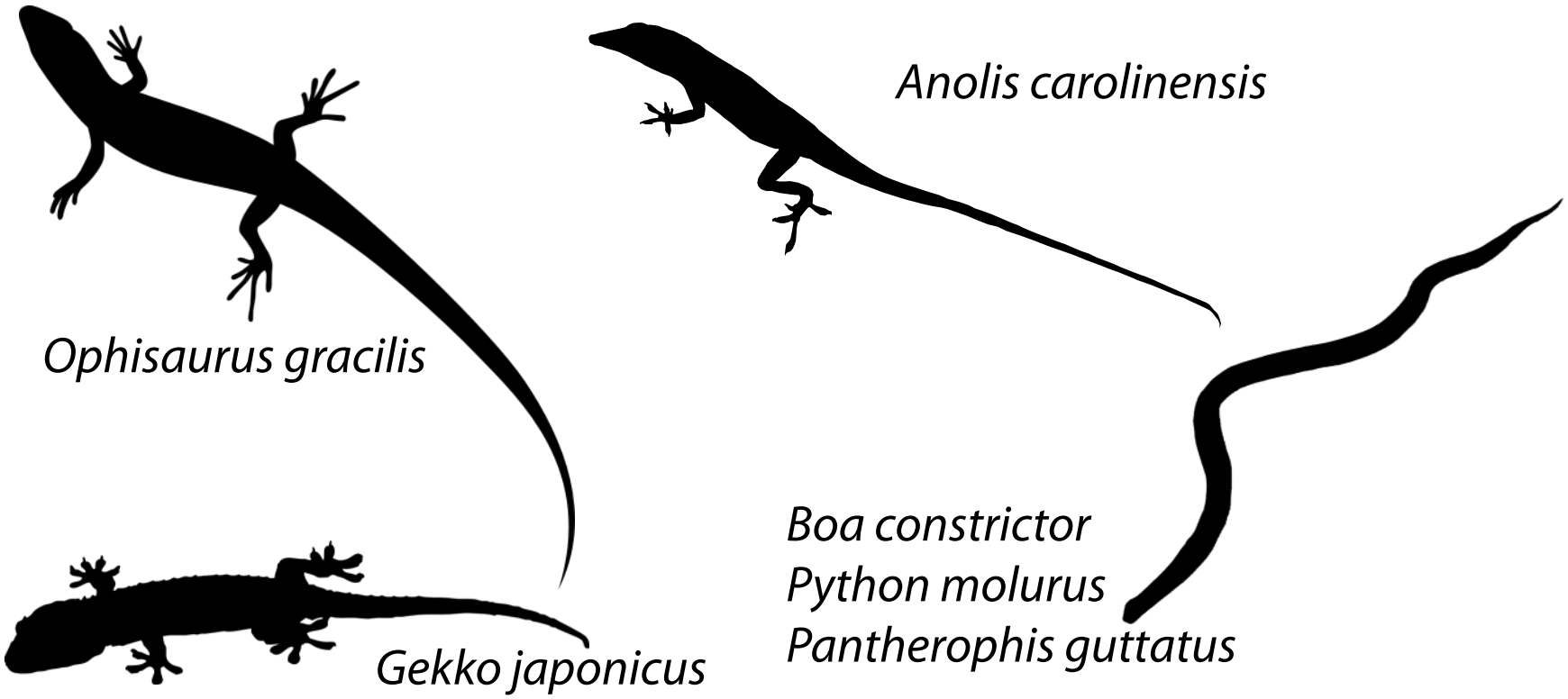




But the markers they generate are similar.



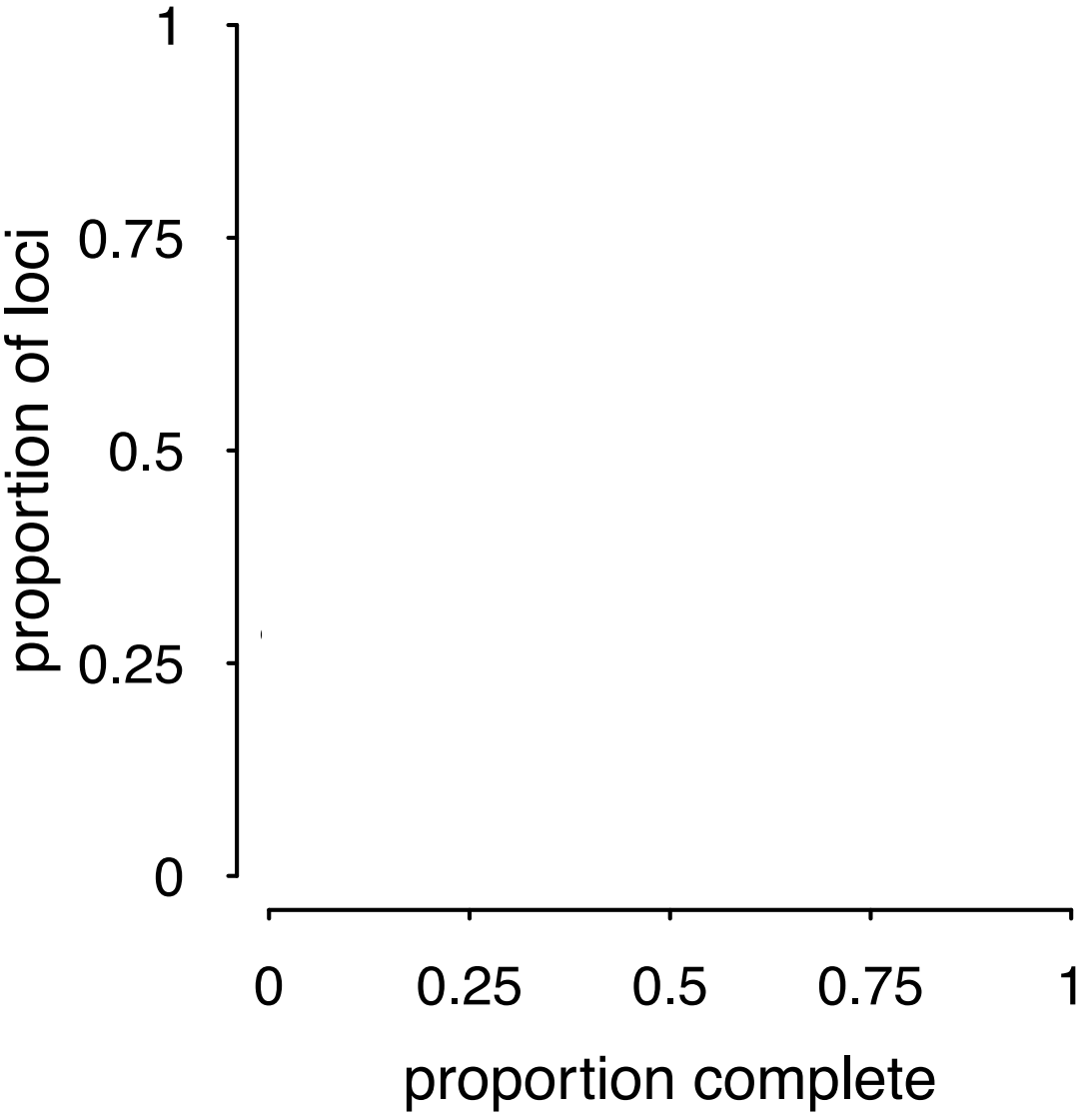
We filtered & refined the markers based on available squamate genomes.



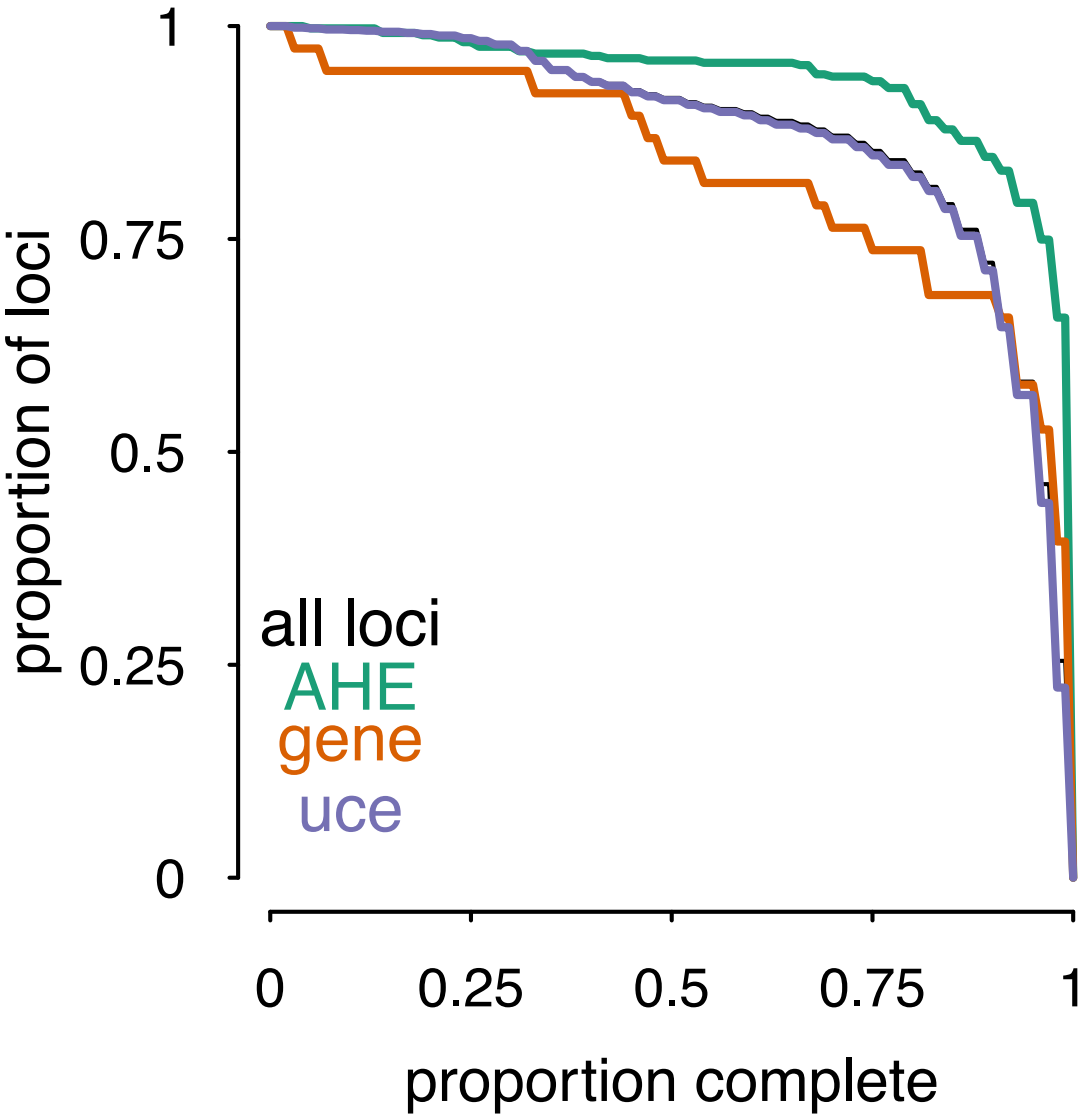
We tested this approach with Brazilian squamates.

wikimedia.org

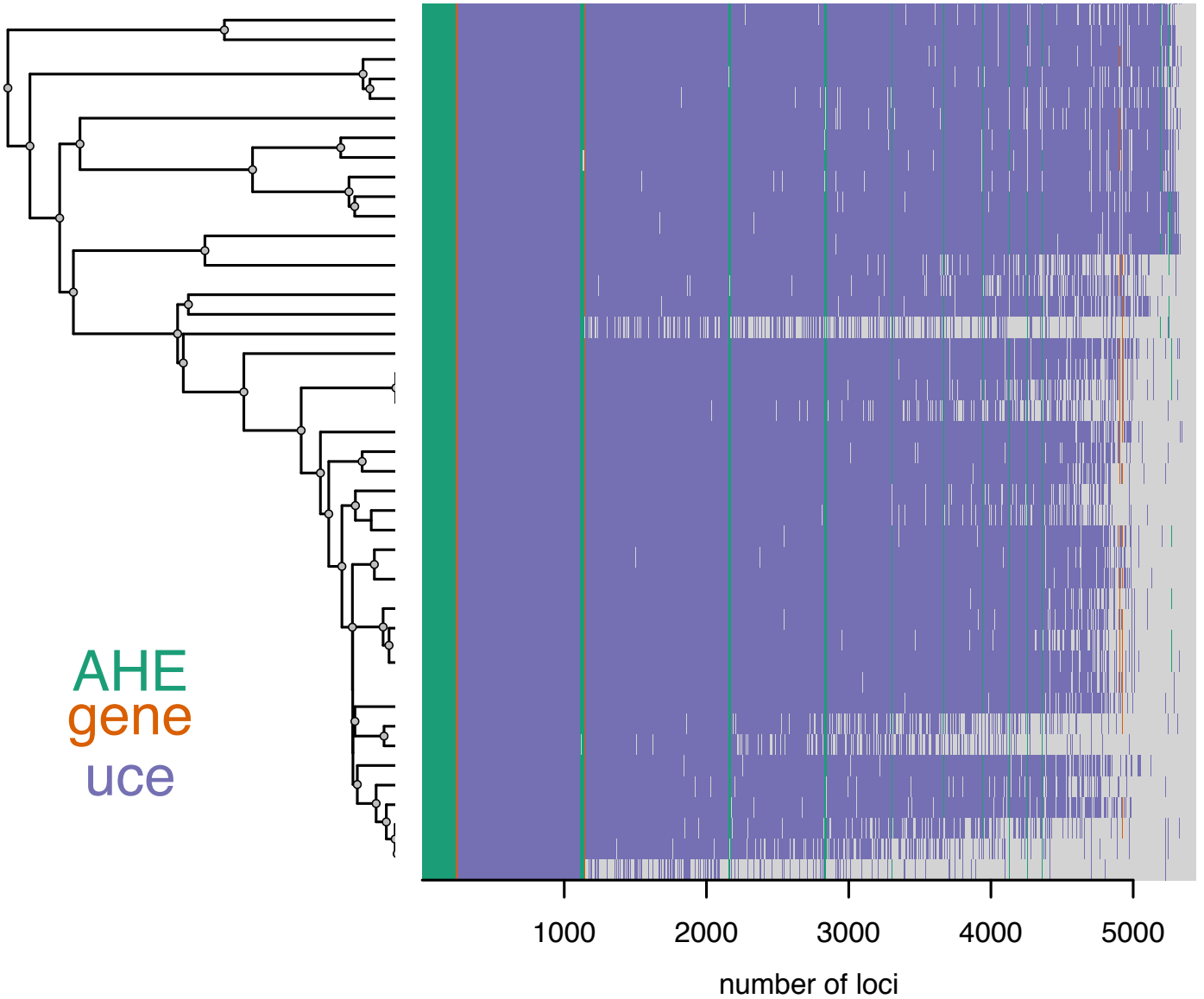
It works! (Quite well.)



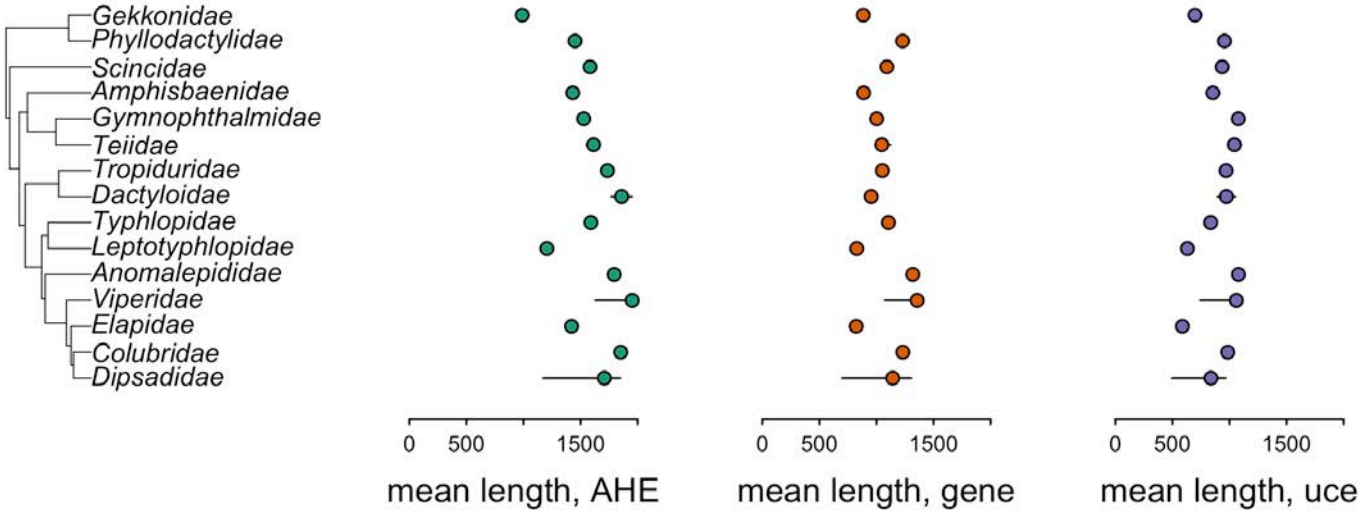
It works! (Quite well.)

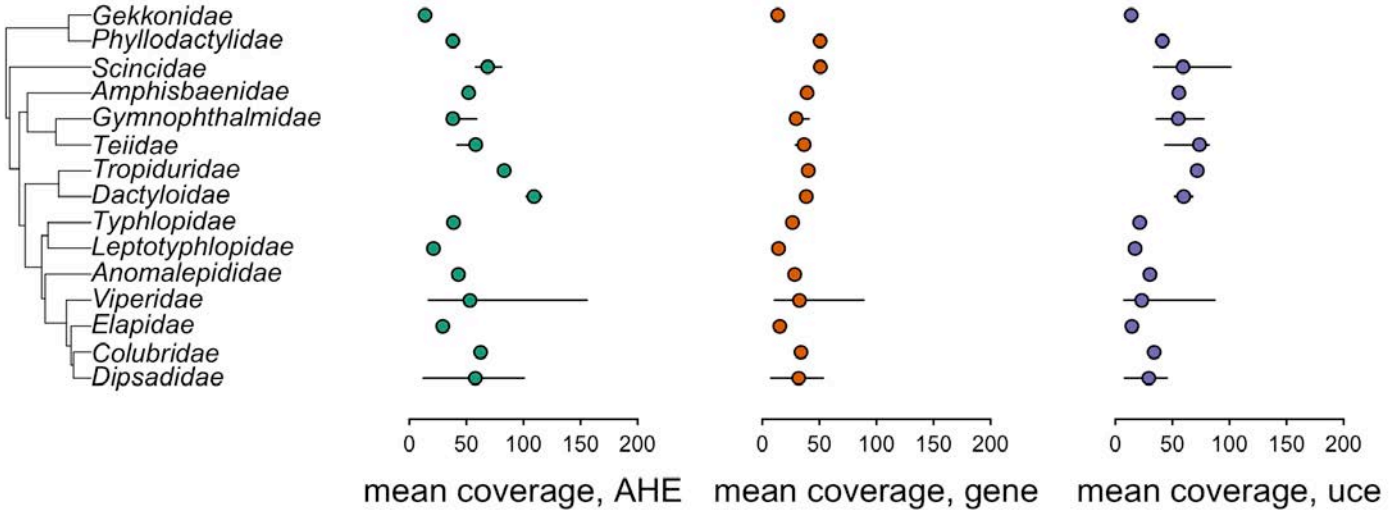


We generated a fairly complete data matrix.

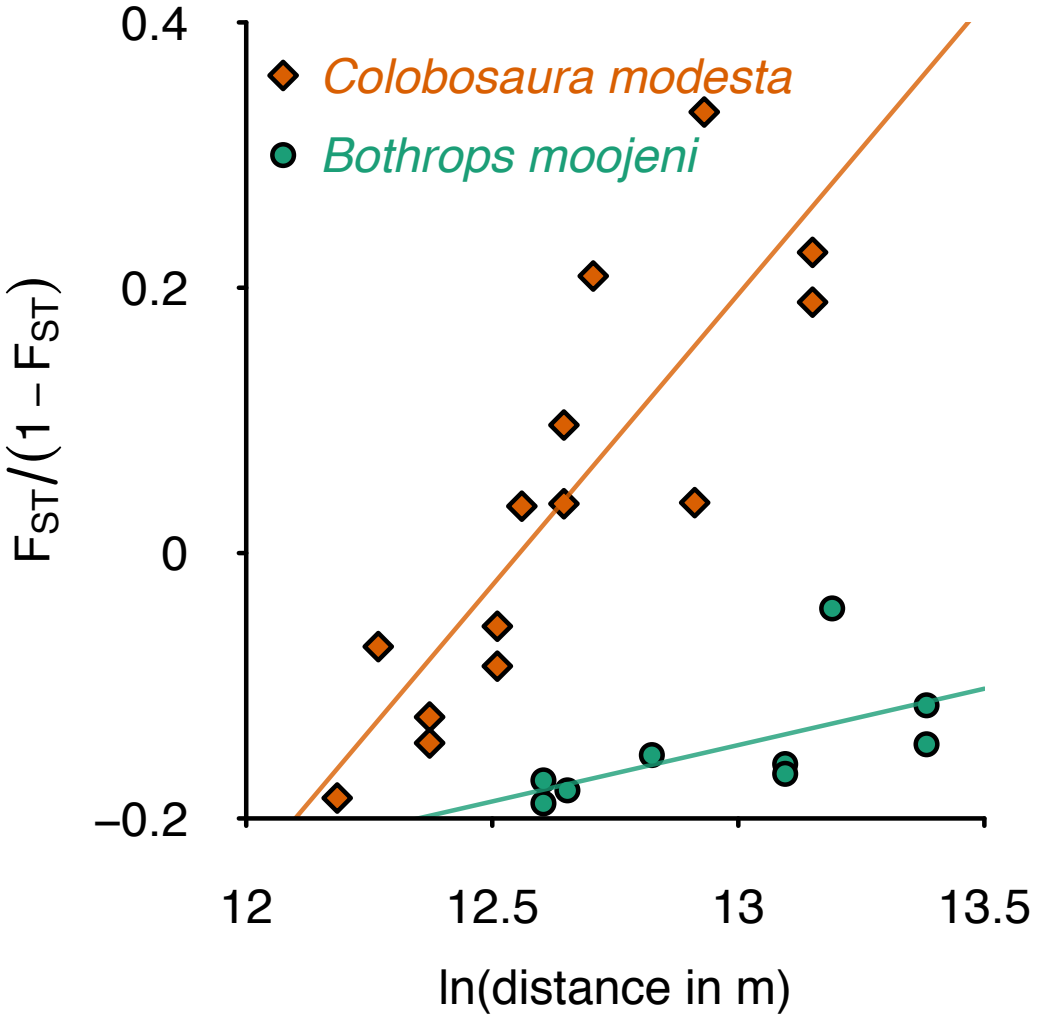


Loci are long & have high coverage.





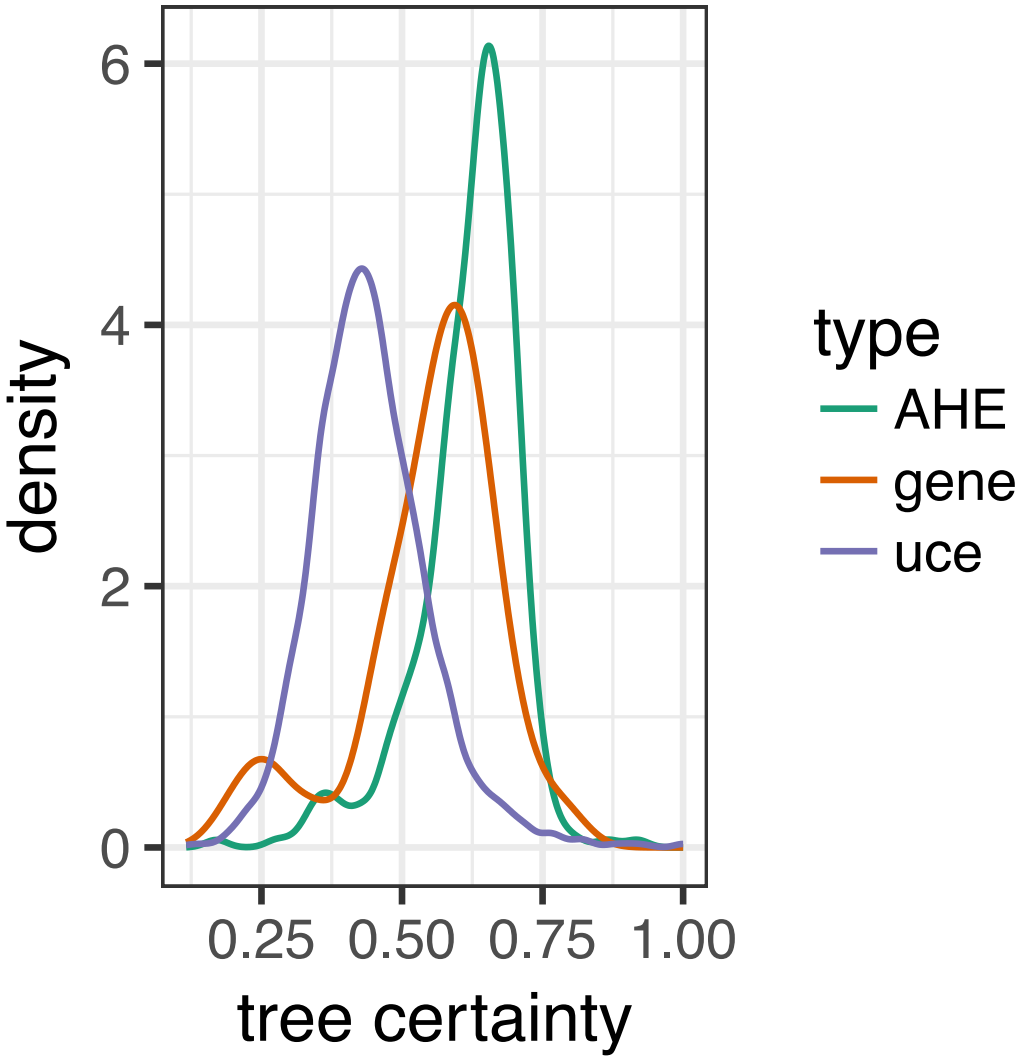
It is useful for population genetic applications, as well.



wikimedia.org

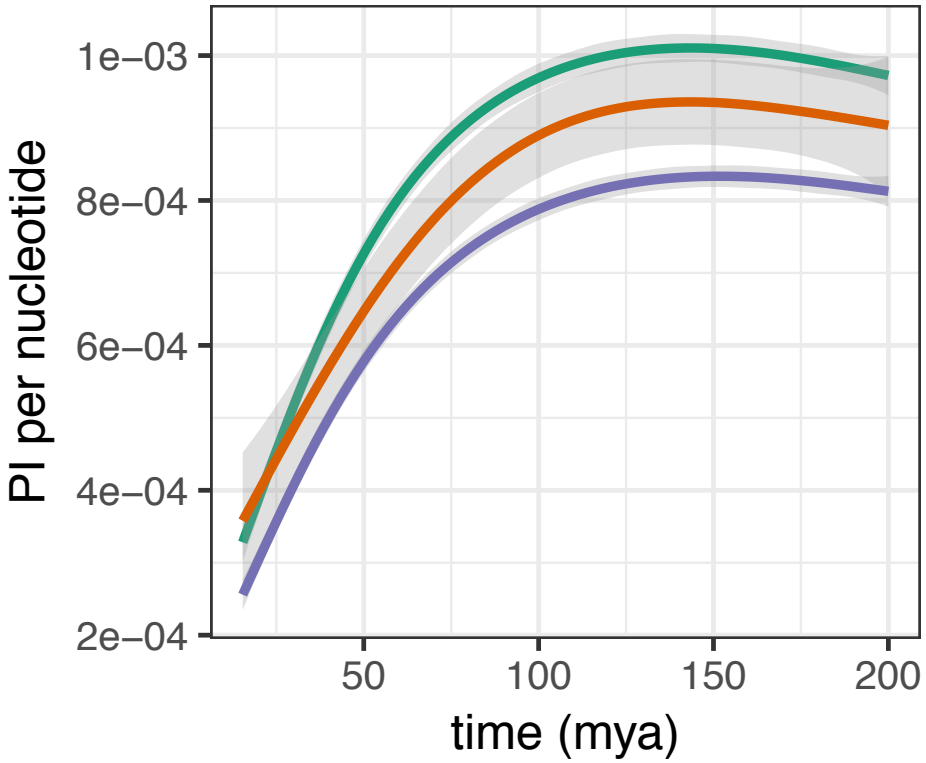
To be clear, most evolutionary genetic applications don’t require thousands of markers.

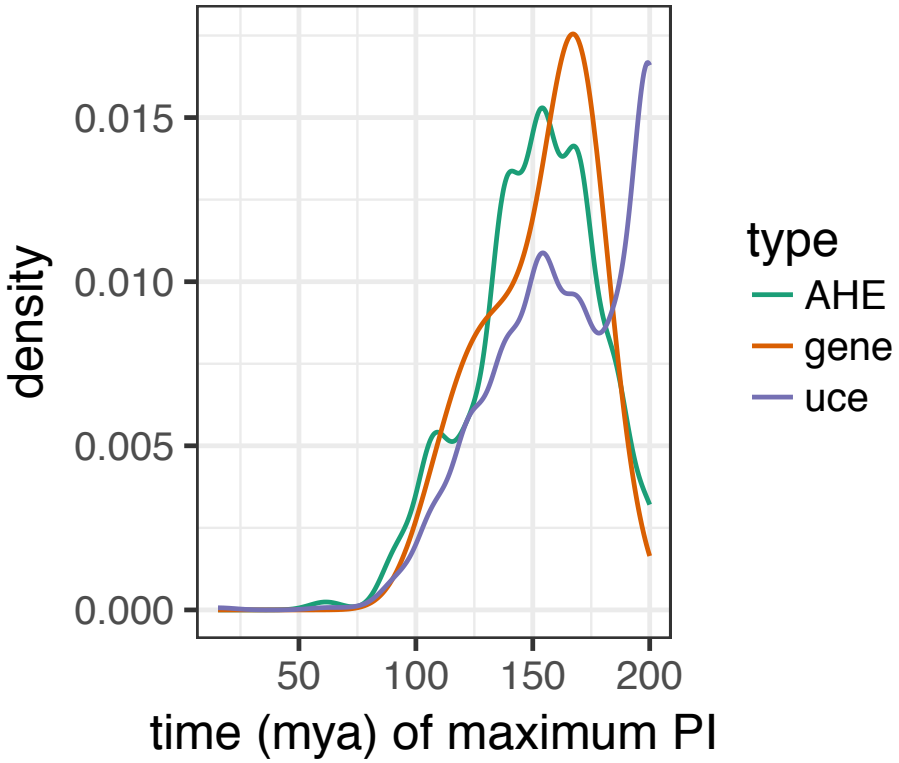
More markers, more filtering.



Salichos and Rokas 2013;
Blom et al 2016

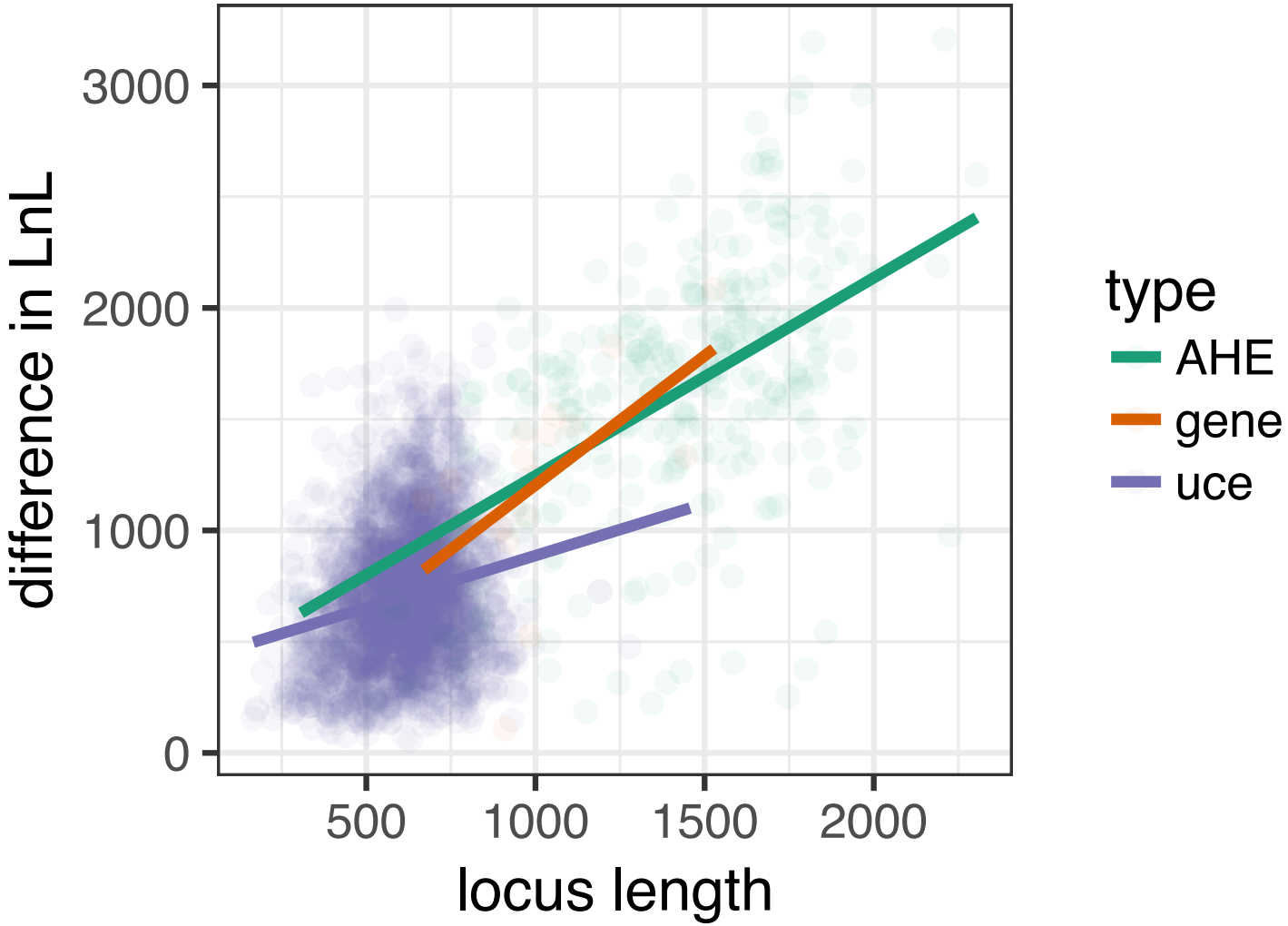
More markers, more filtering.





Townsend 2007;
Gilbert et al 2015

More markers, more filtering (this figure is useless!).



Doyle et al 2015;
Smith et al 2017

So why not take an inclusive approach?

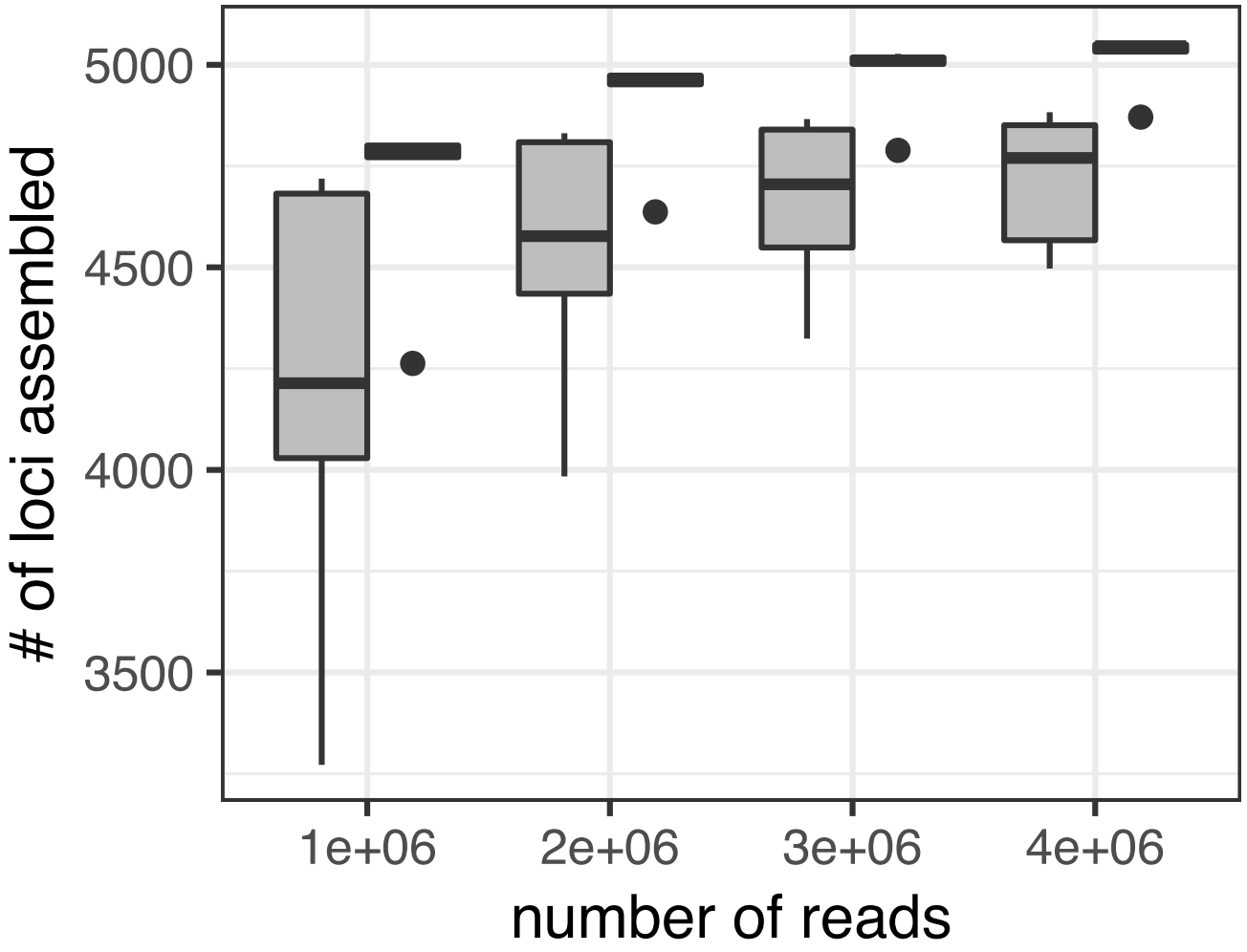
money

time

wikimedia.org

It took no extra time, and it increased the cost per sample 6%.

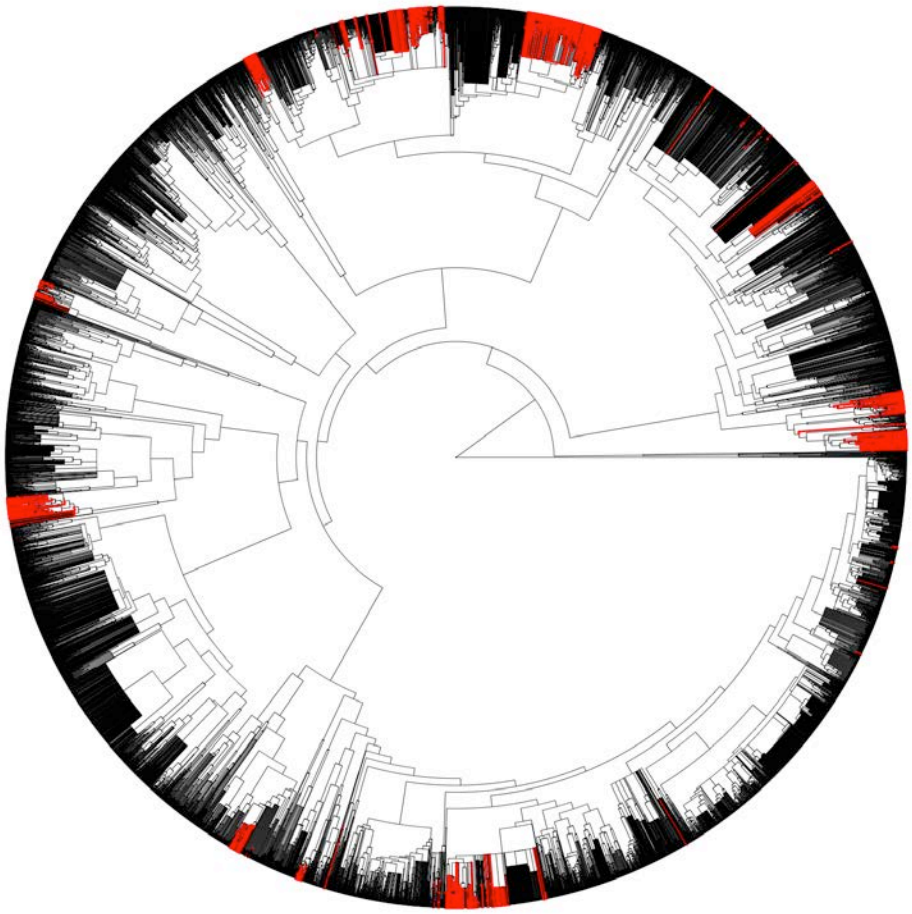
And, there are more cost-savings to be had.



This is extensible to other vertebrate systems.

2446 UCEs
71 AHEs
9 "traditional" genes

We are now using it to build a well-dated phylogeny for Australian squamates.



Tonini et al 2016

Do Macrophylogenies Yield Stable Macroevolutionary Inferences? An Example from Squamate Reptiles

Pascal O. Title; Daniel L. Rabosky

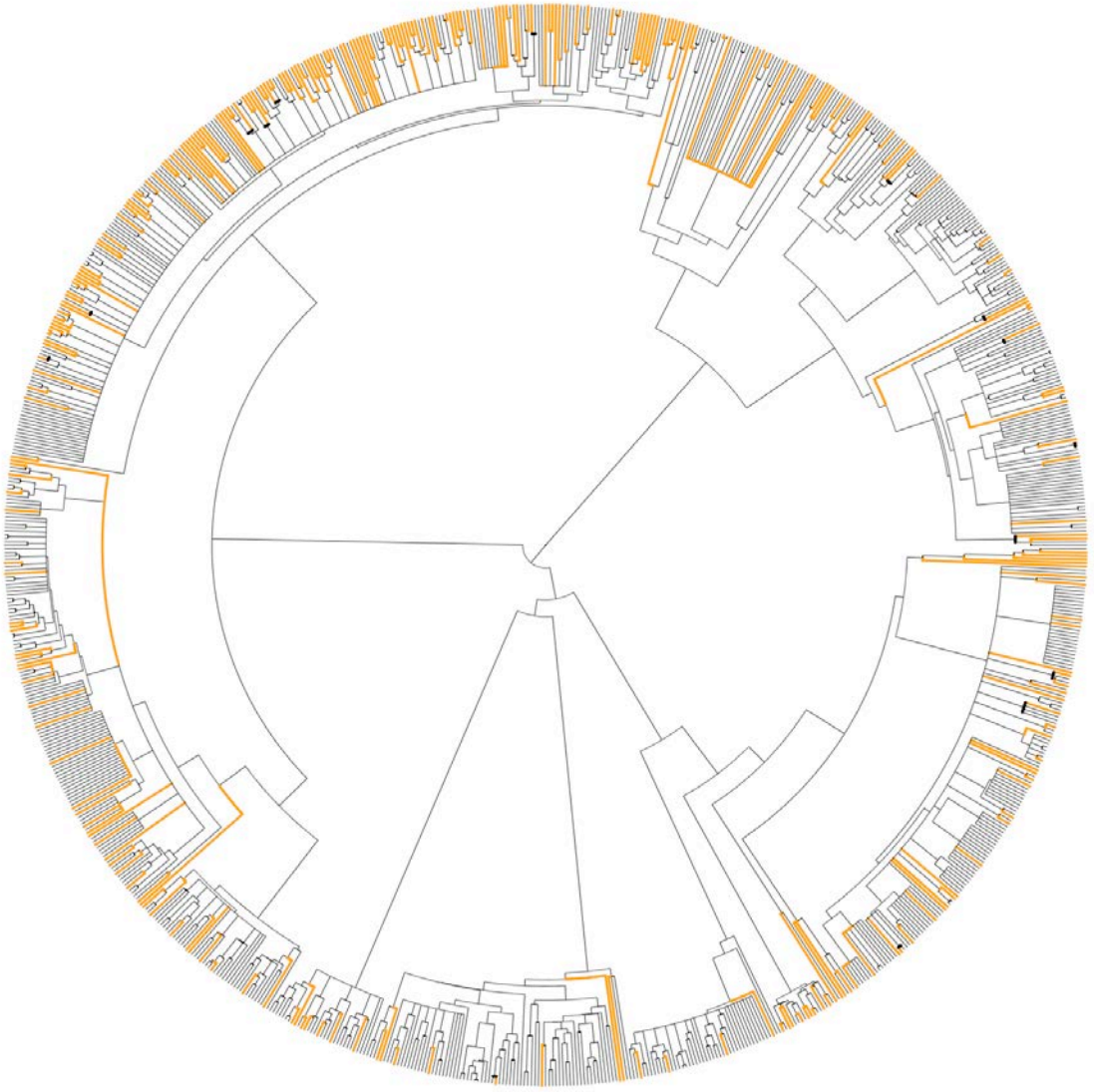
Do Macrophylogenies Yield Stable Macroevolutionary Inferences? An Example from Squamate Reptiles

Pascal O. Title; Daniel L. Rabosky

No.

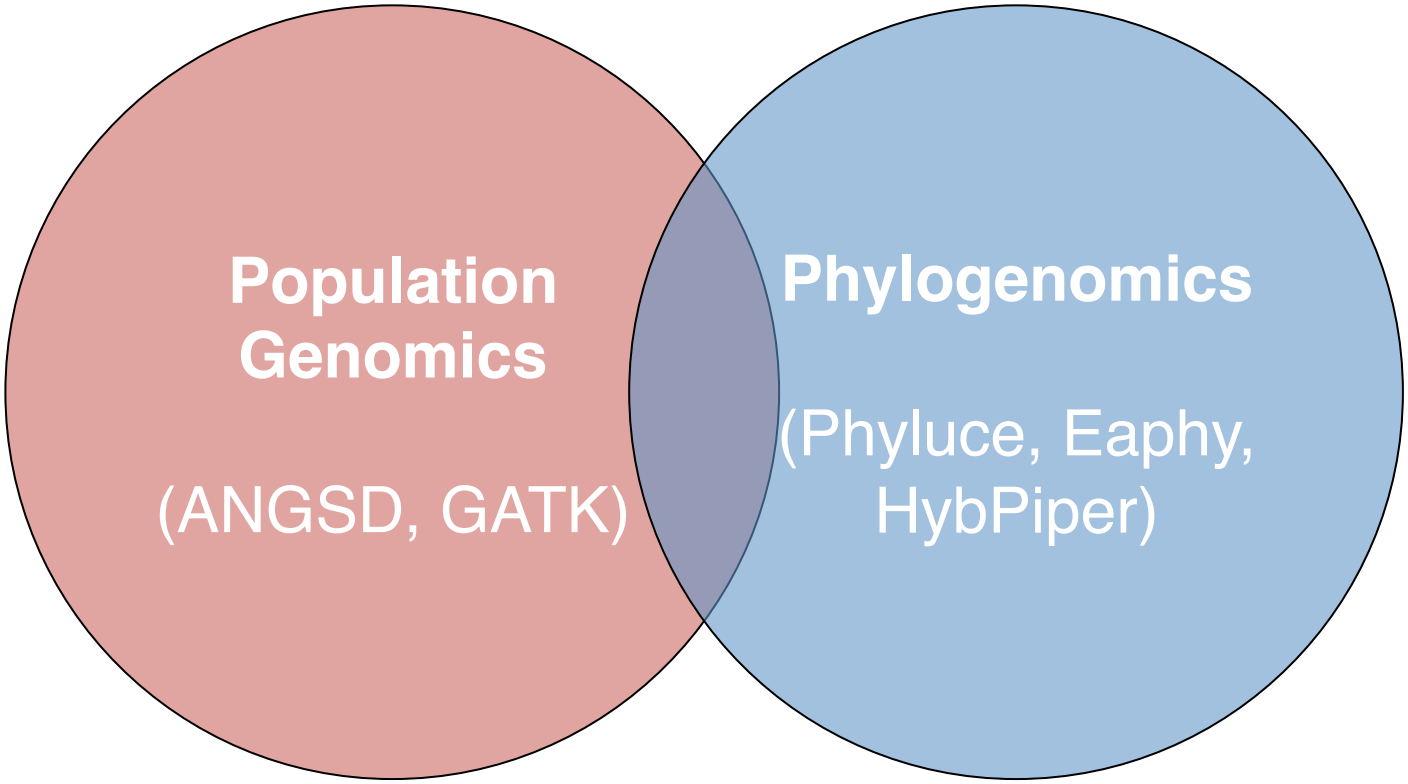
(Short answer.)

We are sampling about ~30% of Oz squamates.

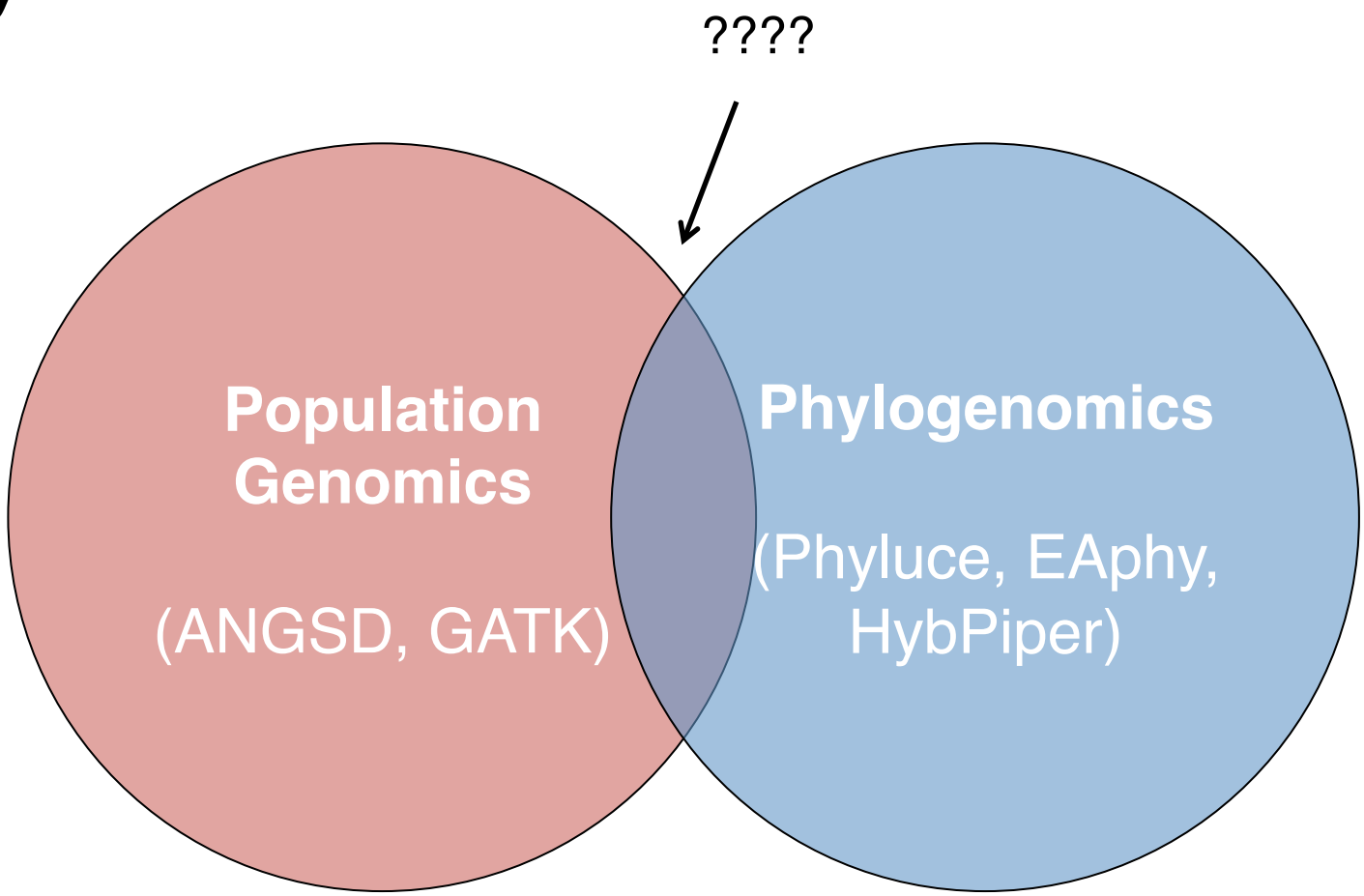


Once we get the data, how do we analyze it?

Once we get the data, how do we analyze it?



Once we get the data, how do we analyze it?



Challenges with the interface

- Phasing & recombination

Challenges with the interface

- Phasing & recombination
- Reference bias

Challenges with the interface

- Phasing & recombination
- Reference bias
- Multi-species SNP data sets

Challenges with the interface

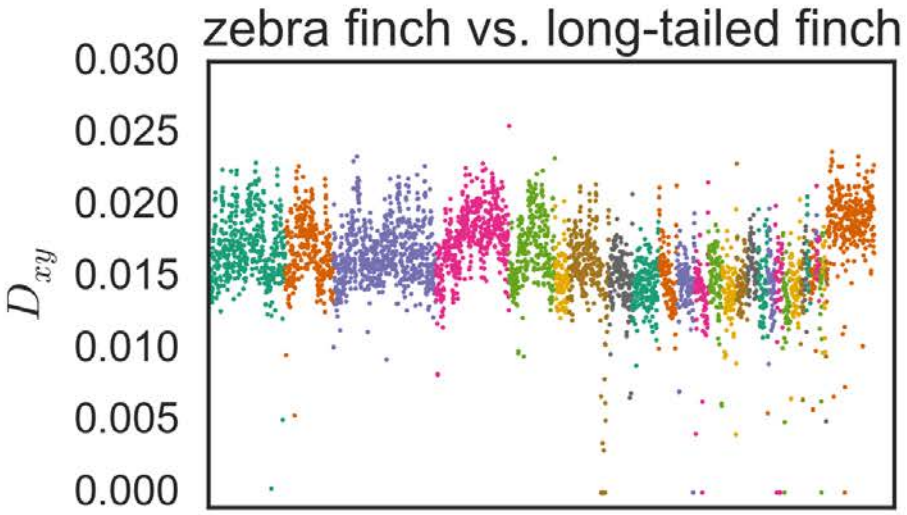
- Phasing & recombination
- Reference bias
- Multi-species SNP data sets
- Allowing for incomplete lineage sorting

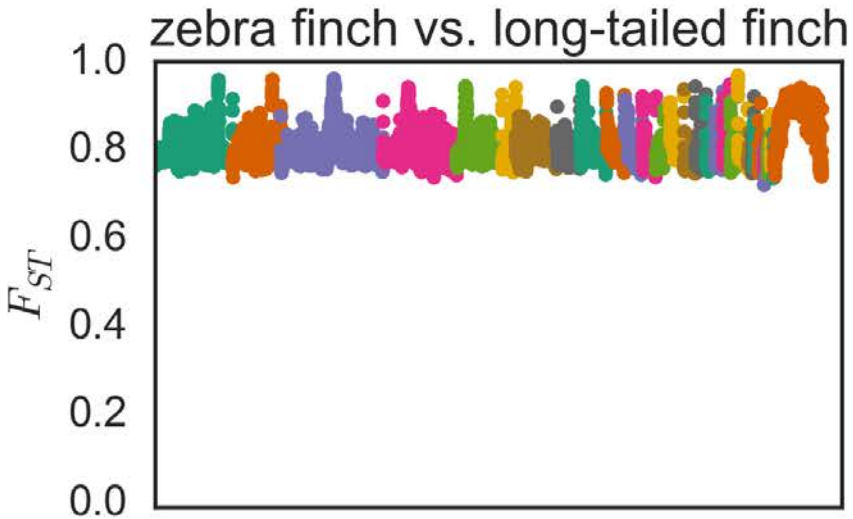
Challenges with the interface

- Phasing & recombination
- Reference bias
- Multi-species SNP data sets
- Allowing for incomplete lineage sorting
- Subsampling data for full-likelihood programs

And a persistent challenge across all these methods …

Plotting & visualizing data is non-trivial.





Things we might want to plot

- Short read alignments, coverage, SNP quality

Things we might want to plot

- Short read alignments, coverage, SNP quality
- Diversity, differentiation, and linkage disequilibrium across the genome

Things we might want to plot

- Short read alignments, coverage, SNP quality
- Diversity, differentiation, and linkage disequilibrium across the genome
- Multi-species alignments

Things we might want to plot

- Short read alignments, coverage, SNP quality
- Diversity, differentiation, and linkage disequilibrium across the genome
- Multi-species alignments
- Variation in gene trees

How can we improve these methods to allow researchers more intuition over their own data?

Collaborators & Funding

Pascal Title

Daniel Rabosky

Maggie Grundler

the David & Lucile Packard FOUNDATION