

Phylogenetic Biology

Week 7

Biology 1425

Professor: Casey Dunn, dunnlab.org

Brown University

2013

Front matter...

All original content in this document is distributed under the following license:



Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License
(http://creativecommons.org/licenses/by-nc-sa/3.0/deed.en_US)

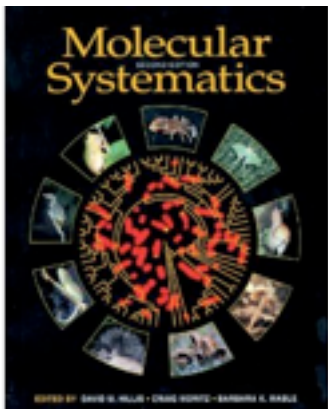
See sources for copyright of non-original content

Sources

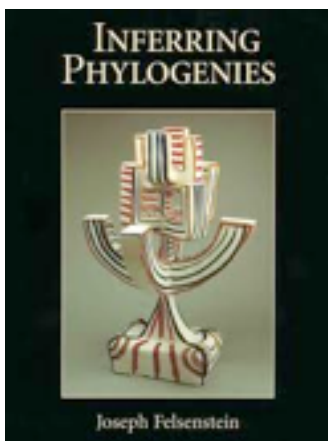
Some non-original content is drawn from:



Baum, D and S. Smith (2012) Tree Thinking: and Introduction to Phylogenetic Biology. Roberts and Company Publishers. ISBN 9781936221165



Swofford, D. L., Olsen, G. J., Waddell, P. J., & Hillis, D. M. (1996). Phylogenetic inference. In: Molecular Systematics, Second Edition. eds: D. M. Hillis, C Moritz, & B. K. Mable. Sinauer Associates. ISBN 9780878932825



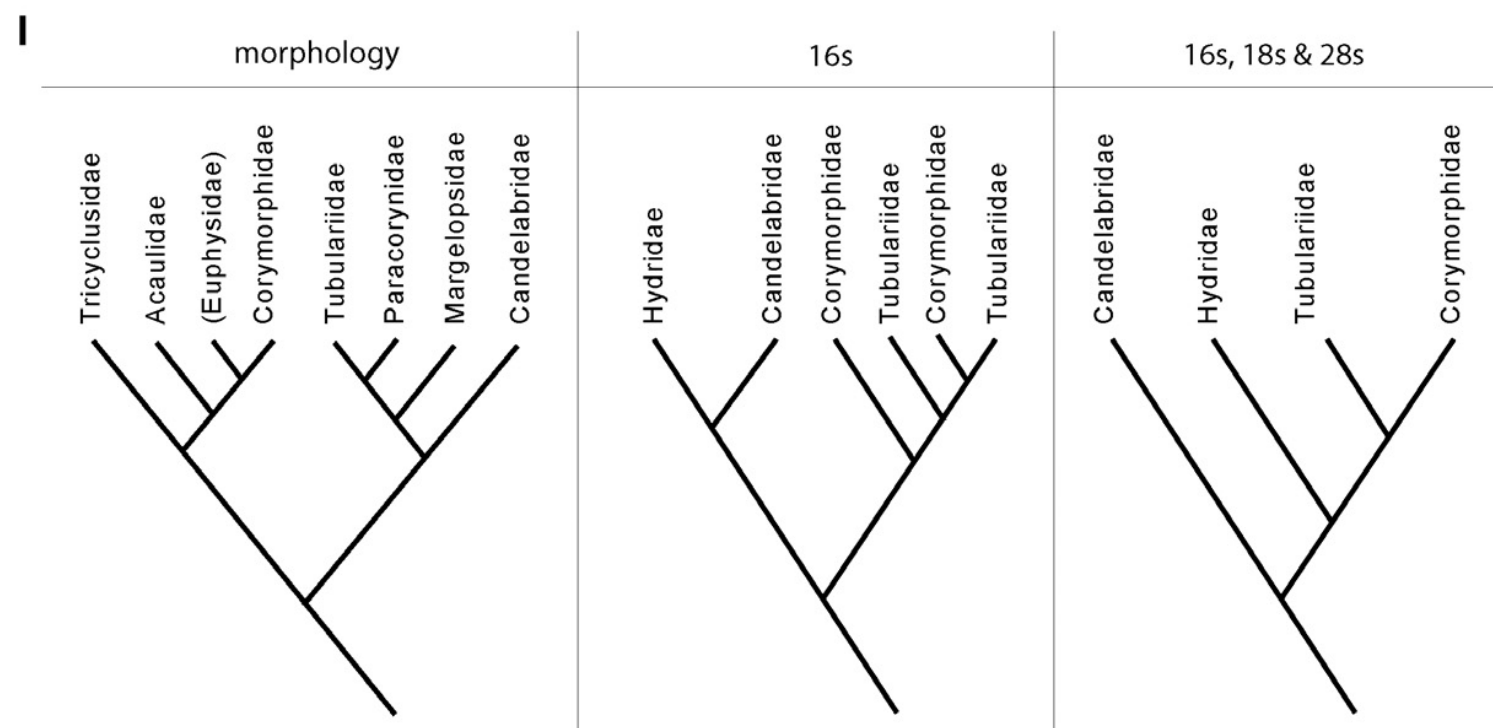
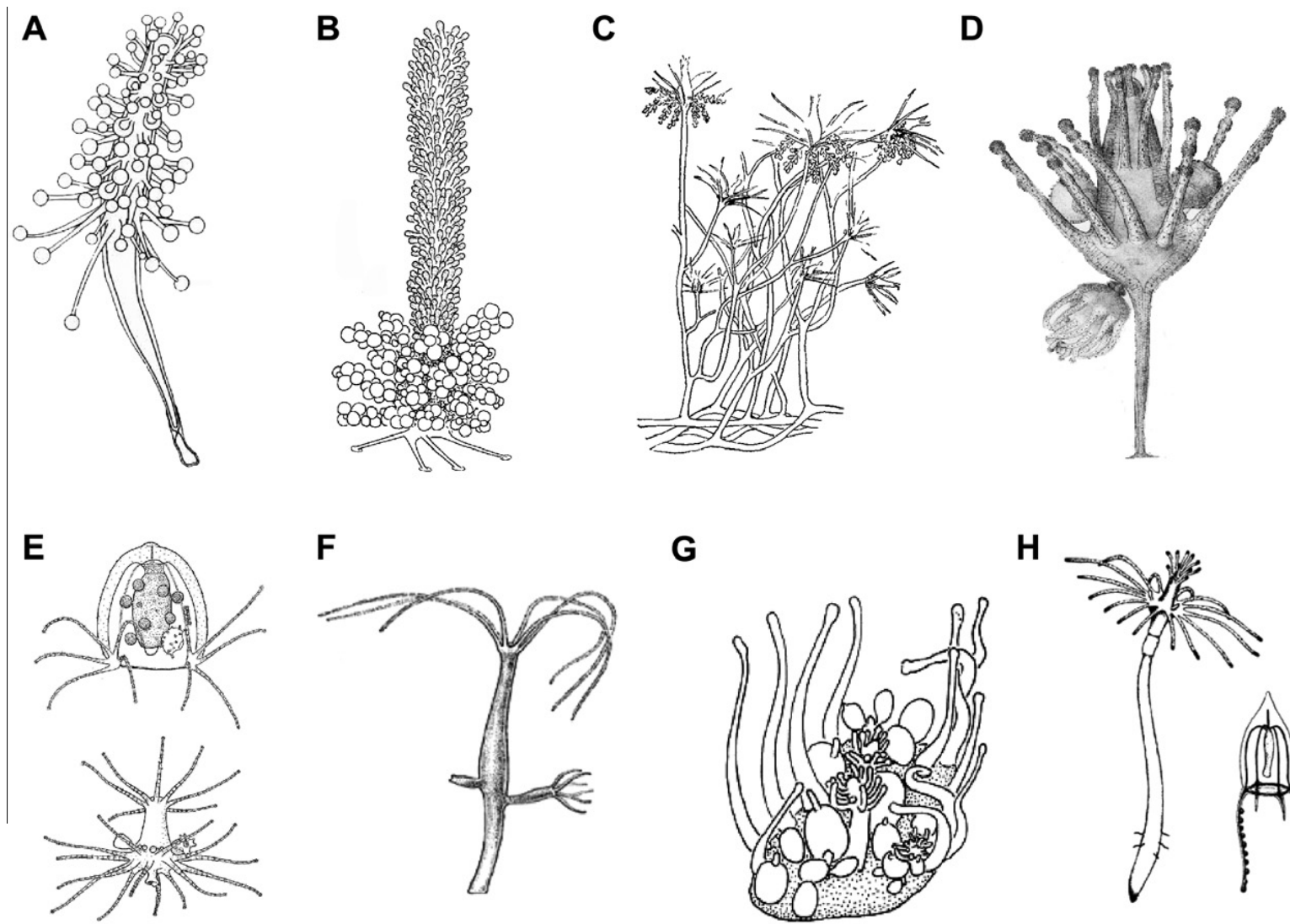
Felsenstein, J. (2003) Inferring Phylogenies. Sinauer Associates. ISBN 978-0878931774

Other non-original content is referenced by url.

**Phylogenetics in the wild-
Project design and the
challenges that arise when
executing a study**

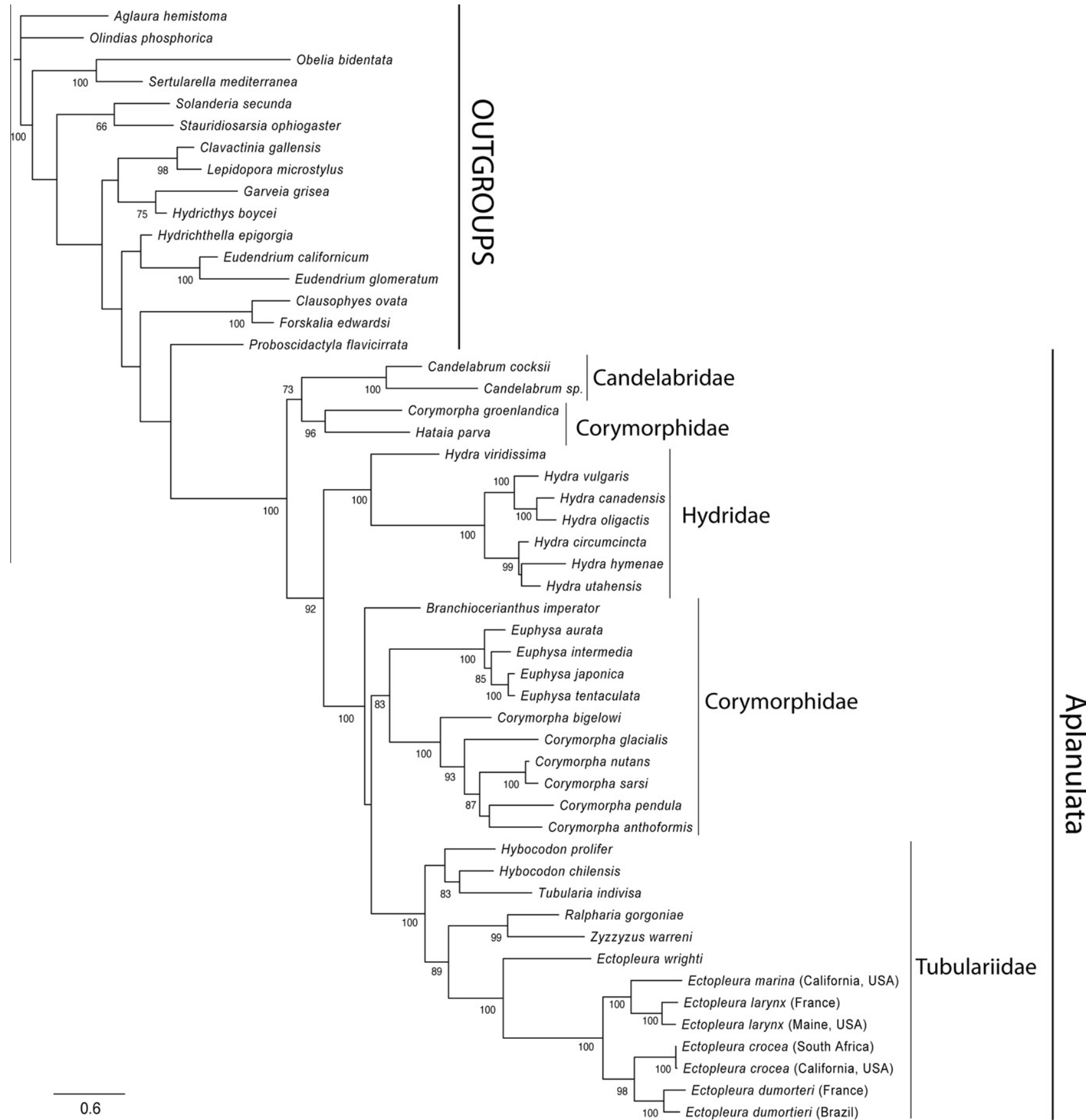
An example study

Nawrocki, A. M., Collins, A. G., Hirano, Y. M., Schuchert, P. & Cartwright, P. **Phylogenetic placement of *Hydra* and relationships within Aplanulata (Cnidaria: Hydrozoa)**. *Molecular Phylogenetics and Evolution* 67, 60–71 (2013). <http://dx.doi.org/10.1016/j.ympev.2012.12.016>



Nawrocki et al 2013,
Figure 1

Bootstrap, all markers



Nawrocki et al 2013,
Figure 3

**How important is taxon
sampling?**

How important is taxon sampling?

The better your taxon sampling, the more questions you can address.

How important is taxon sampling?

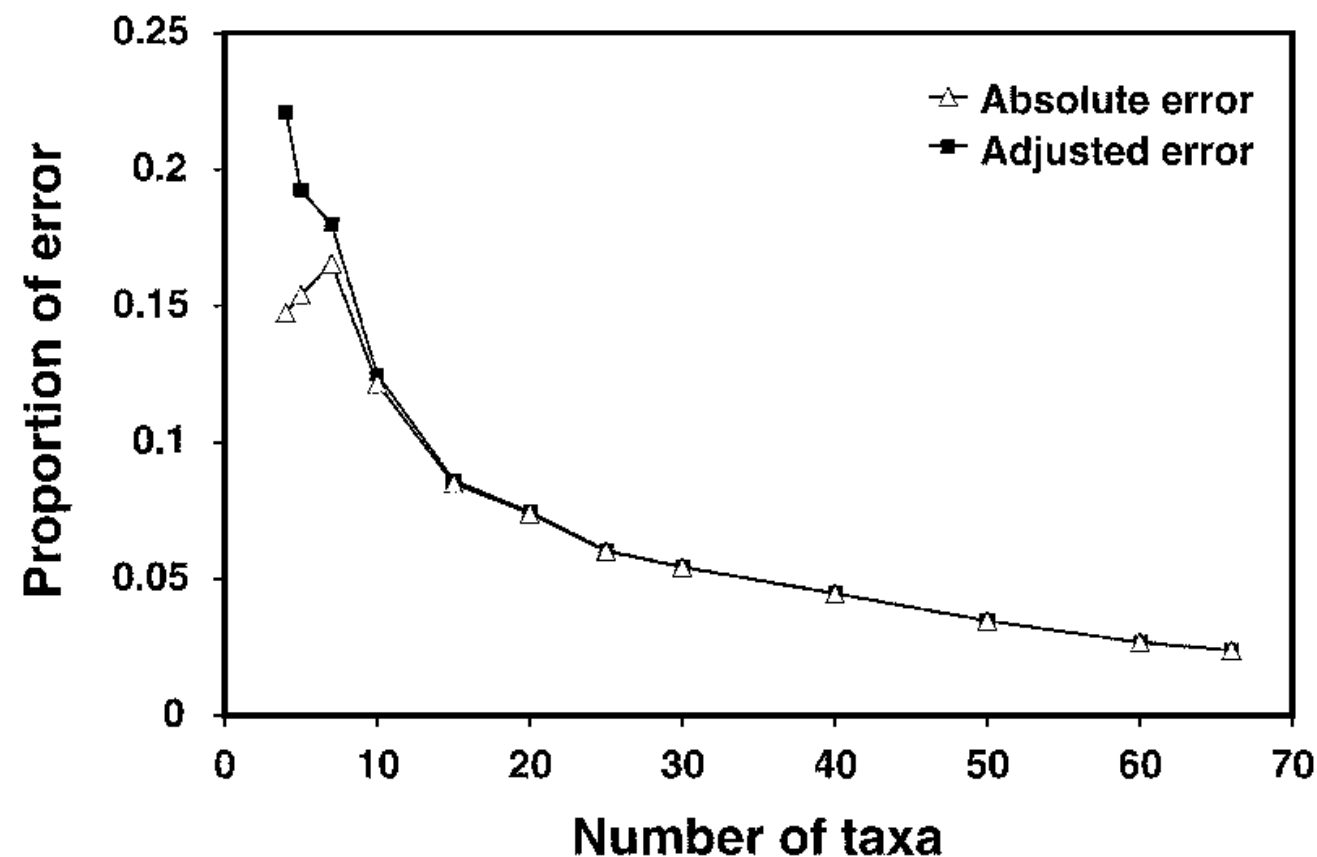
The better your taxon sampling, the more questions you can address.

Better taxon sampling can improve accuracy

Increased Taxon Sampling Greatly Reduces Phylogenetic Error

DERRICK J. ZWICKL AND DAVID M. HILLIS

*Section of Integrative Biology and Center for Computational Biology and Bioinformatics, University of Texas,
Austin, Texas 78712, USA; E-mail: zwickl@mail.utexas.edu and dhillis@mail.utexas.edu*



Zwickl and Hillis 2002,
<http://dx.doi.org/10.1080/10635150290102339>

**How important are
outgroups?**

How important are outgroups?

Poor outgroup sampling is one of the biggest rookie mistakes in phylogenetic analyses

**How important is character
sampling?**

How important is character sampling?

Better character sampling can
improve accuracy

How important is character sampling?

Better character sampling can improve accuracy

Better character sampling can provide more rigorous evaluation of hypotheses

**How does missing data
impact phylogenetic
analyses?**

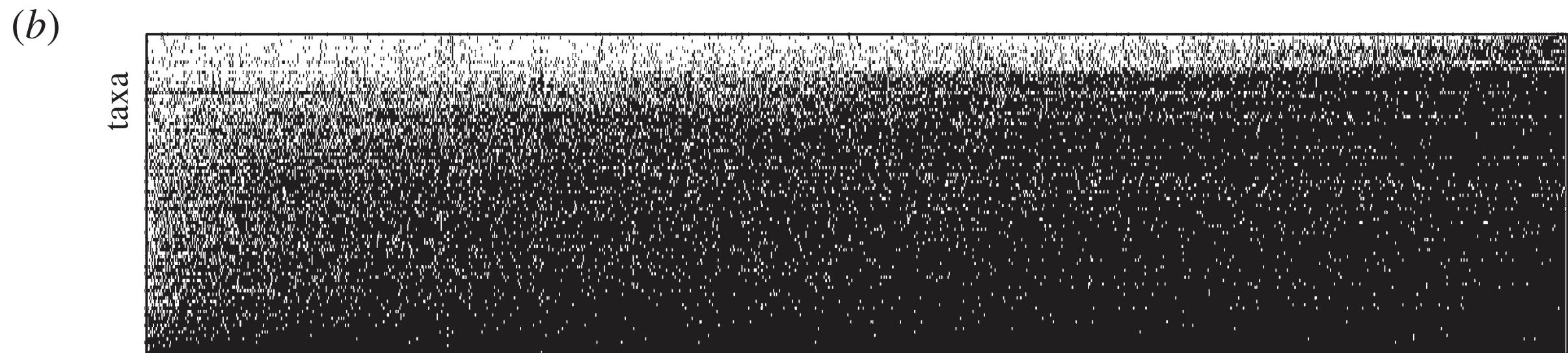
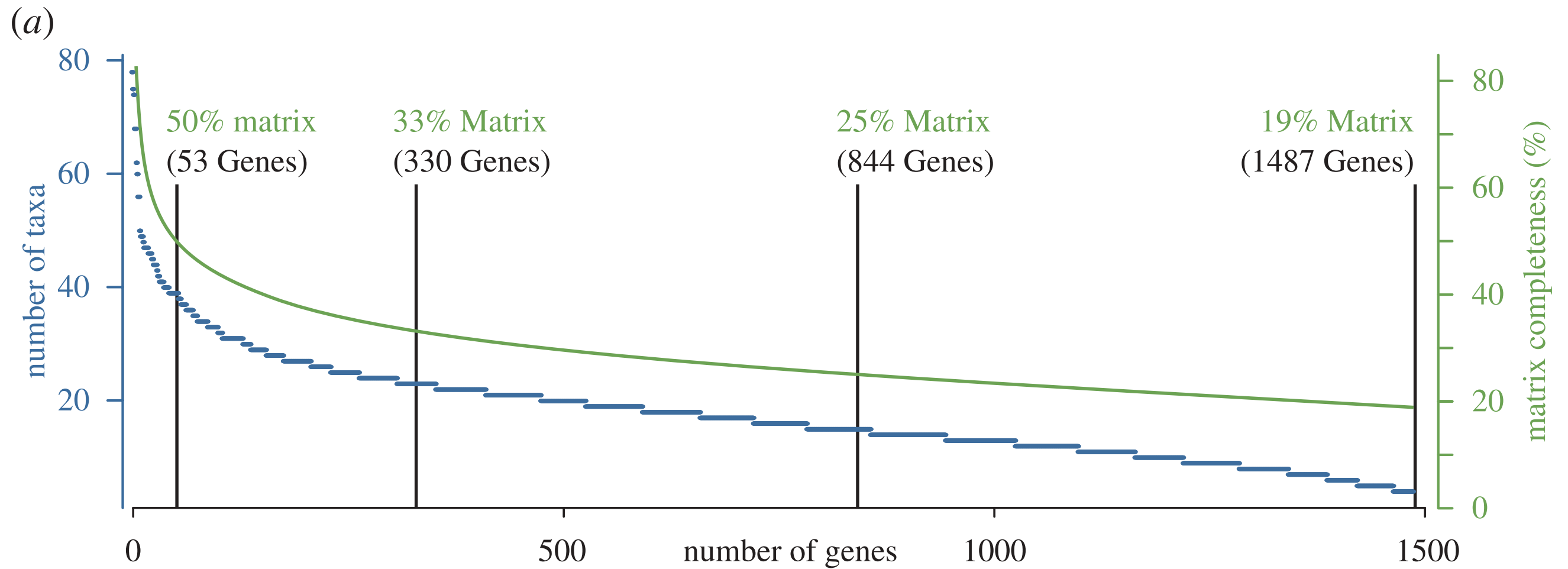
**Missing data leads to loss of
signal.**

Missing data leads to loss of signal.

Can it also lead to inconsistency?

Missing data in phylogenetic analyses are rarely distributed randomly.

Some genes are better sampled than others, some taxa are better sampled than others.



**Should you add poorly
sampled genes and taxa to
an analysis, or have a
smaller but more complete
analysis?**

Amount of missing data:

less	more
Eliminate poorly sampled taxa	Broader taxon sampling
Eliminate poorly sampled genes	Broader gene sampling
Additional expense	Reduced expense
	Could be impacted by inconsistency

Eliminate poorly
sampled taxa

Eliminate poorly
sampled genes

Additional
expense

more

Broader taxon
sampling

Broader gene
sampling

Reduced expense

Could be impacted
by inconsistency



Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

Journal of Biomedical Informatics 39 (2006) 34–42

Journal of
Biomedical
Informatics

www.elsevier.com/locate/yjbin

Methodological Review

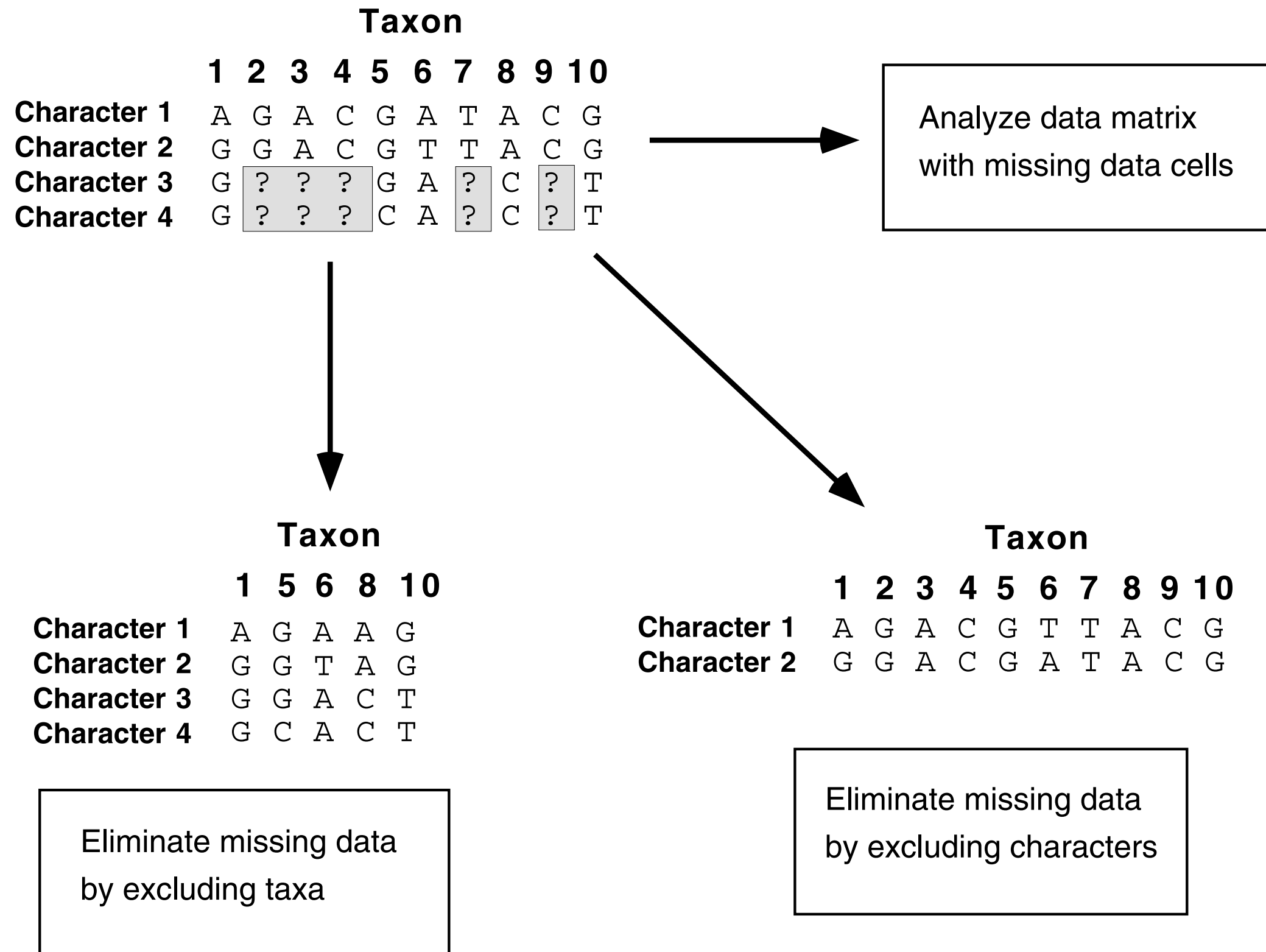
Missing data and the design of phylogenetic analyses

John J. Wiens *

Department of Ecology and Evolution, Stony Brook University, Stony Brook, NY 11794-5245, USA

Received 5 March 2005

Available online 6 May 2005



“Recent simulations suggest that highly incomplete taxa can be accurately placed in phylogenies, as long as many characters have been sampled overall. Furthermore, adding incomplete taxa can dramatically improve results in some cases by subdividing misleading long branches. Adding characters with missing data can also improve accuracy, although there is a risk of long-branch attraction in some cases.”

The Effect of Ambiguous Data on Phylogenetic Estimates Obtained by Maximum Likelihood and Bayesian Inference

ALAN R. LEMMON^{1,2,3,*}, JEREMY M. BROWN¹, KATHRIN STANGER-HALL⁴, AND EMILY MORIARTY LEMMON^{1,3}

¹*Section of Integrative Biology, University of Texas at Austin, 1 University Station C0930, Austin, TX 78712, USA;*

²*Present address: Department of Scientific Computing, Florida State University, Dirac Science Library, Tallahassee, FL 32306-4120, USA;*

³*Present address: Department of Biological Science, Florida State University, Tallahassee, FL 32306, USA;*

⁴*Plant Biology Department, University of Georgia, 403 Biosciences Building, Athens, GA 30602, USA;*

**Correspondence to be sent to: Department of Scientific Computing, Florida State University, Dirac Science Library, Tallahassee, FL 32306-4120, USA; E-mail: alemmon@evotutor.org.*

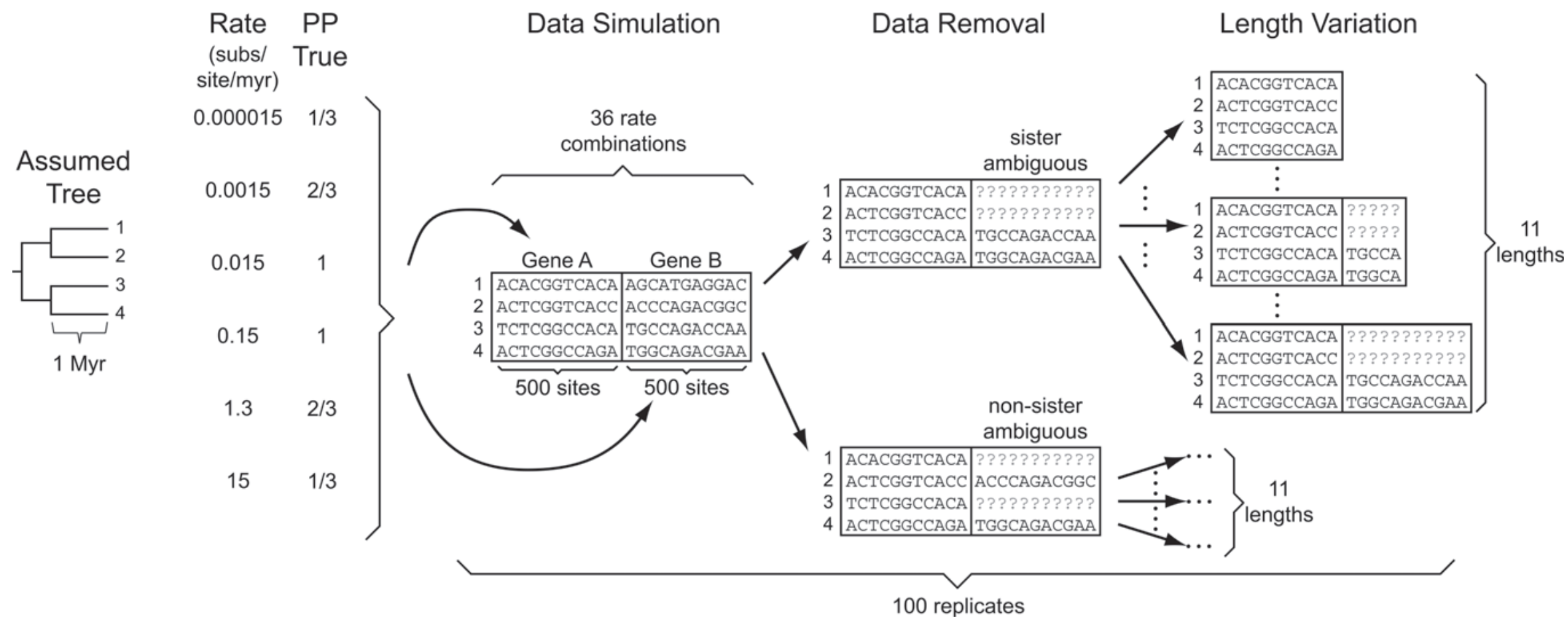


FIGURE 1. Simulation design. Among-site rate variation was simulated using 6 rates of evolution (chosen to produce the desired PP for the true tree with 500 sites) combined across 2 genes to form 36 rate combinations. Gene A contained unambiguous sites, whereas Gene B contained ambiguous sites. Ambiguous characters were present for either sister or nonsister taxa. Although Gene A always contained 500 sites, the length of Gene B varied from 0 to 500 sites. Note that Gene B contained no topological information, regardless of the rate of evolution. PP = posterior probabilities.

“We find that in both ML and Bayesian frameworks, among-site rate variation can interact with ambiguous data to produce misleading estimates of topology and branch lengths.”

Research article

Open Access

Using ESTs for phylogenomics: Can one accurately infer a phylogenetic tree from a gappy alignment?

Stefanie Hartmann^{1,2} and Todd J Vision^{* 1}

Address: ¹Department of Biology, University of North Carolina, Chapel Hill, NC 27599, USA and ²Institute for Biochemistry and Biology, Karl-Liebknecht-Strasse 24-25, University of Potsdam, 14476 Potsdam, Germany

Email: Stefanie Hartmann - stefanie.hartmann@uni-potsdam.de; Todd J Vision^{*} - tjv@bio.unc.edu

^{*} Corresponding author

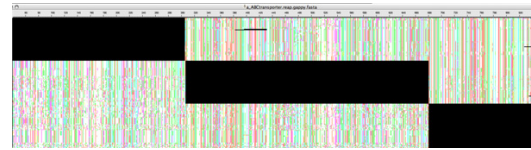
Published: 26 March 2008

BMC Evolutionary Biology 2008, **8**:95 doi:10.1186/1471-2148-8-95

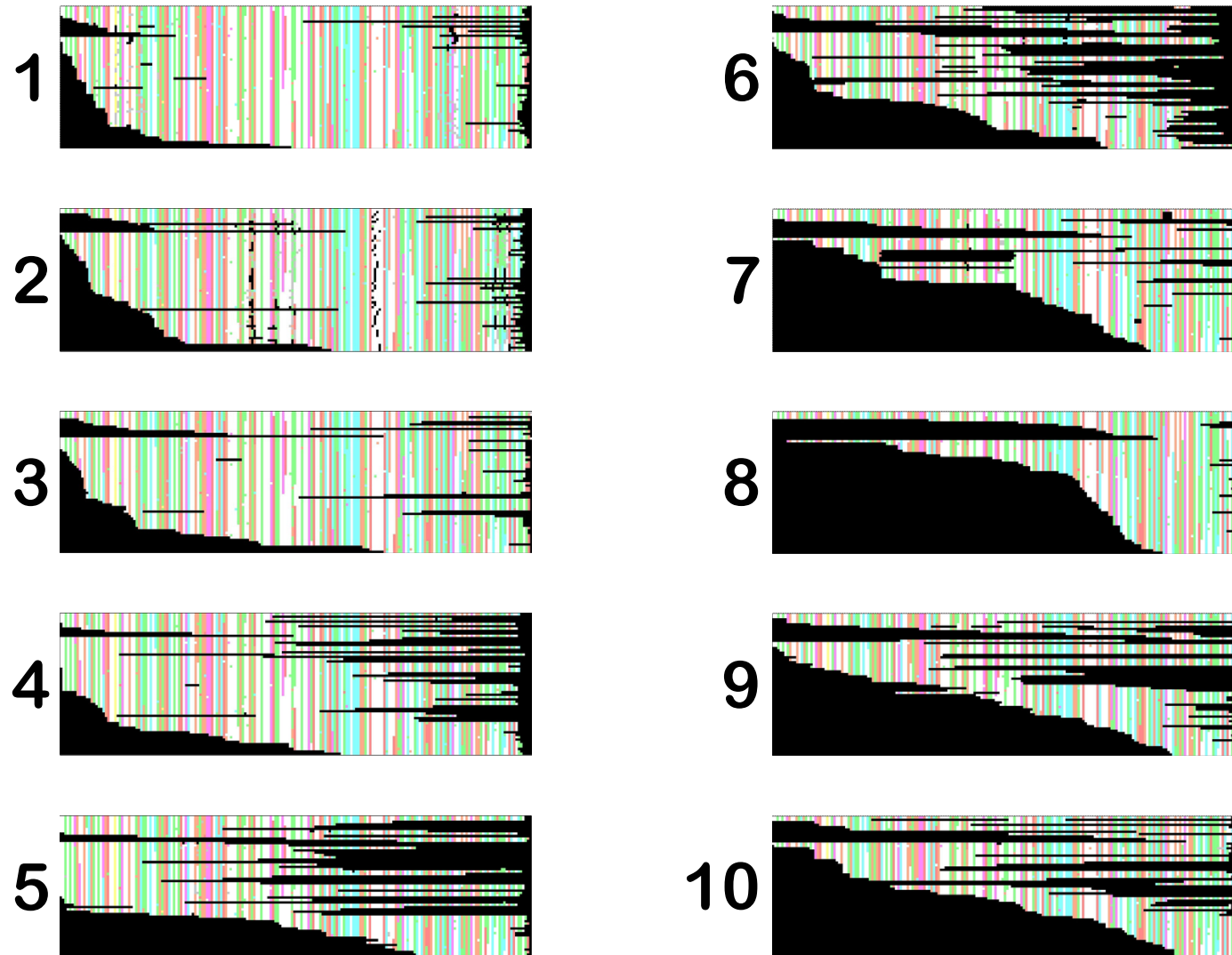
Received: 27 October 2007

Accepted: 26 March 2008

A

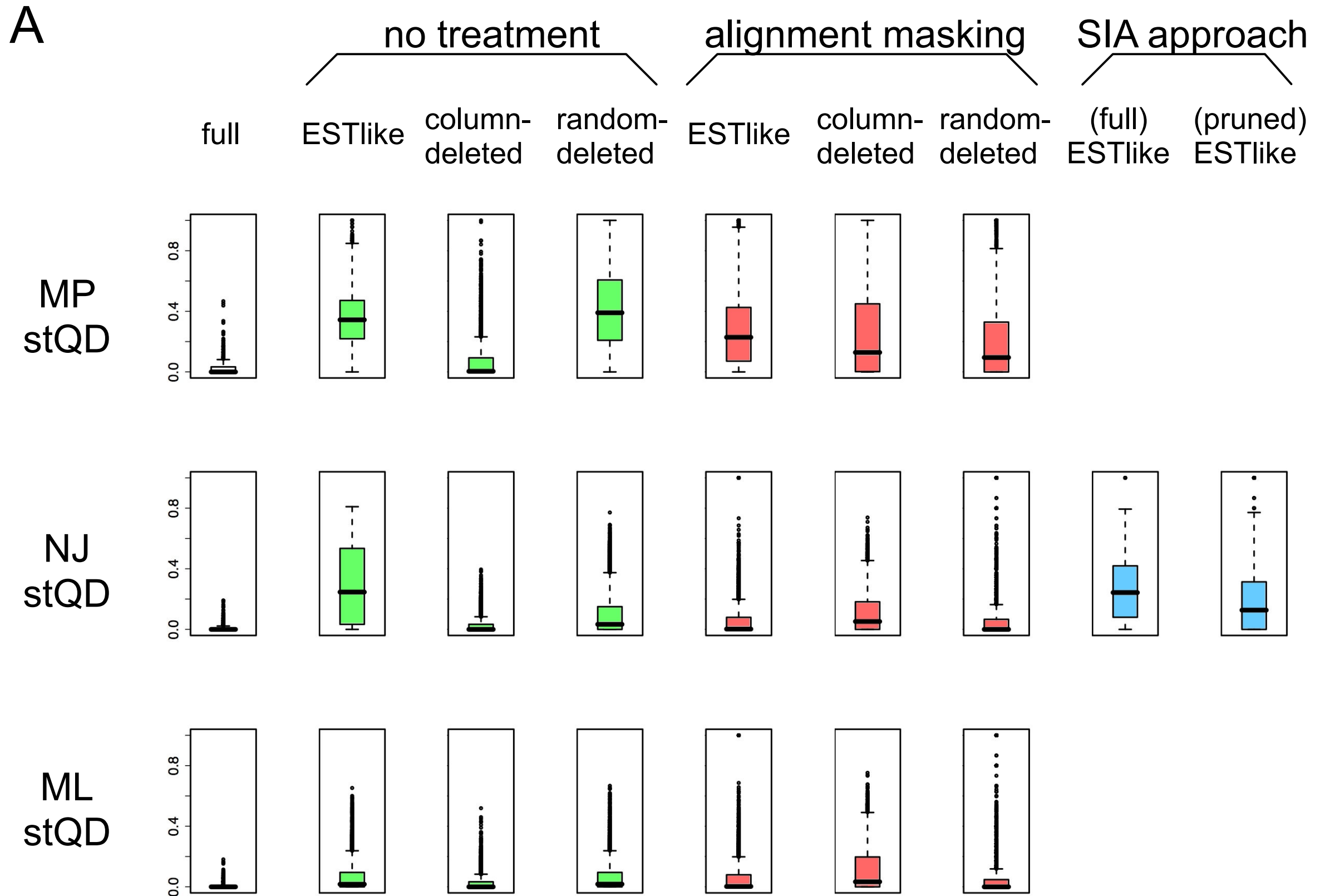


B



C



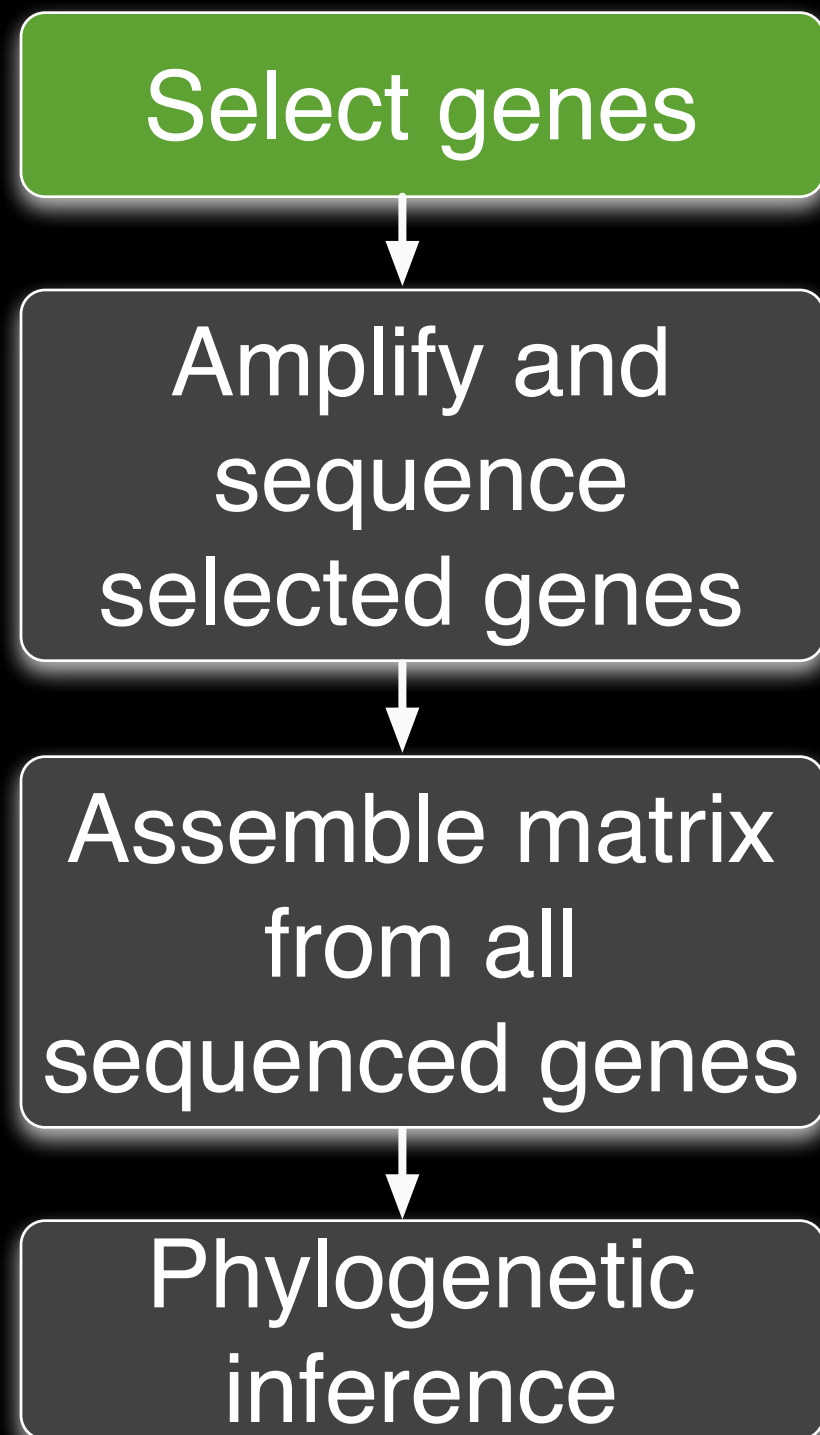


How to collect your molecular data?

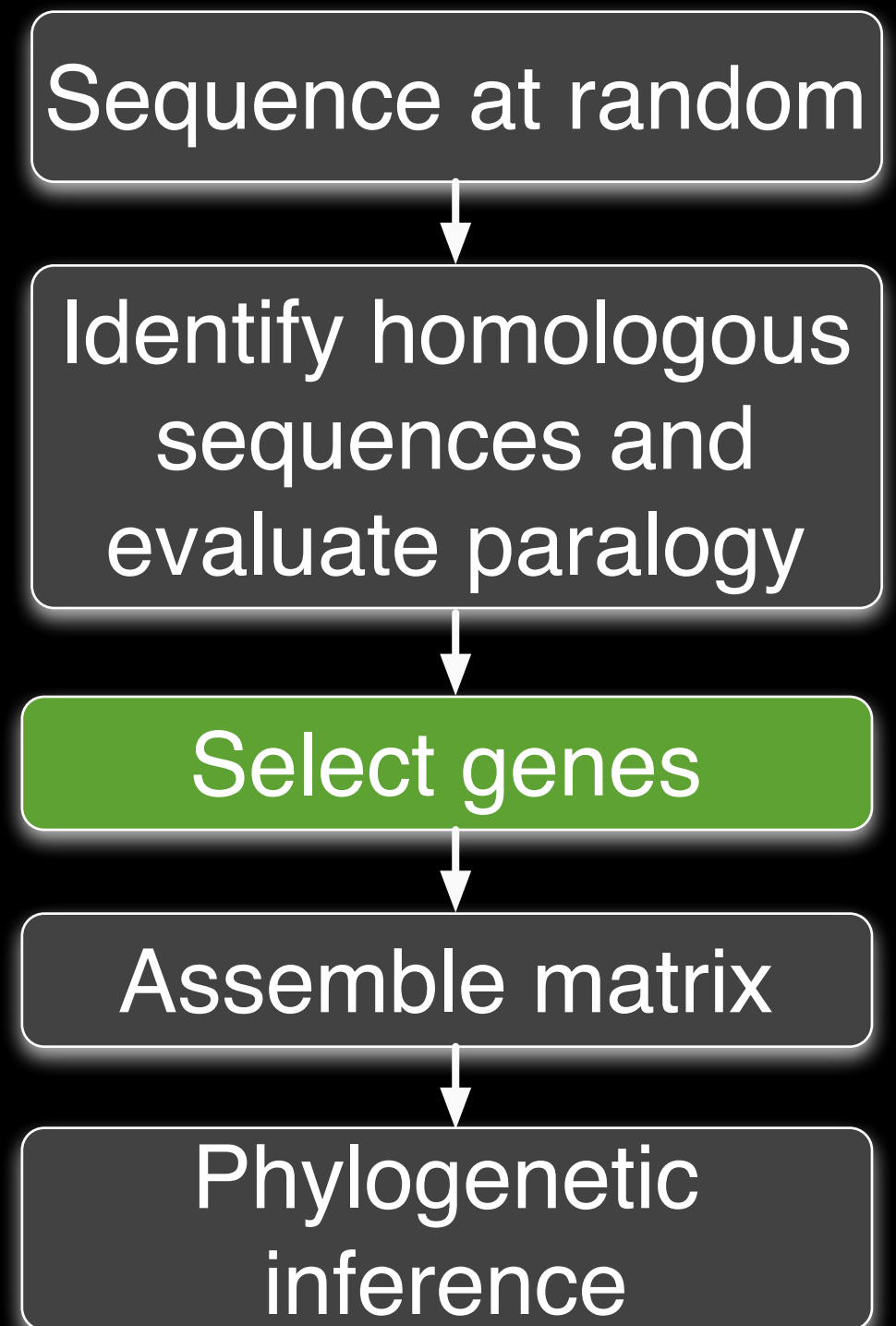
Directed sequencing - pick genes first, then sequence just them

Phylogenomic analyses - use high throughput transcriptome and genome data to broadly sample many genes

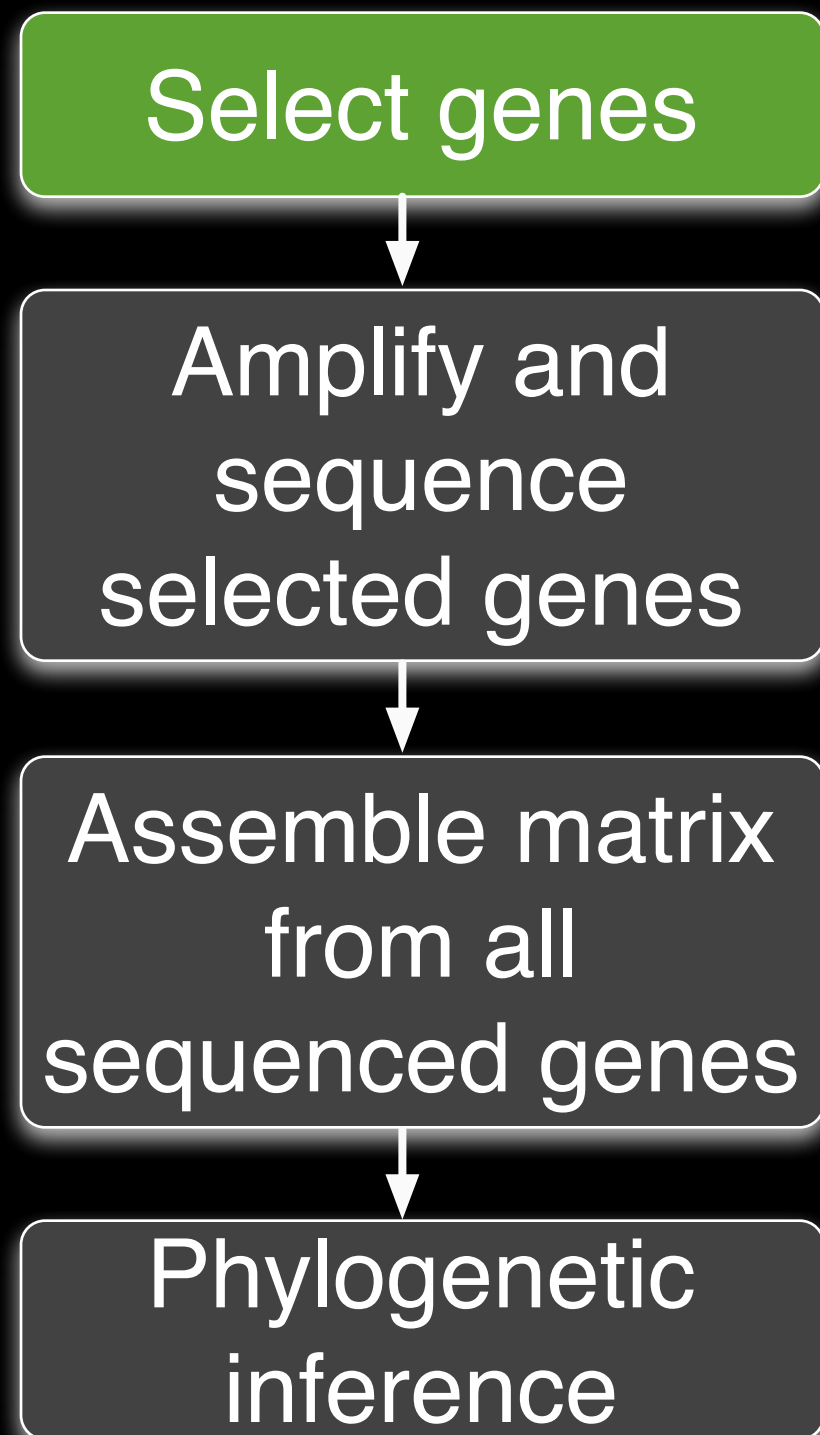
Gene selection as part of project design



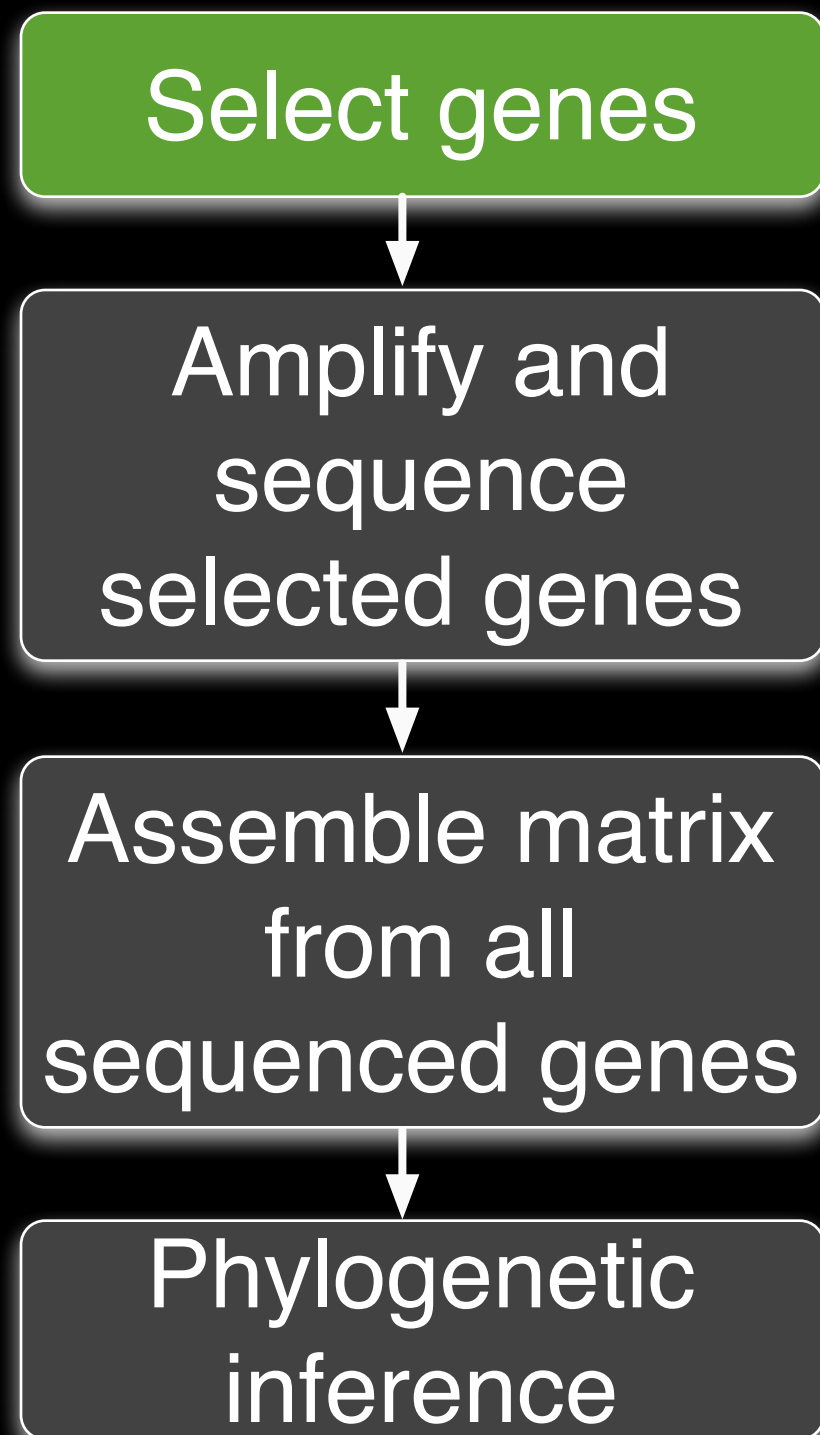
Gene selection as part of analysis



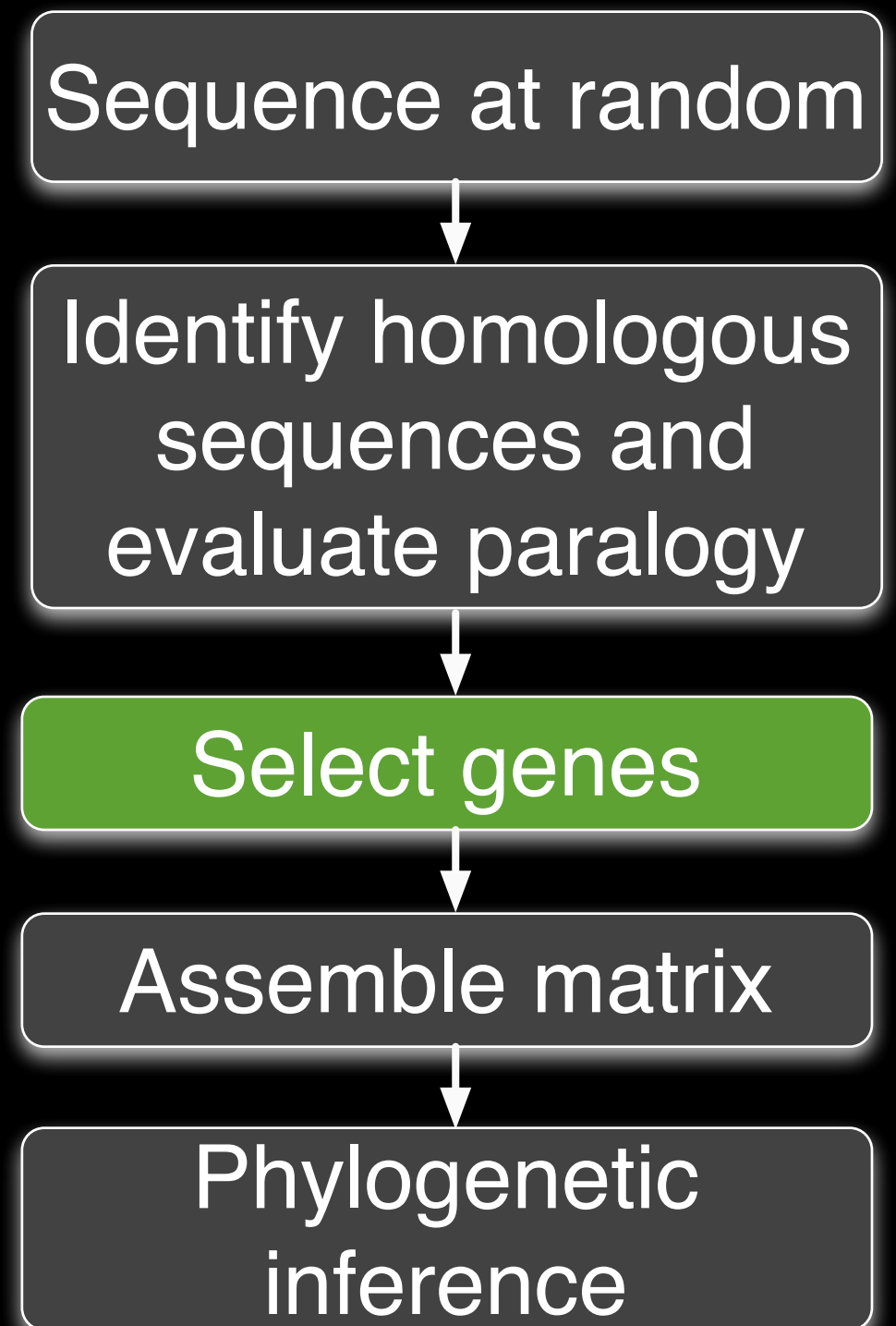
Gene selection as part of project design



Gene selection as part of project design



Gene selection as part of analysis



**Should you include
morphological data in your
analysis?**

**Should you include
morphological data in your
analysis?**

If you are analyzing fossils, it is
essential.

Common tools for inferring trees:

Parsimony

PAUP*

Poy

TNT

Likelihood

PAUP*

RAxML

GARLI

MrBayes

Bayesian

MrBayes

PhyloBayes