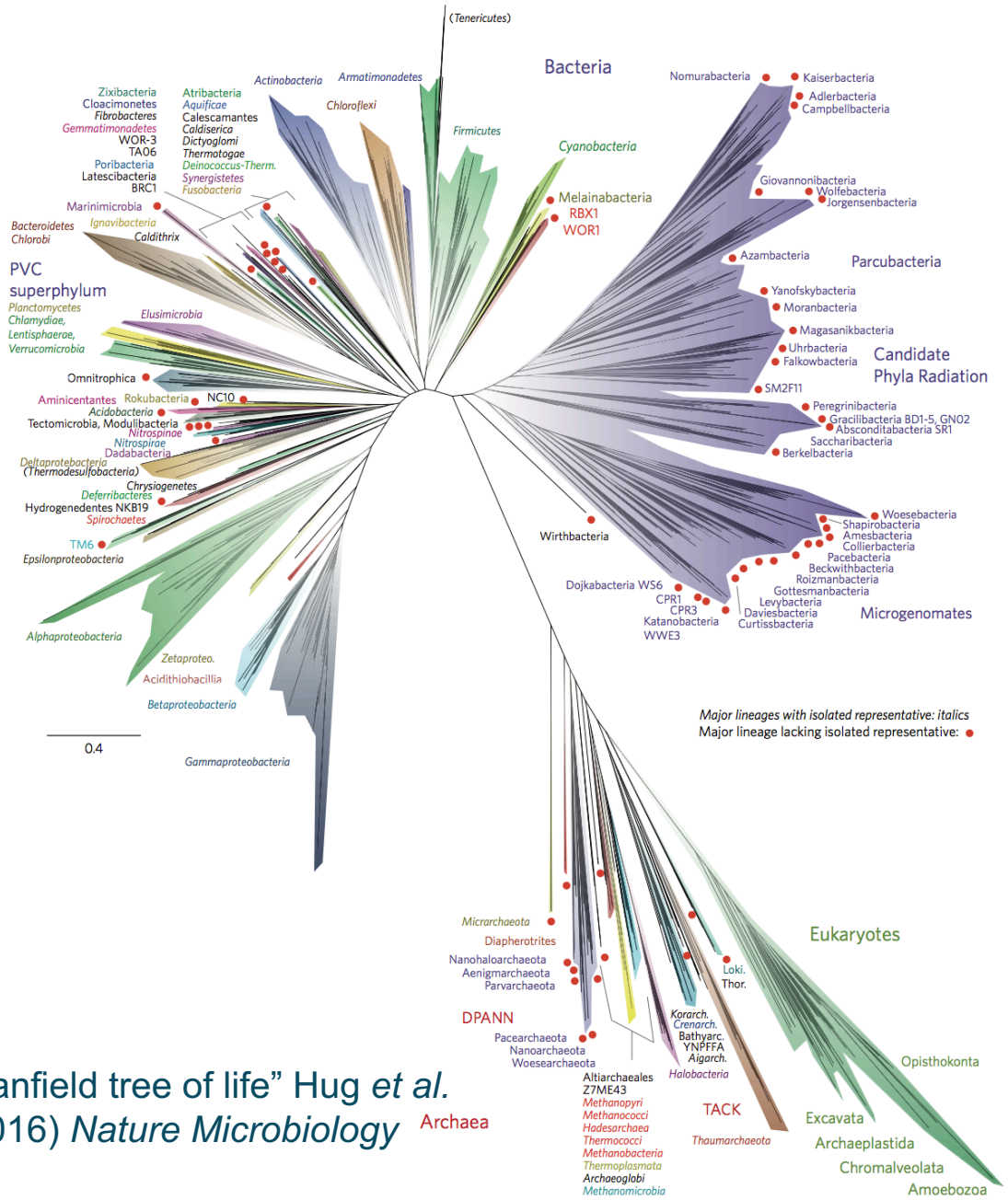
Evolution: Deeper into phylogenetics

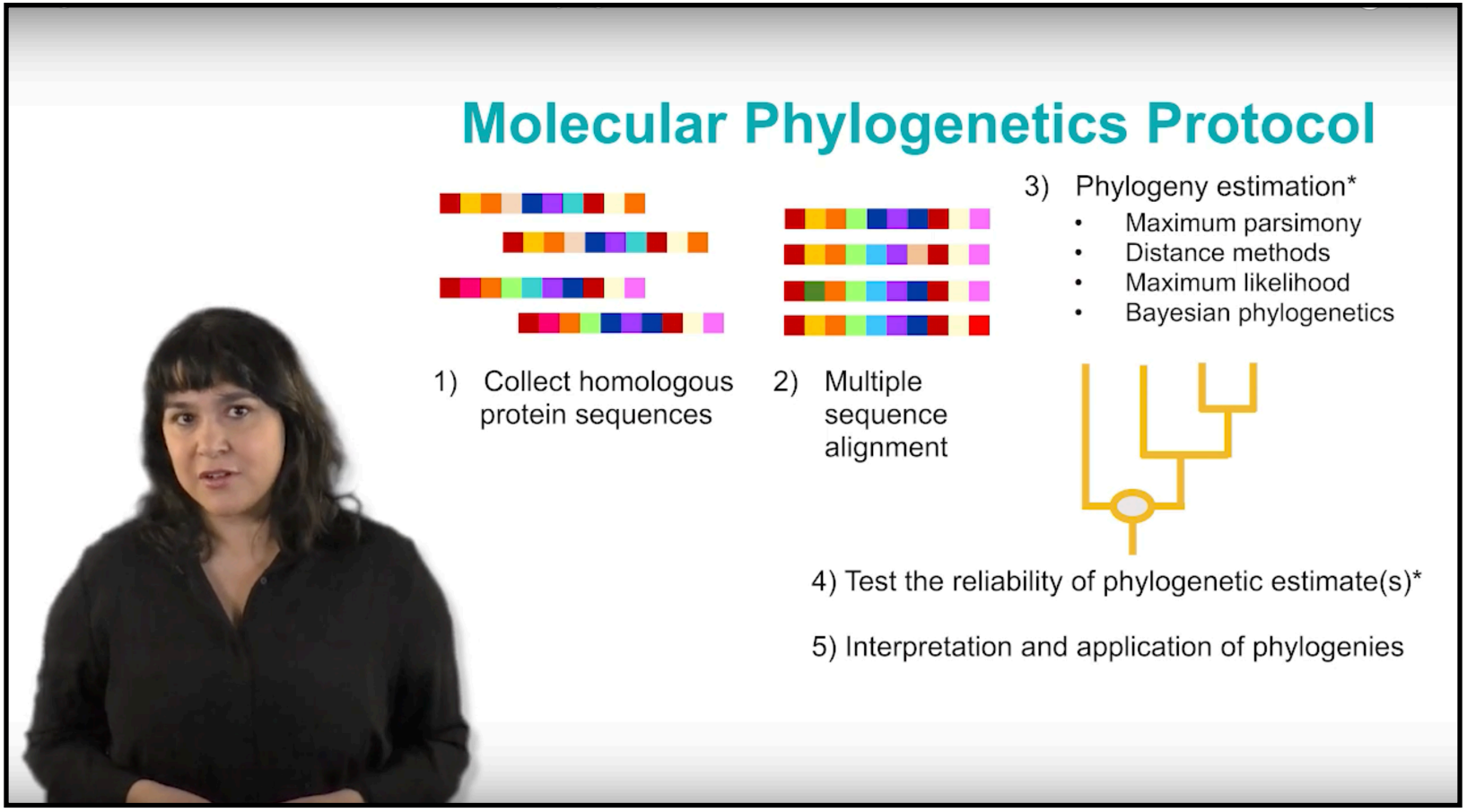
Andrew Rominger
Omidyar Postdoctoral Fellow
Santa Fe Institute

- How to infer a phylogeny
- Why the deep past is difficult

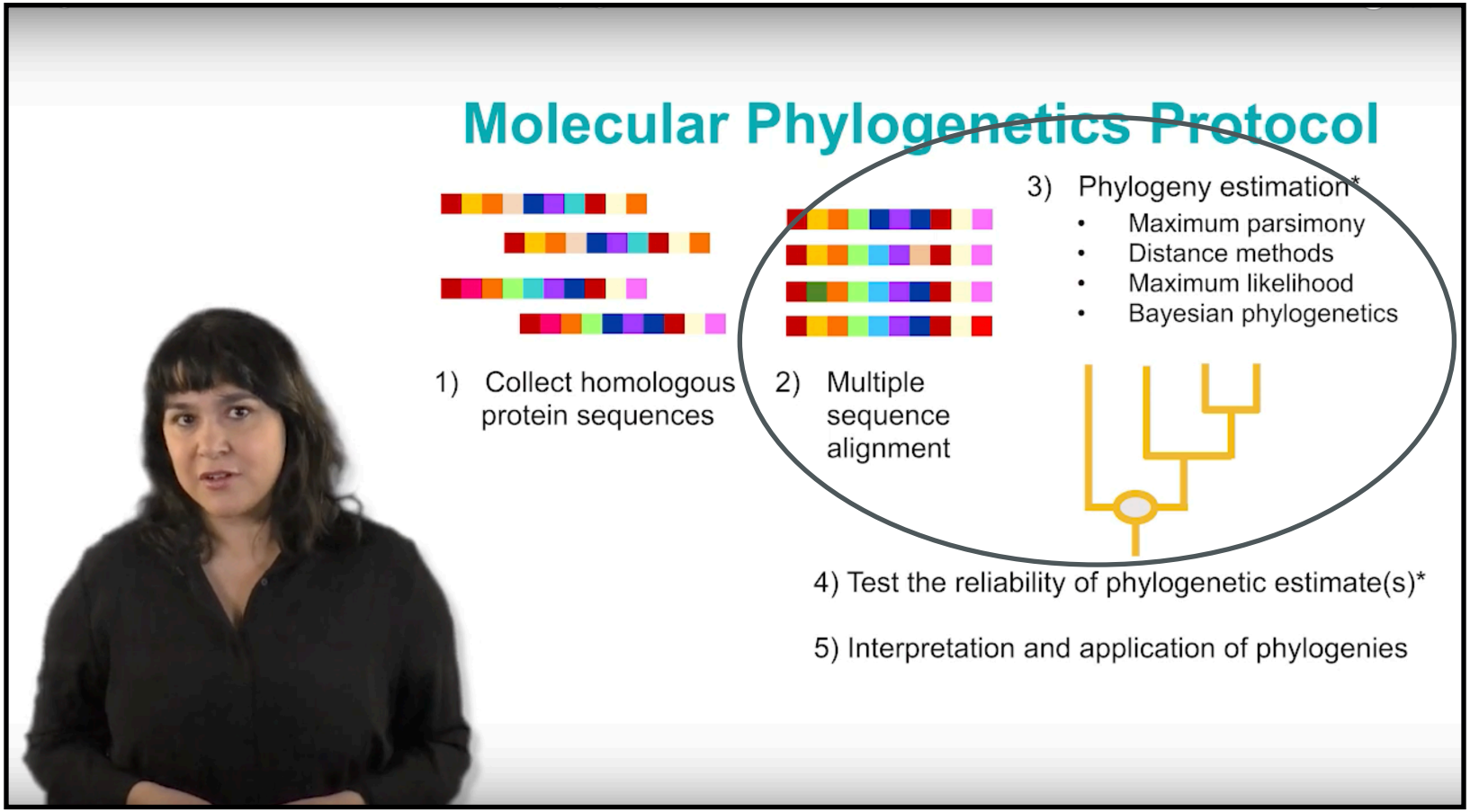


"Banfield tree of life" Hug *et al.* (2016) *Nature Microbiology*

Inferring a phylogeny

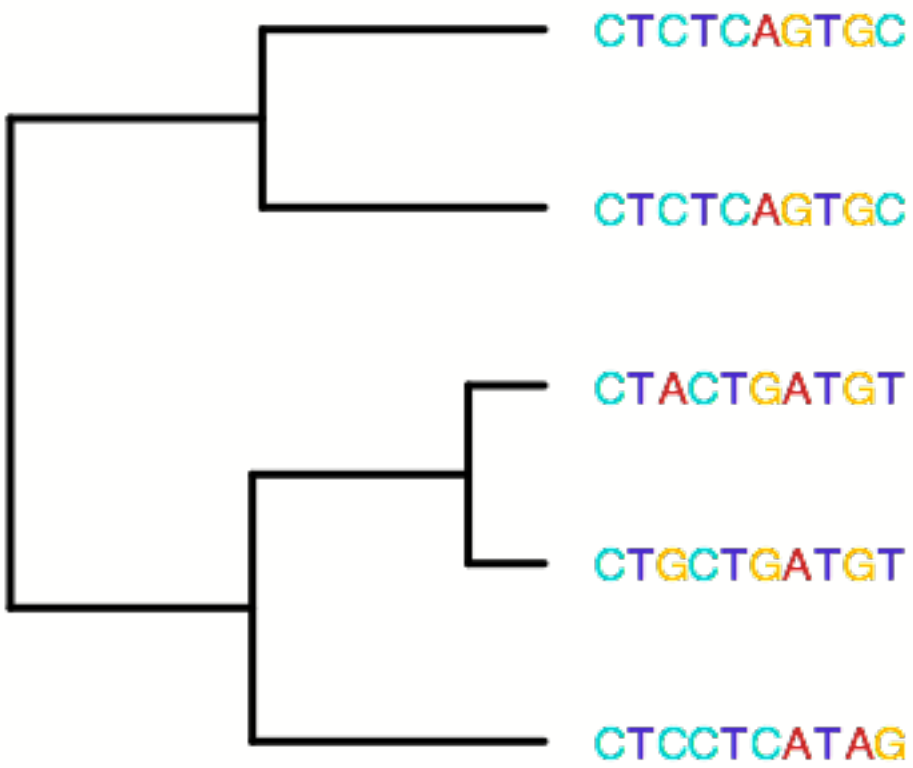


Inferring a phylogeny



Inferring a phylogeny

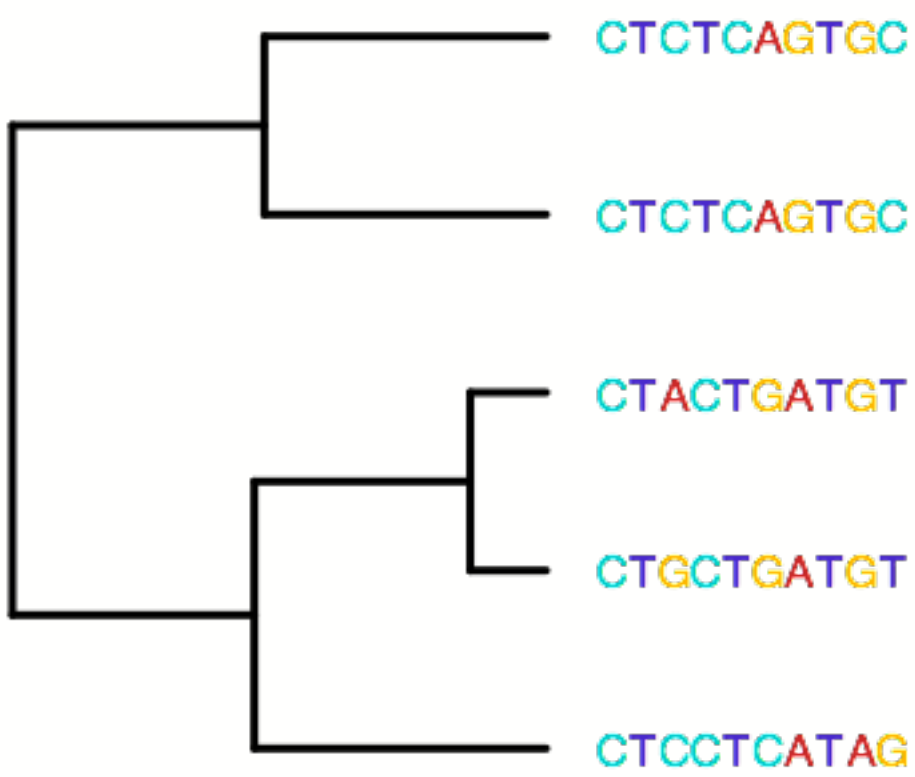
Goal: find the process leading to the data



Inferring a phylogeny

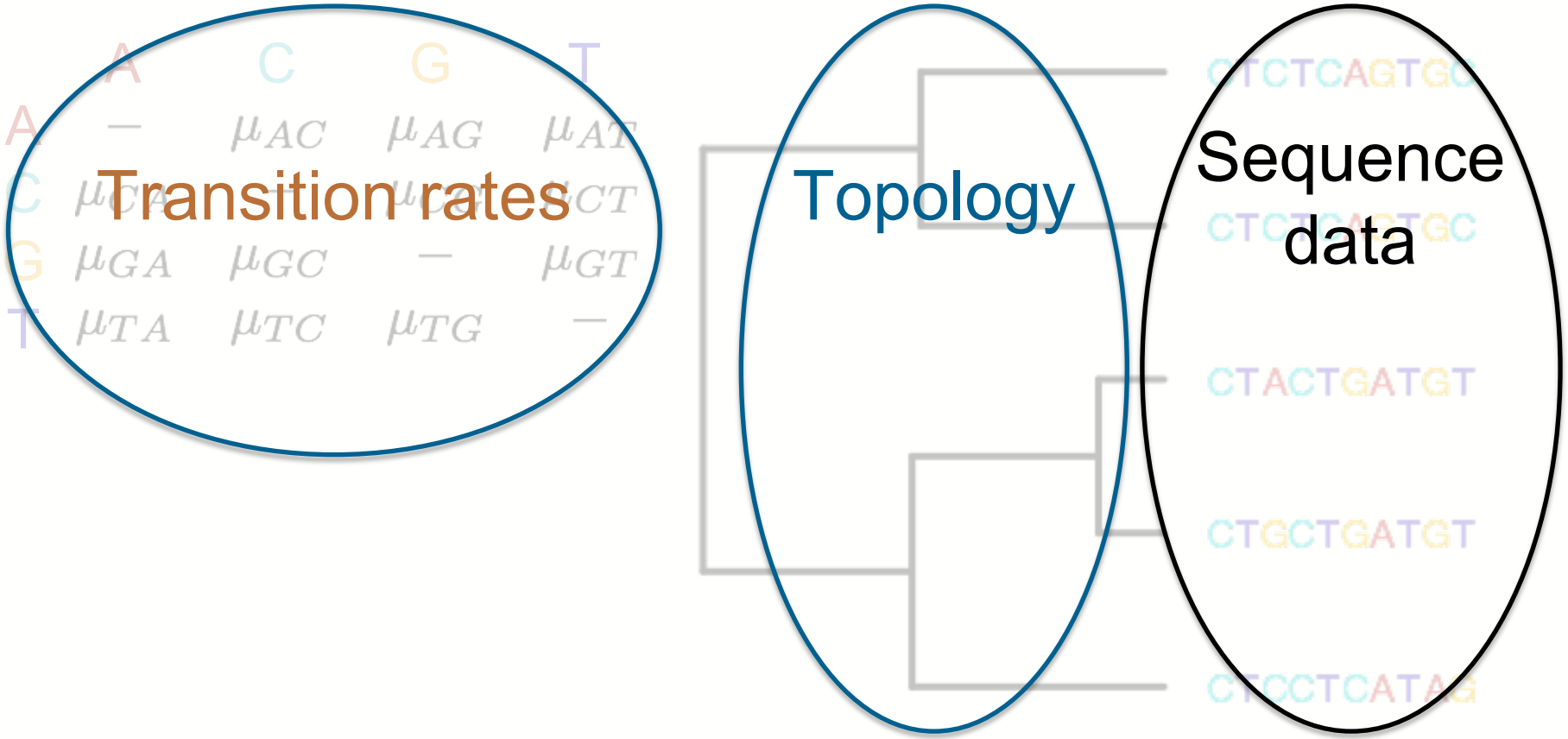
Goal: find the process leading to the data

	A	C	G	T
A	$-$	μ_{AC}	μ_{AG}	μ_{AT}
C	μ_{CA}	$-$	μ_{CG}	μ_{CT}
G	μ_{GA}	μ_{GC}	$-$	μ_{GT}
T	μ_{TA}	μ_{TC}	μ_{TG}	$-$



Inferring a phylogeny

Goal: find the process leading to the data



Inferring a phylogeny

Goal: find the process leading to the data

Likelihood:

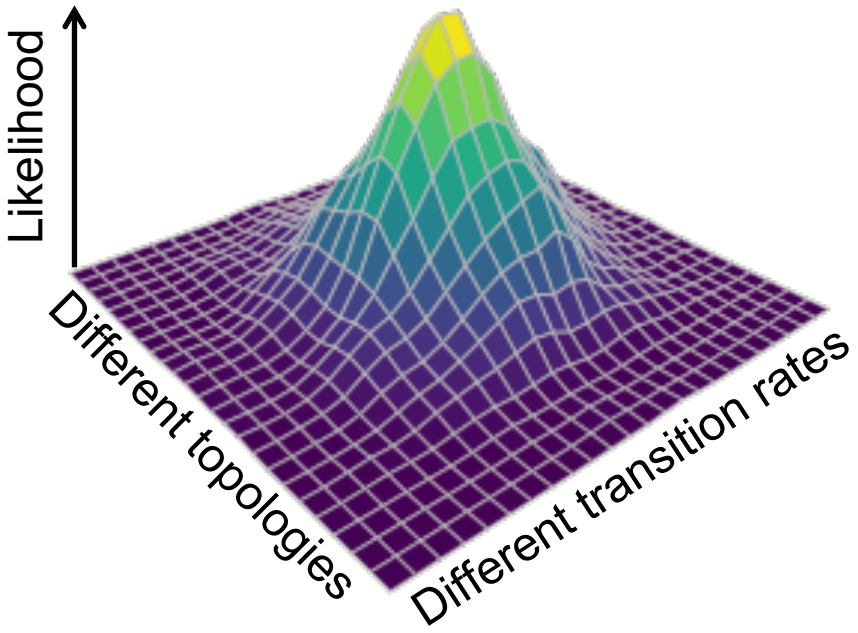
$$P(\text{Sequence data} \mid \text{Transition rates}, \text{Topology})$$

Inferring a phylogeny

Goal: find the process leading to the data

Likelihood:

$$P(\text{Sequence data} \mid \text{Transition rates}, \text{Topology})$$

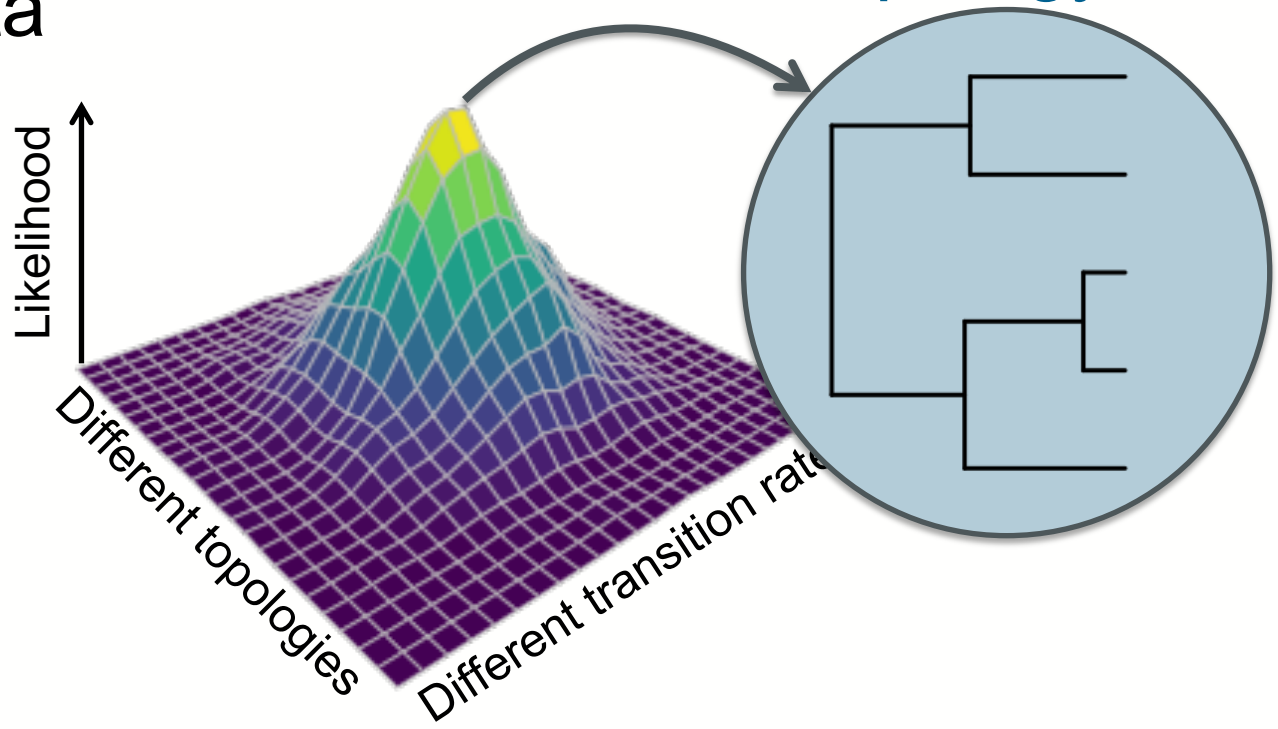


Inferring a phylogeny

Goal: find the process leading to the data

Likelihood:

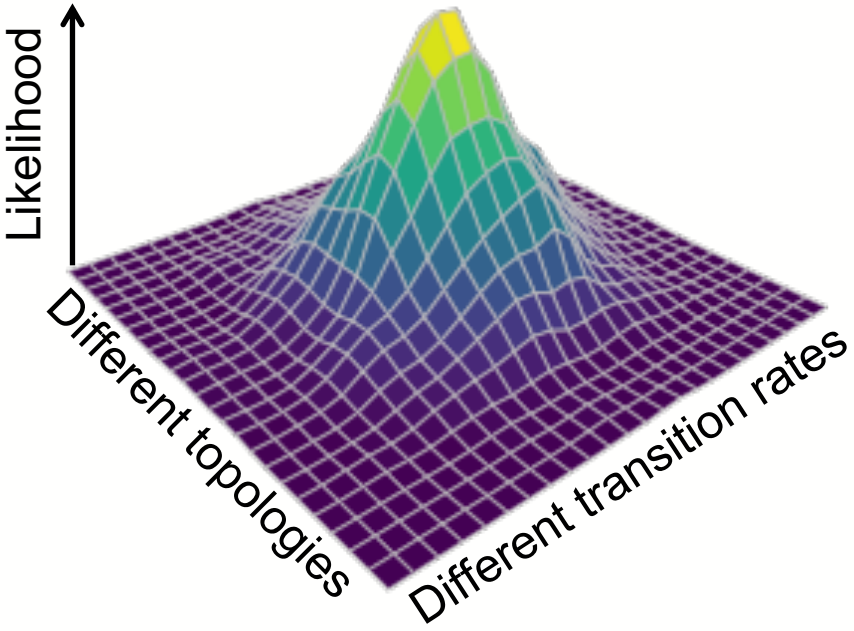
$$P(\text{Sequence data} \mid \text{Transition rates}, \text{Topology})$$



Inferring a phylogeny

Goal: find the process leading to the data

Bayesian inference: traversing the complex topology and rate matrix space lends itself to Bayesian algorithms



Inferring a phylogeny

Goal: find the process leading to the data

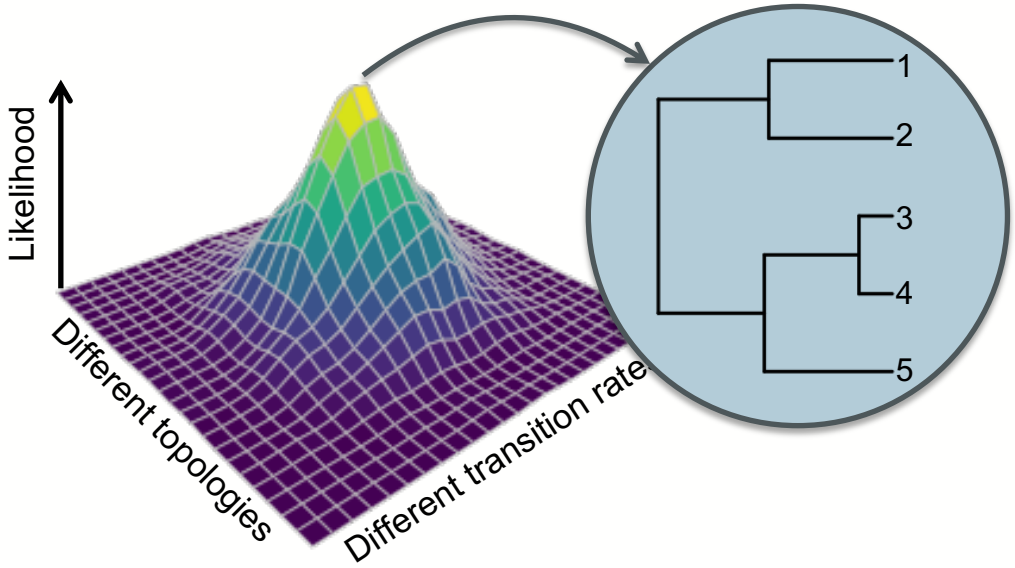
Information on parsimony and distance-based methods are left for the Further Reading section

Inferring a phylogeny

Caveats: inferring the past is hard

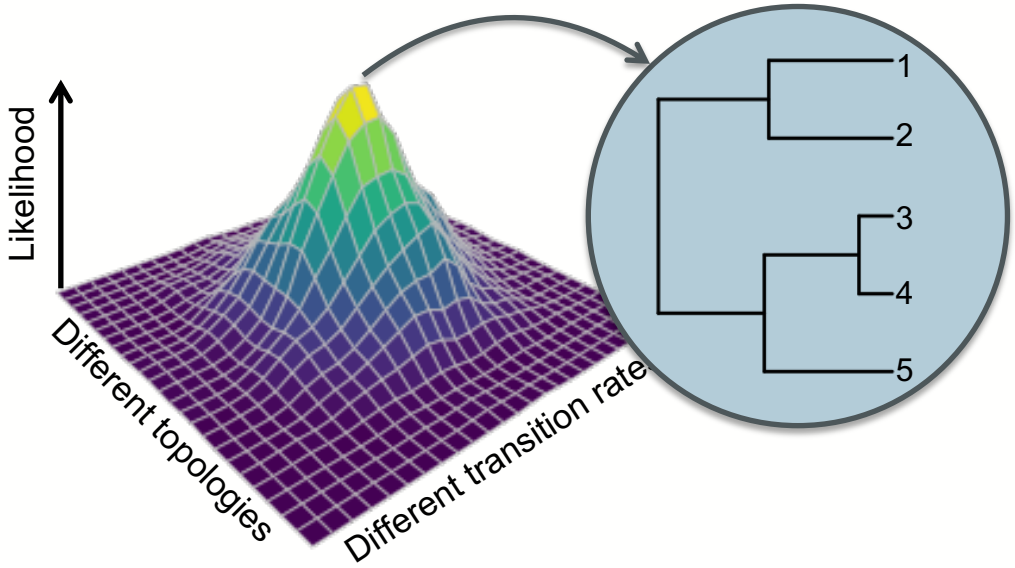
Inferring a phylogeny

Caveats: inferring the past is hard

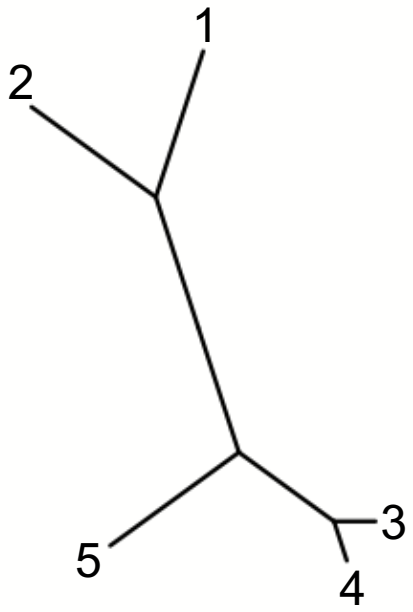


Inferring a phylogeny

Caveats: inferring the past is hard

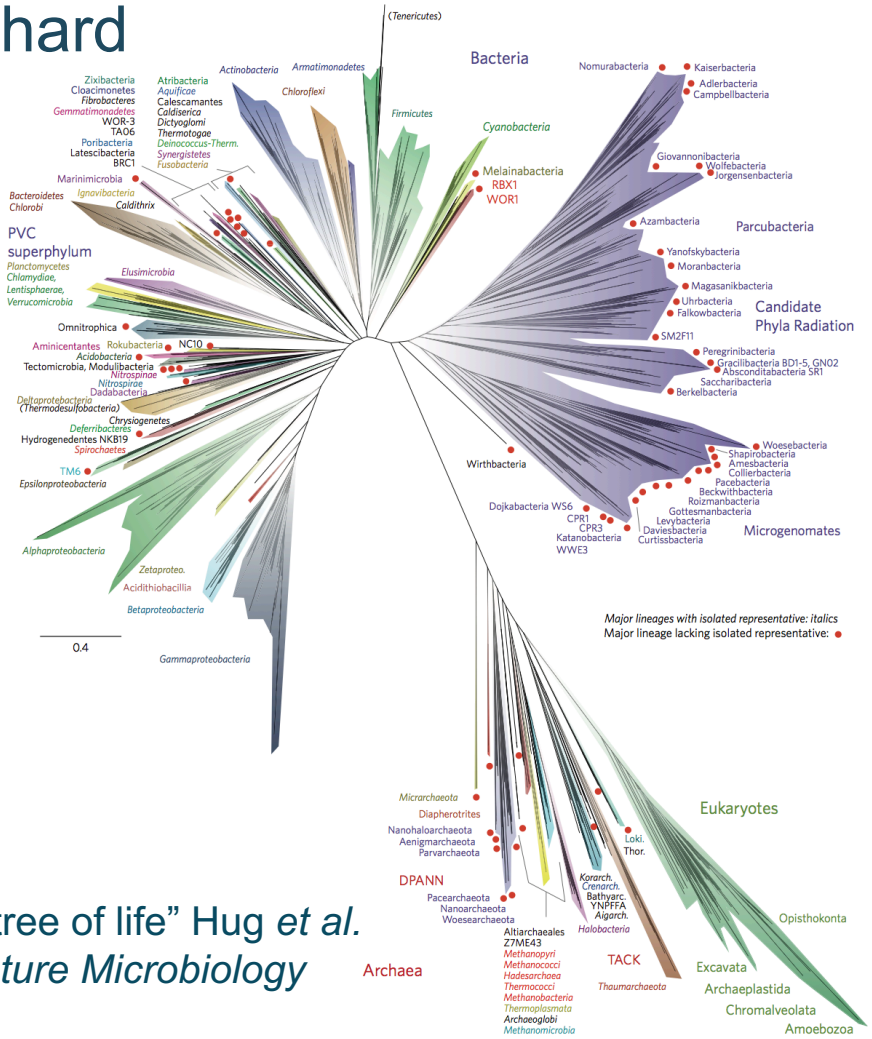


"Unrooted" phylogeny arrises because mutation process is reversible in our models



Inferring a phylogeny

Caveats: inferring the past is hard

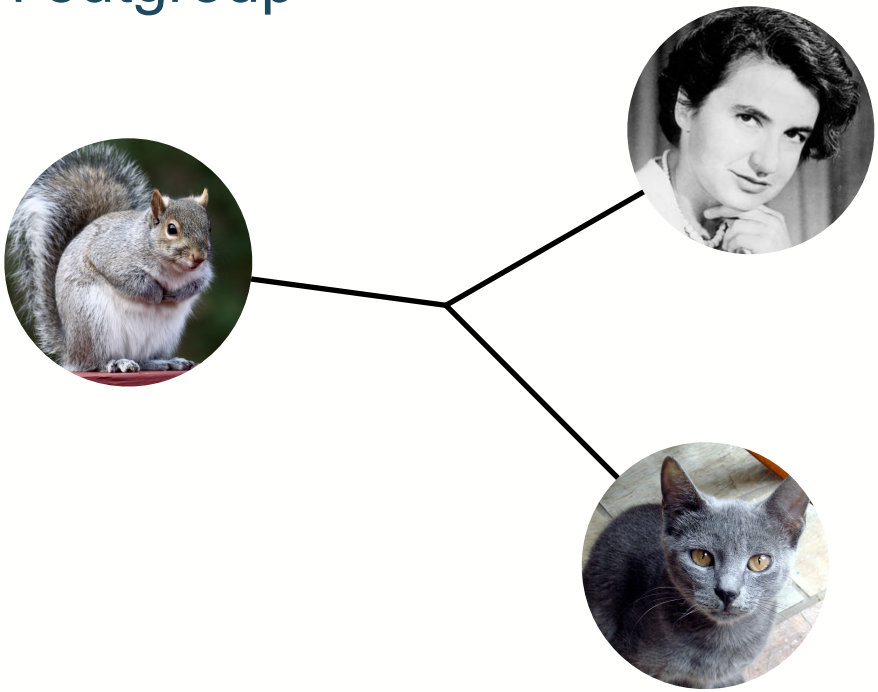


"Banfield tree of life" Hug *et al.* (2016) *Nature Microbiology*

Inferring a phylogeny

Caveats: inferring the past is hard

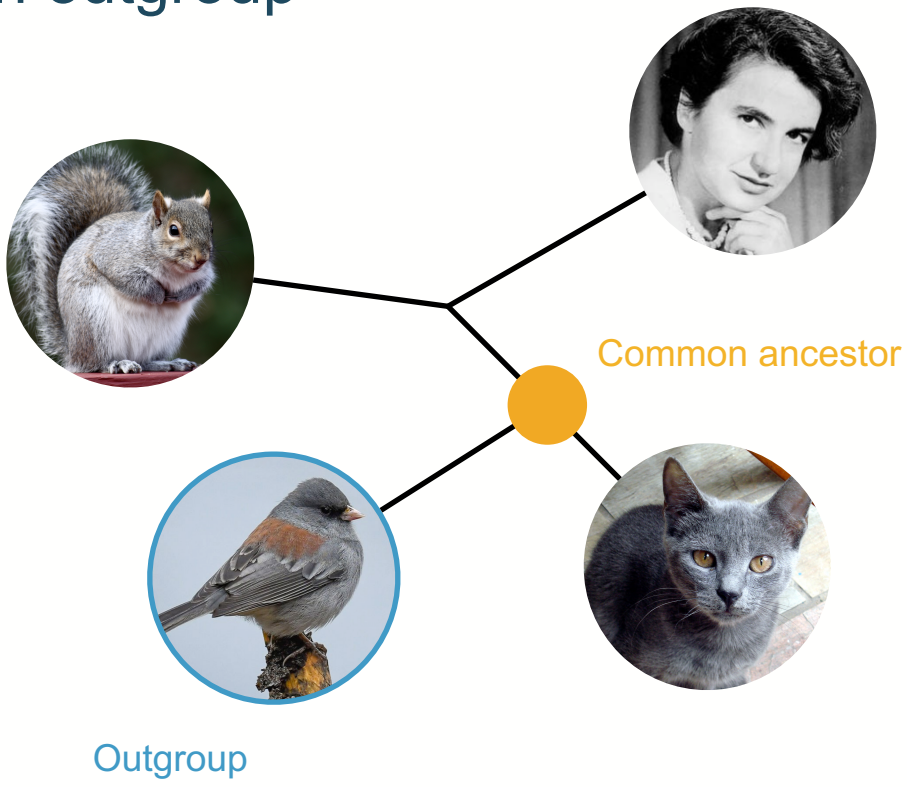
- Rooting the tree with an outgroup



Inferring a phylogeny

Caveats: inferring the past is hard

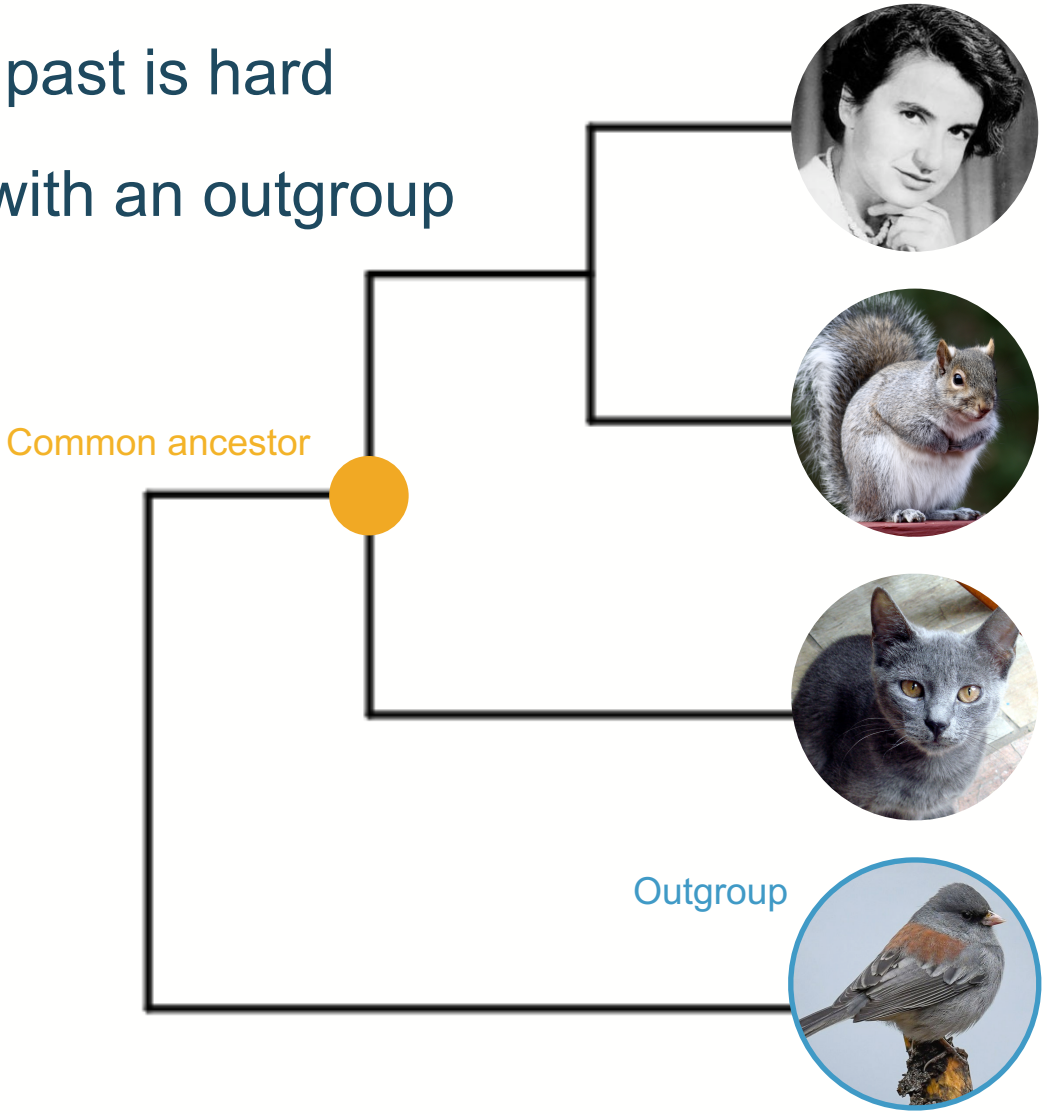
- Rooting the tree with an outgroup



Inferring a phylogeny

Caveats: inferring the past is hard

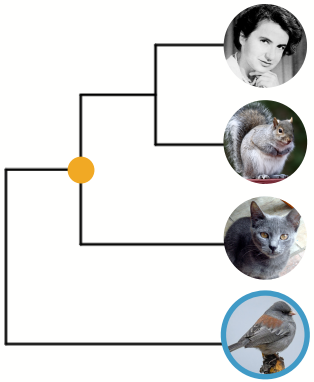
- Rooting the tree with an outgroup



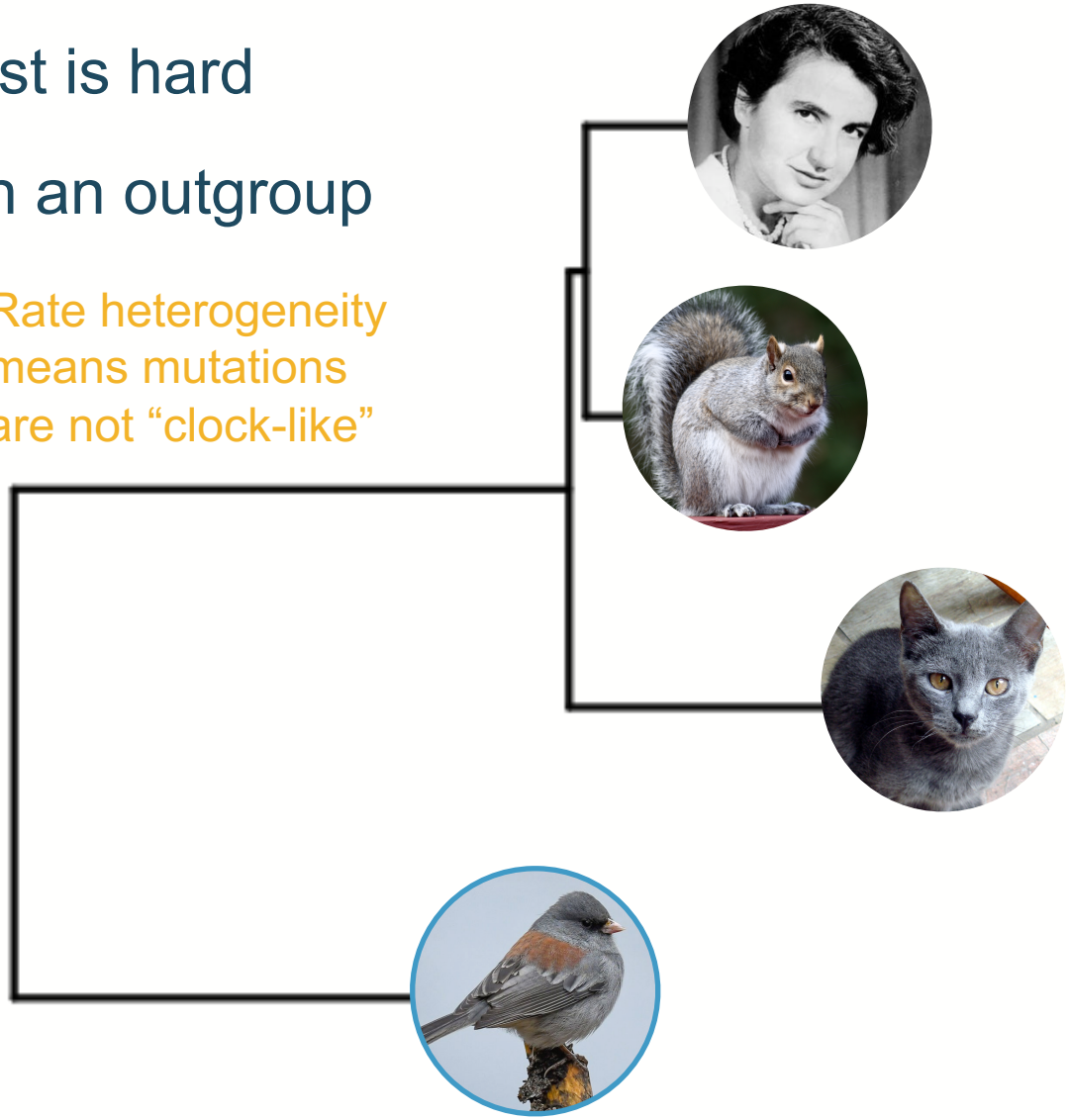
Inferring a phylogeny

Caveats: inferring the past is hard

- Rooting the tree with an outgroup



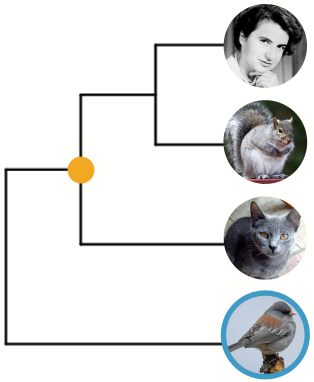
Rate heterogeneity means mutations are not “clock-like”

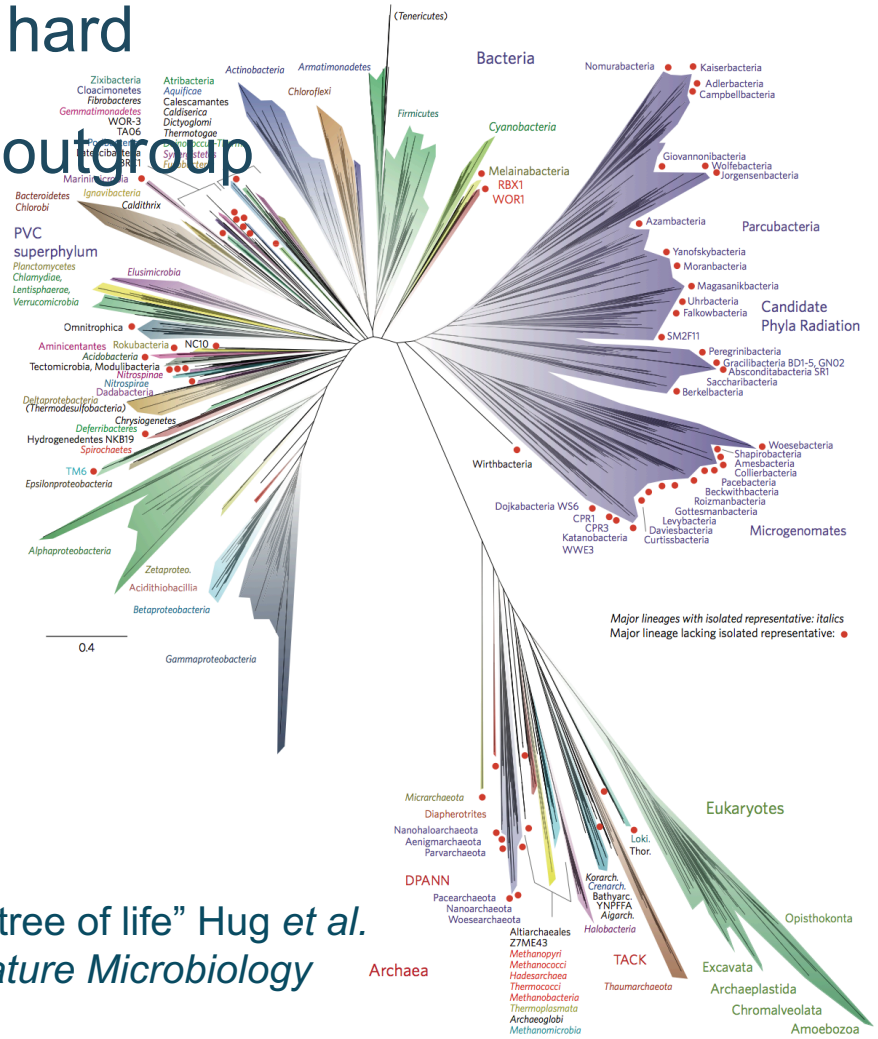


Inferring a phylogeny

Caveats: inferring the past is hard

- Rooting the tree with an outgroup



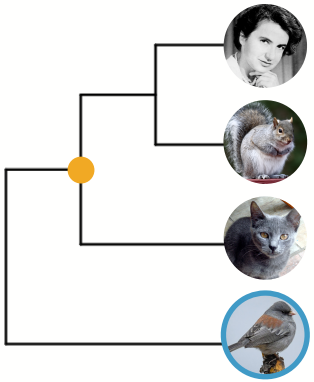


"Banfield tree of life" Hug *et al.* (2016) *Nature Microbiology*

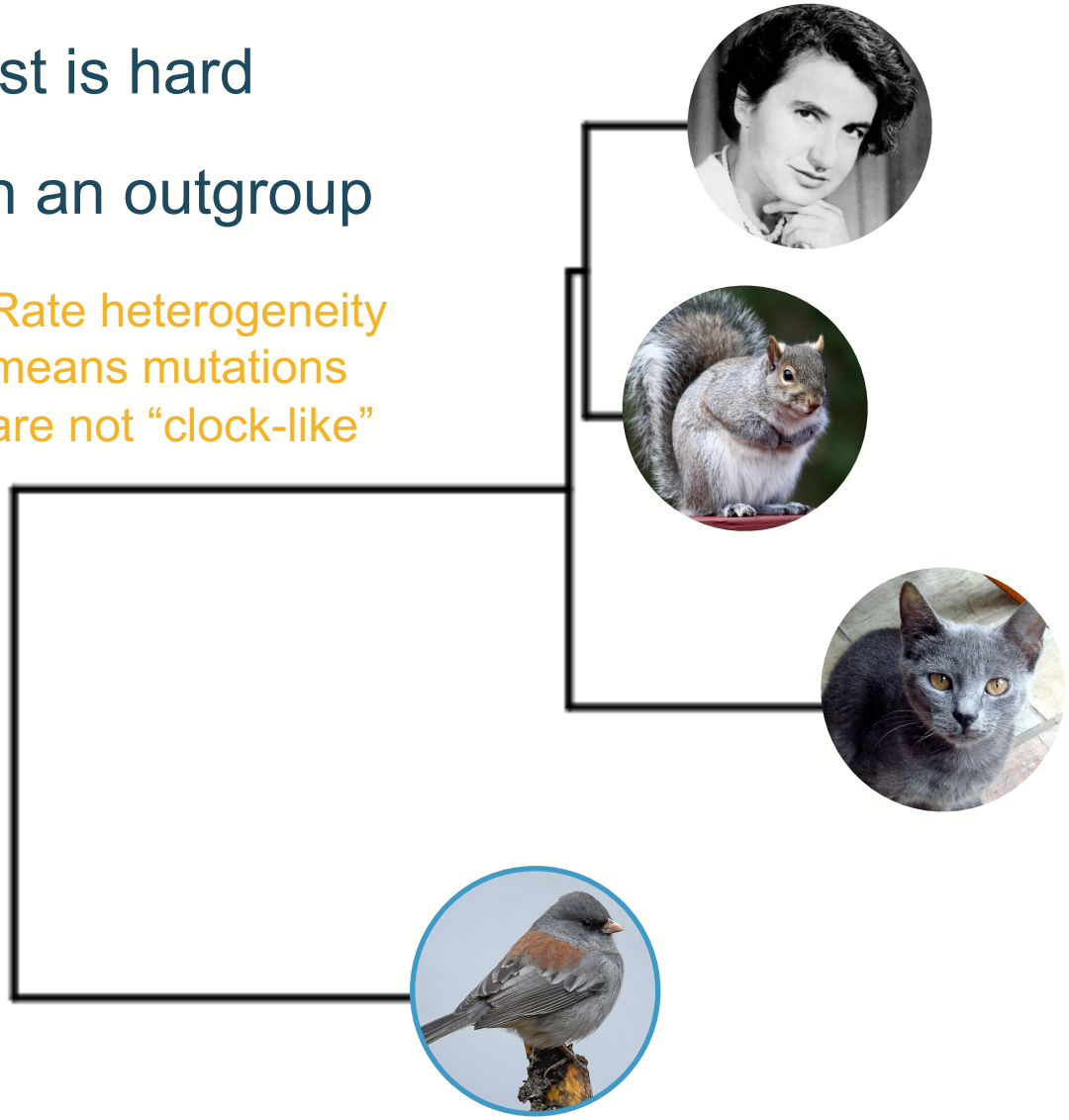
Inferring a phylogeny

Caveats: inferring the past is hard

- Rooting the tree with an outgroup



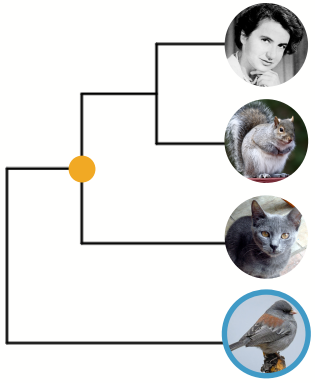
Rate heterogeneity means mutations are not “clock-like”

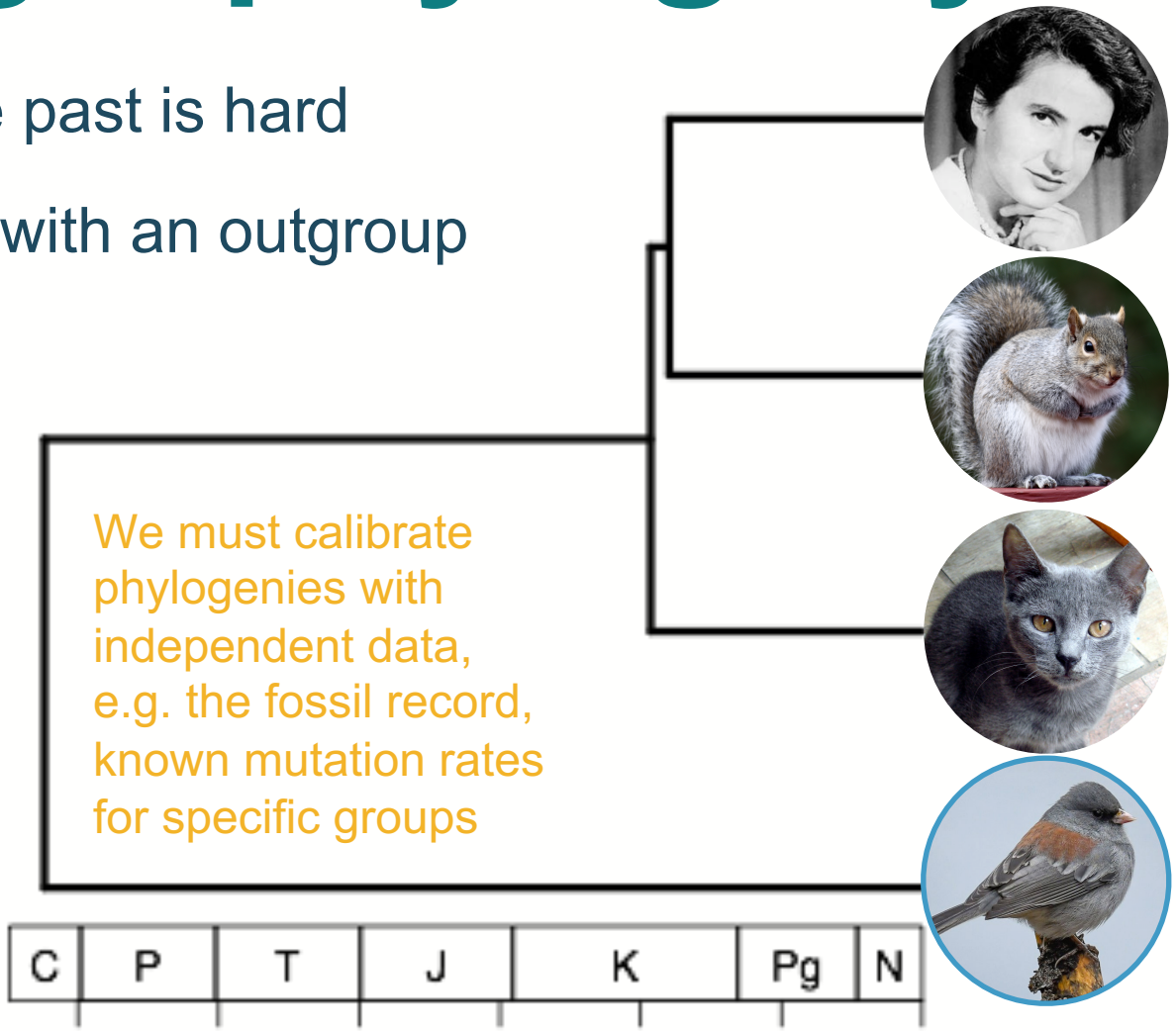


Inferring a phylogeny

Caveats: inferring the past is hard

- Rooting the tree with an outgroup
- Time calibration





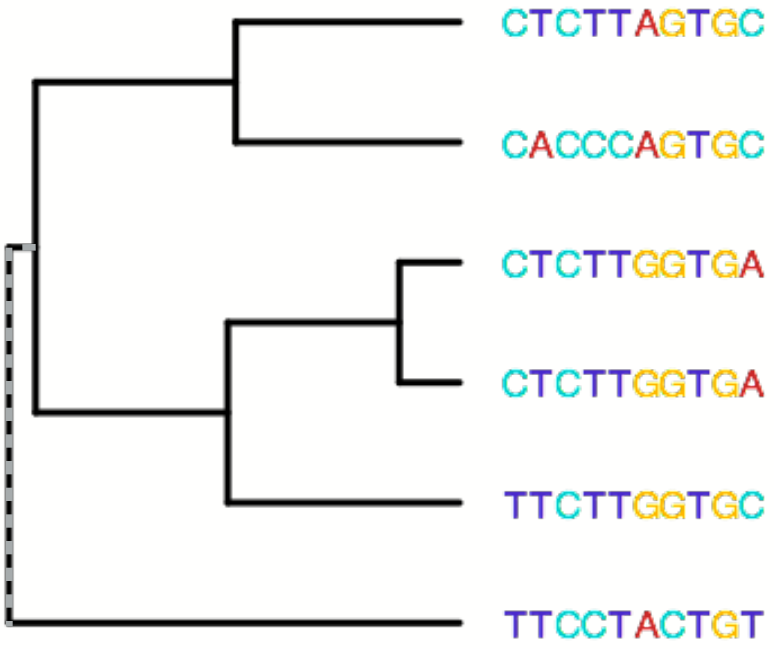
Inferring a phylogeny

Caveats: inferring the past is hard

- Rooting the tree with an outgroup
- Time calibration

	A	C	G	T
A	$-$	μ_{AC}	μ_{AG}	μ_{AT}
C	μ_{CA}	$-$	μ_{CG}	μ_{CT}
G	μ_{GA}	μ_{GC}	$-$	μ_{GT}
T	μ_{TA}	μ_{TC}	μ_{TG}	$-$

Long branches “attract” each other



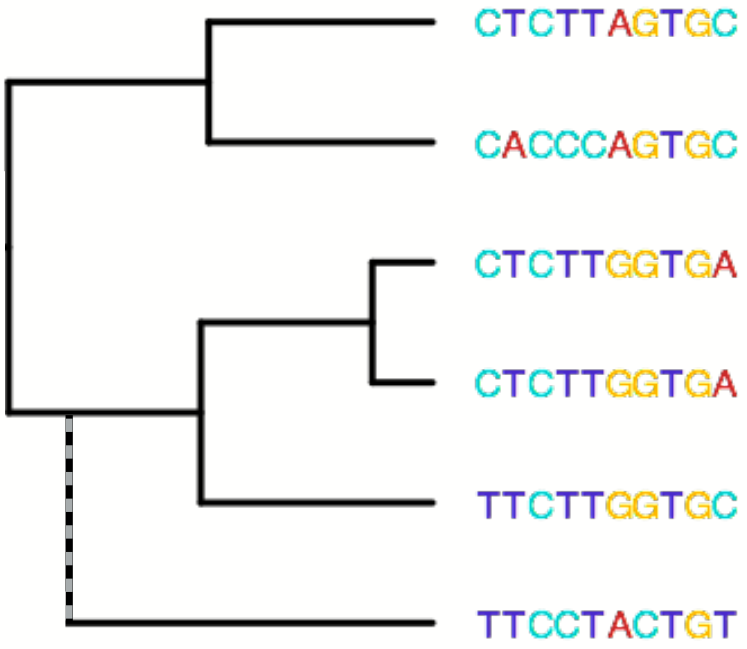
Inferring a phylogeny

Caveats: inferring the past is hard

- Rooting the tree with an outgroup
- Time calibration

	A	C	G	T
A	$-$	μ_{AC}	μ_{AG}	μ_{AT}
C	μ_{CA}	$-$	μ_{CG}	μ_{CT}
G	μ_{GA}	μ_{GC}	$-$	μ_{GT}
T	μ_{TA}	μ_{TC}	μ_{TG}	$-$

Long branches “attract” each other



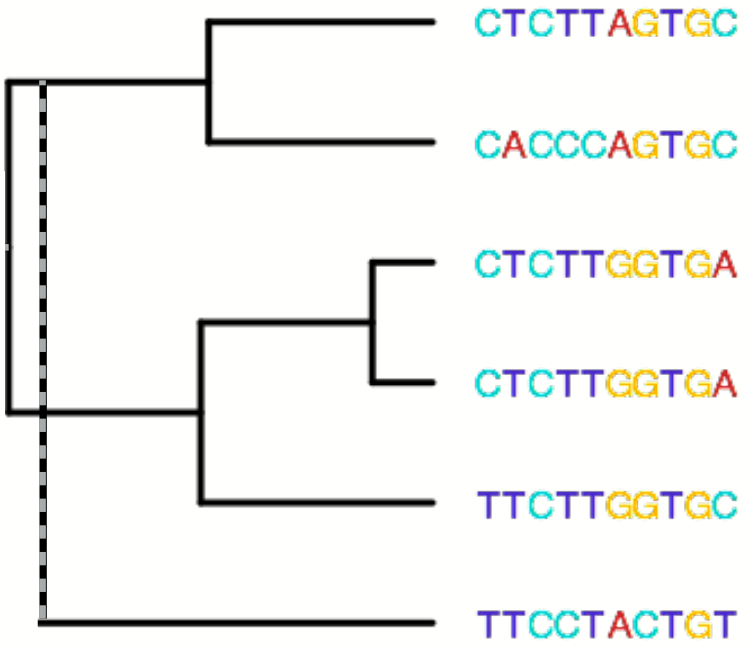
Inferring a phylogeny

Caveats: inferring the past is hard

- Rooting the tree with an outgroup
- Time calibration

	A	C	G	T
A	$-$	μ_{AC}	μ_{AG}	μ_{AT}
C	μ_{CA}	$-$	μ_{CG}	μ_{CT}
G	μ_{GA}	μ_{GC}	$-$	μ_{GT}
T	μ_{TA}	μ_{TC}	μ_{TG}	$-$

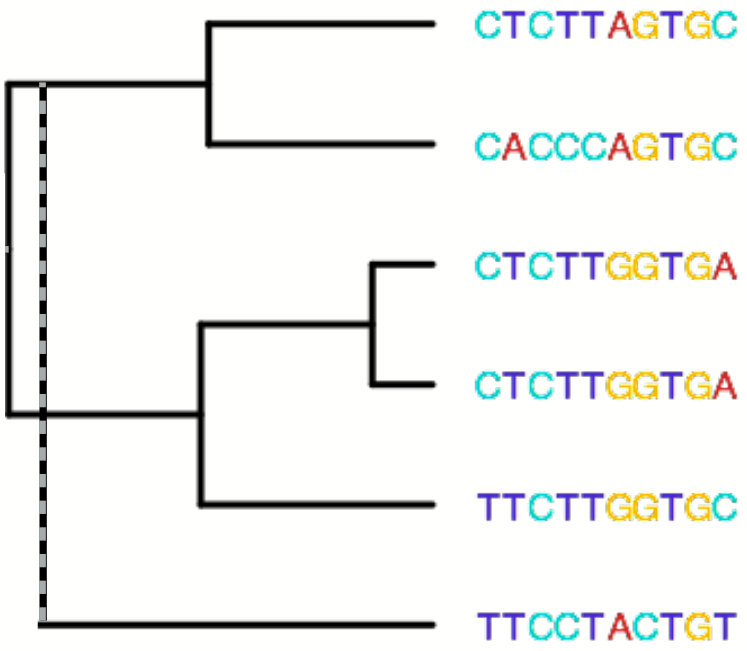
Long branches “attract” each other



Inferring a phylogeny

Caveats: inferring the past is hard

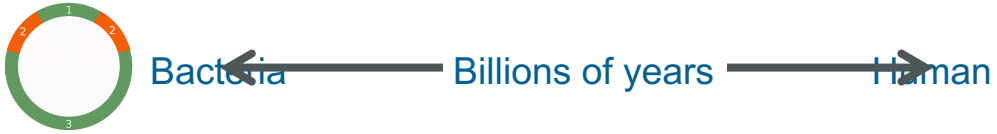
- Rooting the tree with an outgroup
- Time calibration
- Long branch attraction

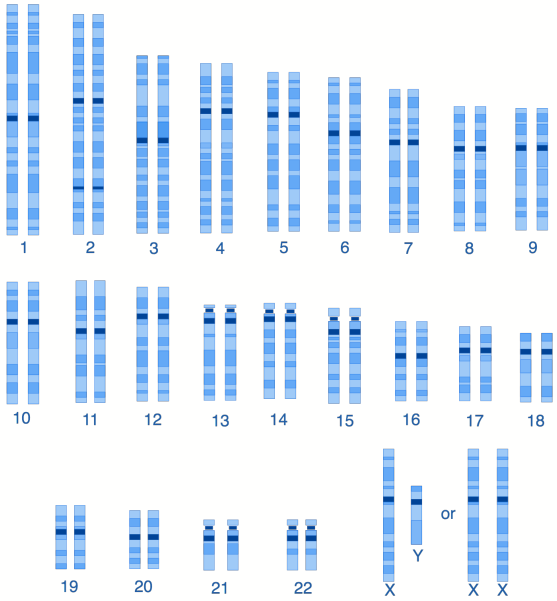


Inferring a phylogeny

Caveats: inferring the past is hard (especially in deep time)

- Rooting the tree with an outgroup
- Time calibration
- Long branch attraction
- What genetic information goes back to LUCA?

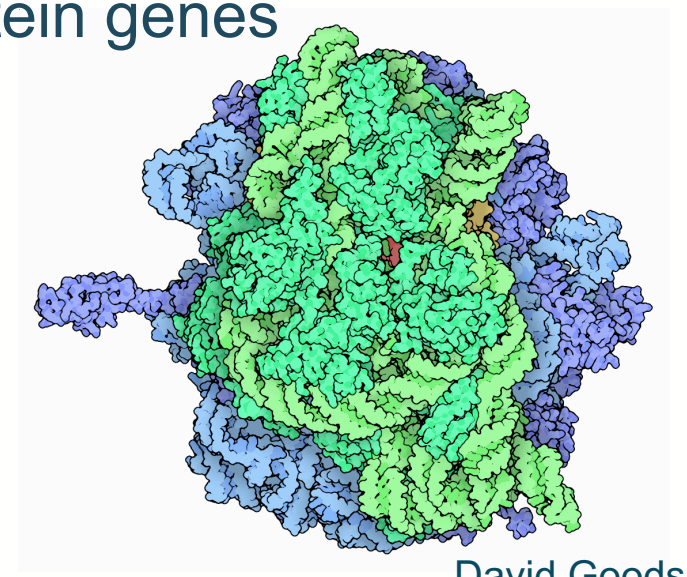




Inferring a phylogeny

Caveats: inferring the past is hard (especially in deep time)

- Rooting the tree with an outgroup
- Time calibration
- Long branch attraction
- What genetic information goes back to LUCA?
  - Ribosomal RNA and protein genes



David Goodsell

References and Suggested Reading

Huelsenbeck, J. P., & Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*, 17, 754-755.

Huelsenbeck, J. P., & Crandall, K. A. (1997). Phylogeny estimation and hypothesis testing using maximum likelihood. *Annual Review of Ecology and Systematics*, 28, 437-466.

Hug, L. A., Baker, B. J., Anantharaman, K., ... & Banfield, J. F. (2016). A new view of the tree of life. *Nature Microbiology*, 1, 16048.

Philippe, H., Zhou, Y., Brinkmann, H., Rodrigue, N., & Delsuc, F. (2005). Heterotachy and long-branch attraction in phylogenetics. *BMC Evolutionary Biology*, 5, 50.

Quast, C., Pruesse, E., Yilmaz, P., ... & Glöckner, F. O. (2012). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids*