

Assessing phylogenetic model adequacy

Lecturers:
David A. Duchene

Molecular Ecology, Evolution, and Phylogenetics Laboratory
University of Sydney

09/02/2017



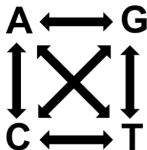
Model uncertainty
Substitution models
Clock models
Model adequacy
Example: subst.
models
Test statistics
Assessing other
models
Concluding notes
References

Data

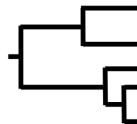
AAACCCAACA
AATCGGTTCA
AATCGGAATT
AAACCCAATA
AATCGGTTCA

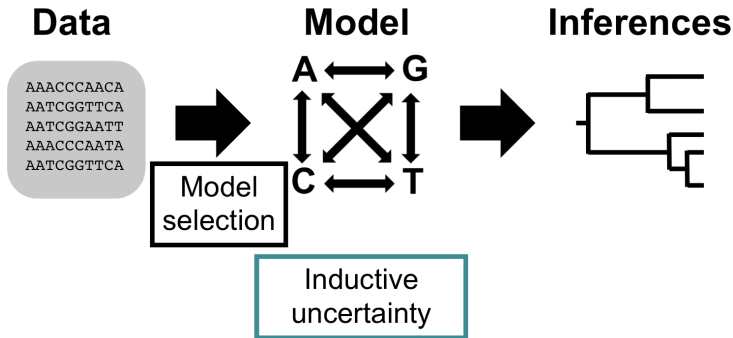


Model

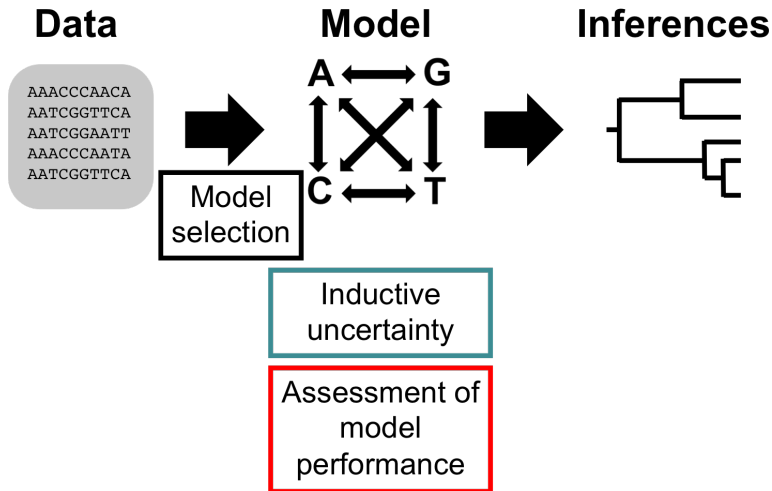


Inferences





Model uncertainty
Substitution models
Clock models
Model adequacy
Example: subst.
models
Test statistics
Assessing other
models
Concluding notes
References

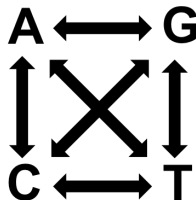


Model uncertainty
Substitution models
Clock models
Model adequacy
Example: subst. models
Test statistics
Assessing other models
Concluding notes
References

The most common models of substitution are those of the GTR family.

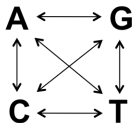
Some assumptions in GTR models include:

- ▶ Substitution events are **independent and identically distributed** events
- ▶ **Reversibility** (the expected outcome is the same regardless of the direction of the process)
- ▶ The parameter values are equal across lineages
 - ▶ Transition rates (assumption of **homogeneity**)
 - ▶ Base frequencies (assumption of **stationarity**)



Even a model that is complex relative to others can be a poor representation of the evolutionary process.

Rate Matrix



Base Frequencies Site Rates

$$\pi_A + \pi_C + \pi_G + \pi_T = 1 \quad + I + G$$

Number of models

$$203 \quad \times \quad 15 \quad \times \quad 4 \quad = \quad 12,180$$

Model uncertainty

Substitution models

Clock models

Model adequacy

Example: subst.
models

Test statistics

Assessing other
models

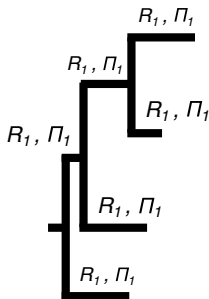
Concluding notes

References

Phylogenetic substitution models

Model
adequacy

GTR models are very simple relative to other possible models.



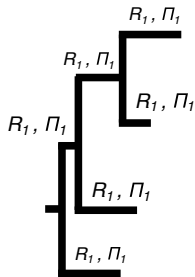
Time-reversible,
Homogeneous,
Stationary

Model uncertainty
Substitution models
Clock models
Model adequacy
Example: subst.
models
Test statistics
Assessing other
models
Concluding notes
References

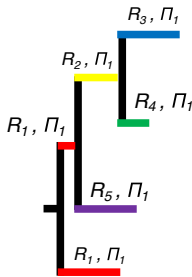
Phylogenetic substitution models

Model
adequacy

GTR models are very simple relative to other possible models.



Time-reversible,
Homogeneous,
Stationary



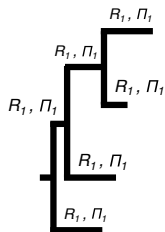
Non-reversible,
Non-homogeneous,
Stationary

Model uncertainty
Substitution models
Clock models
Model adequacy
Example: subst.
models
Test statistics
Assessing other
models
Concluding notes
References

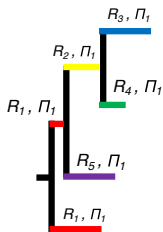
Phylogenetic substitution models

Model
adequacy

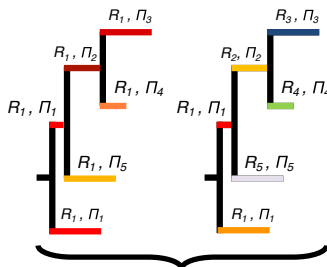
GTR models are very simple relative to other possible models.



Time-reversible,
Homogeneous,
Stationary



Non-reversible,
Non-homogeneous,
Stationary



Non-reversible,
Non-homogeneous,
Non-stationary

Model uncertainty

Substitution models

Clock models

Model adequacy

Example: subst.
models

Test statistics

Assessing other
models

Concluding notes

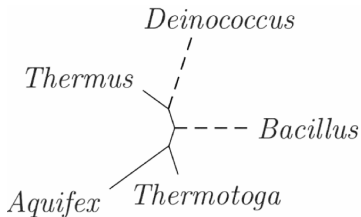
References

Phylogenetic substitution models

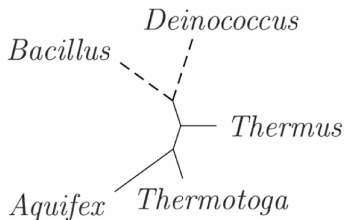
Model
adequacy

The consequences of using an inadequate model can be dramatic. The following is an example using data from bacteria:

True relationships



Inference from 16S region



Model uncertainty

Substitution models

Clock models

Model adequacy

Example: subst.
models

Test statistics

Assessing other
models

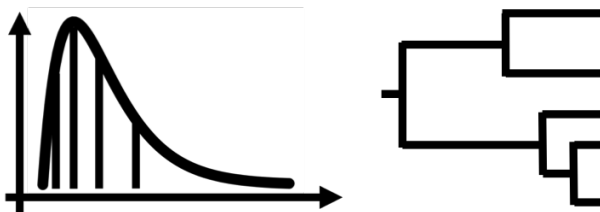
Concluding notes

References

Phylogenetic clock models

Model
adequacy

- ▶ Other models that are fundamental in the context of BEAST are the **rates** and **times** models.
- ▶ Divergence time estimates can be misleading if any of these has a large departure from the underlying evolutionary process.



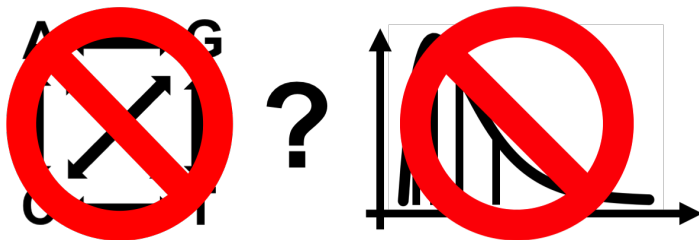
- ▶ One way to assess these two is to investigate the component that links them to the substitutions model: **substitutions per site** (product of rates and times)

Model uncertainty
Substitution models
Clock models
Model adequacy
Example: subst.
models
Test statistics
Assessing other
models
Concluding notes
References

"The data are good, but the models are bad"

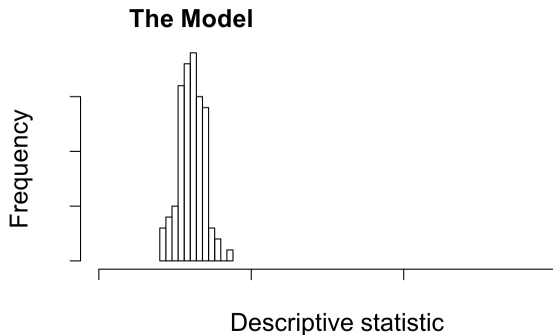
Model
adequacy

How can we assess each model independently for their merit, as opposed to relative to other models?



The most widely accepted method to do this is to assess model adequacy.

Model uncertainty
Substitution models
Clock models
Model adequacy
Example: subst.
models
Test statistics
Assessing other
models
Concluding notes
References



Model uncertainty

Substitution models

Clock models

Model adequacy

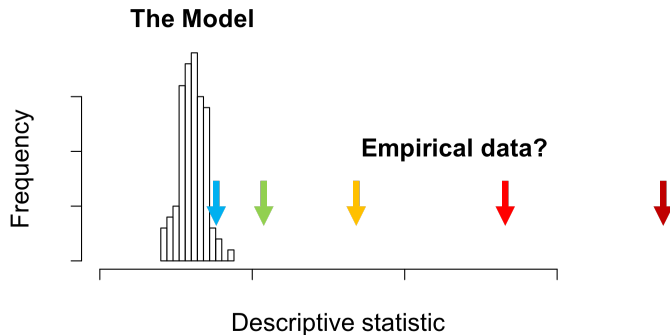
Example: subst.
models

Test statistics

Assessing other
models

Concluding notes

References



Model uncertainty

Substitution models

Clock models

Model adequacy

Example: subst. models

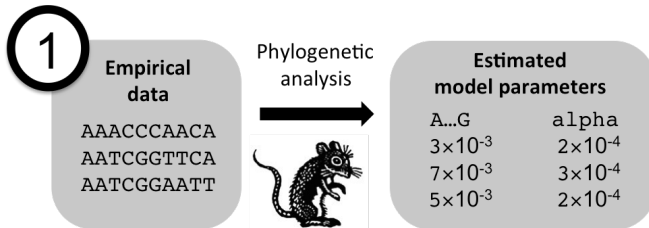
Test statistics

Assessing other models

Concluding notes

References

Step 1. We perform a standard analysis using the candidate model and our empirical data.



Model uncertainty

Substitution models

Clock models

Model adequacy

Example: subst.
models

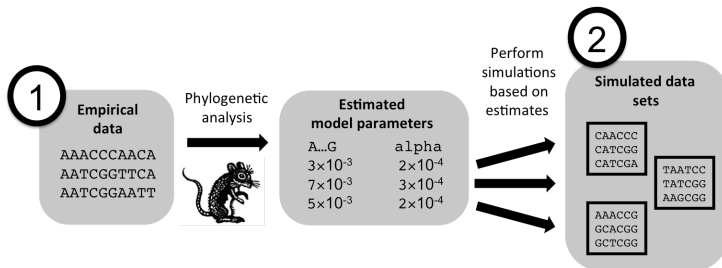
Test statistics

Assessing other
models

Concluding notes

References

Step 2. We **simulate data from the model** that resemble the empirical data in size and parameter values.

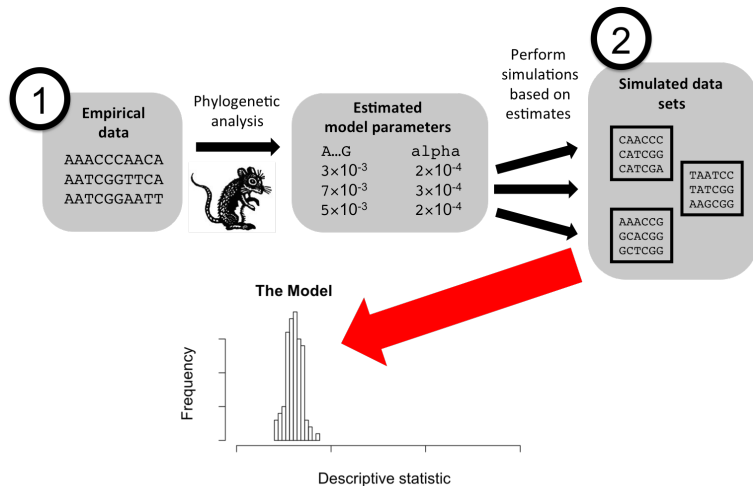


Model uncertainty
Substitution models
Clock models
Model adequacy
Example: subst.
models
Test statistics
Assessing other
models
Concluding notes
References

Assessing substitution model adequacy

Model
adequacy

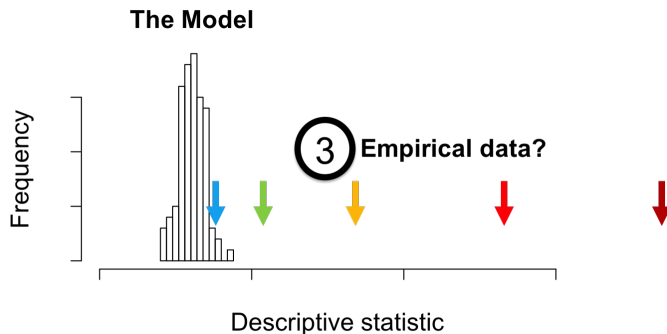
The simulated data are a **sample from the model**.



Model uncertainty
Substitution models
Clock models
Model adequacy
Example: subst. models
Test statistics
Assessing other models
Concluding notes
References

What is a test statistic?

Step 3. We compare the simulated data with the empirical data using a **descriptive** or **test** statistic.



The test statistic is a **metric** about a data set. It should allow us to quantify the **discrepancies** between the simulated data and the empirical data.

Data-based statistics

- **Multinomial likelihood**
- **Chi-squared statistic**

Model uncertainty

Substitution models

Clock models

Model adequacy

Example: subst.
models

Test statistics

Assessing other
models

Concluding notes

References

Data-based statistics

- **Multinomial likelihood**
- **Chi-squared statistic**

Inference-based statistics

- **Mean branch lengths**
- **Mean node support**

Model uncertainty

Substitution models

Clock models

Model adequacy

Example: subst.
models

Test statistics

Assessing other
models

Concluding notes

References

Data-based statistics

- **Multinomial likelihood**
- **Chi-squared statistic**

Inference-based statistics

- **Mean branch lengths**
- **Mean node support**

"Test statistic choice is only limited by a researcher's creativity."

Model uncertainty

Substitution models

Clock models

Model adequacy

Example: subst.
models

Test statistics

Assessing other
models

Concluding notes

References

Model uncertainty

Substitution models

Clock models

Model adequacy

Example: subst.
models

Test statistics

Assessing other
models

Concluding notes

References

Advances in methods for assessing model adequacy now allow us to assess many components of phylogenetic analyses.

- ▶ Substitution model
- ▶ Demographic models
- ▶ Multi-species coalescent model
- ▶ Times and rates model

Model uncertainty

Substitution models

Clock models

Model adequacy

Example: subst.
models

Test statistics

Assessing other
models

Concluding notes

References

- ▶ Most test statistics will not give a total assessment of model adequacy, but assess a specific aspect of a model. Some parameters might still be inadequate if they are not being assessed!
- ▶ Failing an assessment of model adequacy does not mean the model is useless. It can be useful for a subset of the data, or useful to infer at least some of the model parameters.

- Bollback, J. P. (2002). Bayesian model adequacy and choice in phylogenetics. *Molecular Biology and Evolution*, 19(7):1171.
- Brown, J. M. (2014). Detection of implausible phylogenetic inferences using posterior predictive assessment of model fit. *Systematic biology*, 63(3):334–348.
- Drummond, A. J. and Suchard, M. A. (2008). Fully bayesian tests of neutrality using genealogical summary statistics. *BMC genetics*, 9(1):68.
- Duchne, D. A., Duchne, S., Holmes, E. C., and Ho, S. Y. (2015). Evaluating the adequacy of molecular clock models using posterior predictive simulations. *Molecular Biology and Evolution*, 32(11):2986.
- Edwards, C., Holmes, E., Pybus, O., Wilson, D., Viscidi, R., Abrams, E., Phillips, R., and Drummond, A. (2006). Evolution of the human immunodeficiency virus envelope gene is dominated by purifying selection. *Genetics*, 174(3):1441–1453.
- Foster, P. G. (2004). Modeling compositional heterogeneity. *Systematic Biology*, 53(3):485–495.
- Goldman, N. (1993). Statistical tests of models of dna substitution. *Journal of Molecular Evolution*, 36(2):182–198.
- Reid, N. M., Hird, S. M., Brown, J. M., Pelletier, T. A., McVay, J. D., Satler, J. D., and Carstens, B. C. (2014). Poor fit to the multispecies coalescent is widely detectable in empirical data. *Systematic biology*, 63(3):322–333.

Model uncertainty

Substitution models

Clock models

Model adequacy

Example: subst.
models

Test statistics

Assessing other
models

Concluding notes

References