

The state-of-the-art of modeling
the molecular evolution

How large data can we analyze?

- What is the performance and time/memory/storage/bandwidth complexity (both in theory and in practice) as a function of
 - sequence length
 - number of sequences
 - evolutionary model
 - other variables?
- what happens when these numbers get VERY large (e.g. thousands of genome-sized

Priors

- Flat or more informative priors?
- What are the priors that we can choose in a Bayesian analyses?
- What are the pros and cons?

Involved substitution models?

- Site dependent
- Context dependent
- Lineage dependent
- Grammars
- Non-local correlations