Bayesian estimation of time-trees:

A journey through a strange land

Luiz Max Carvalho

Postgraduate Symposium 2017
Institute of Evolutionary Biology

THE UNIVERSITY
of EDINBURGH

Acknowledgements

Andrew Rambaut
UoE

Marc Suchard
UCLA

Guy Baele
KU Leuven

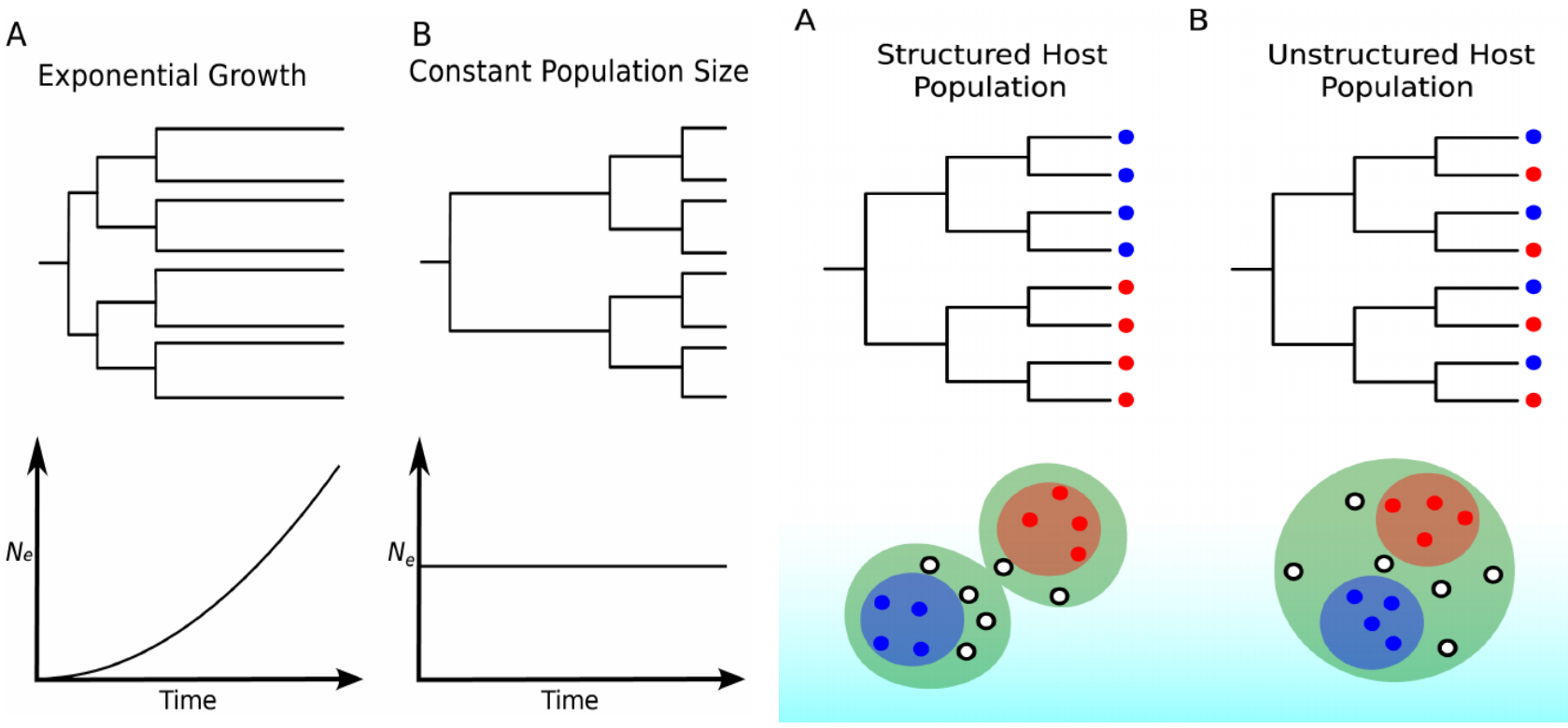
Motivation

Phylodynamics of fast-evolving viruses

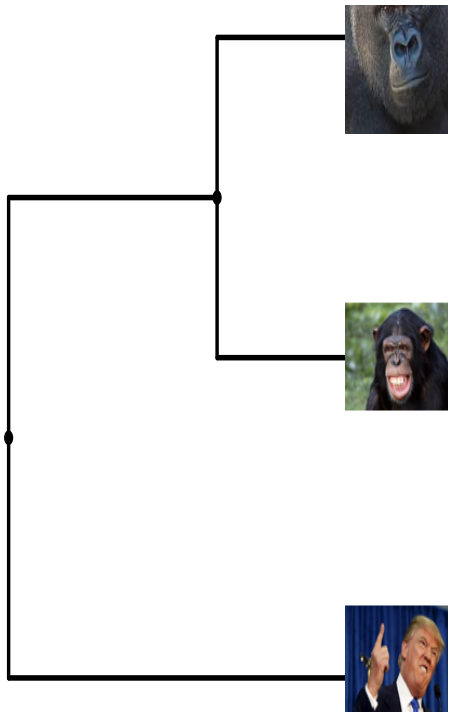
Inferring spatial and temporal dynamics from genomic data:

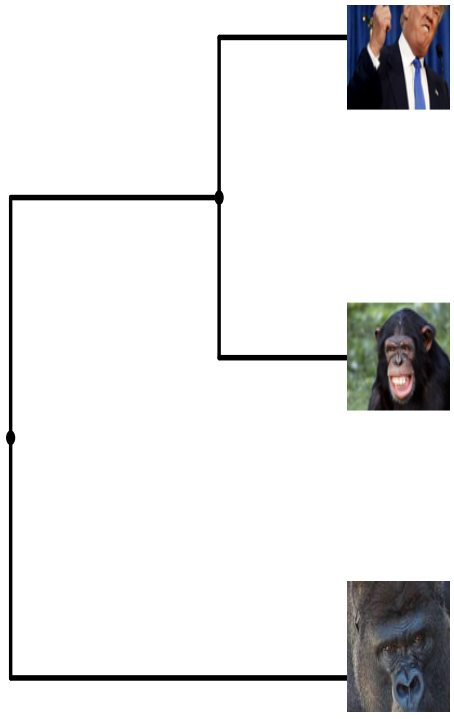
Phylogenies*!

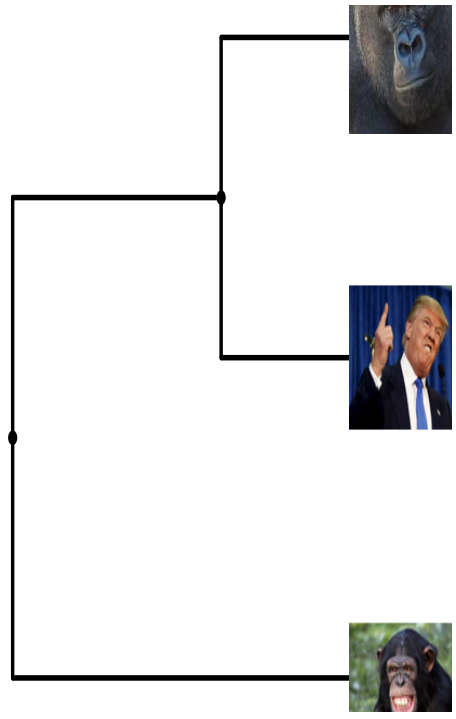
* plus complicated models



Trees are hypotheses



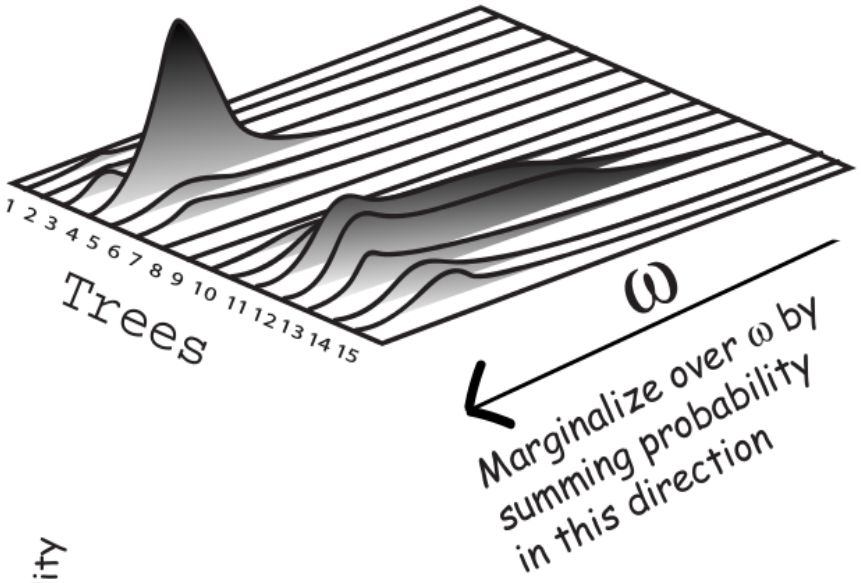


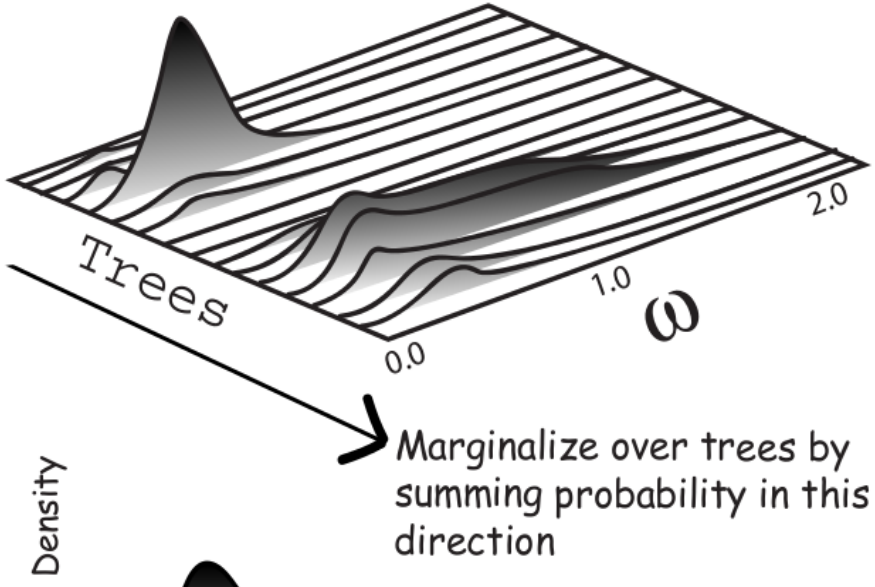


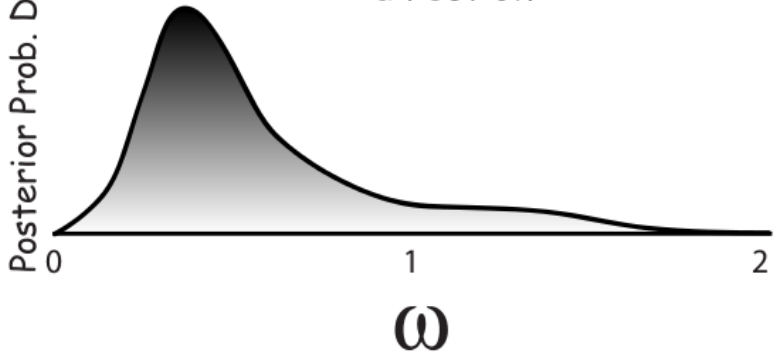
The gist of Bayesian phylogenetics

Bayesian paradigm:

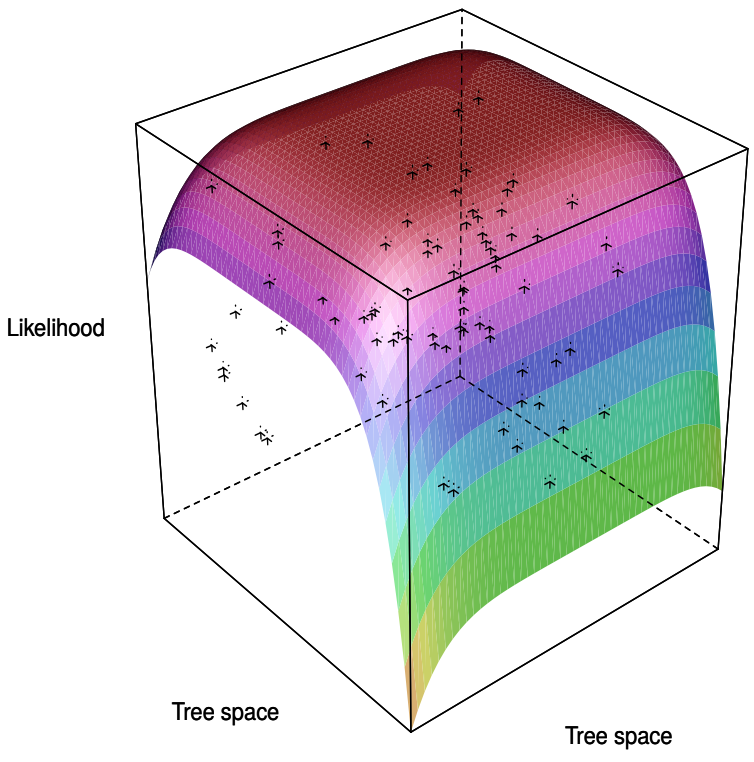
Marginalise (integrate), not maximise

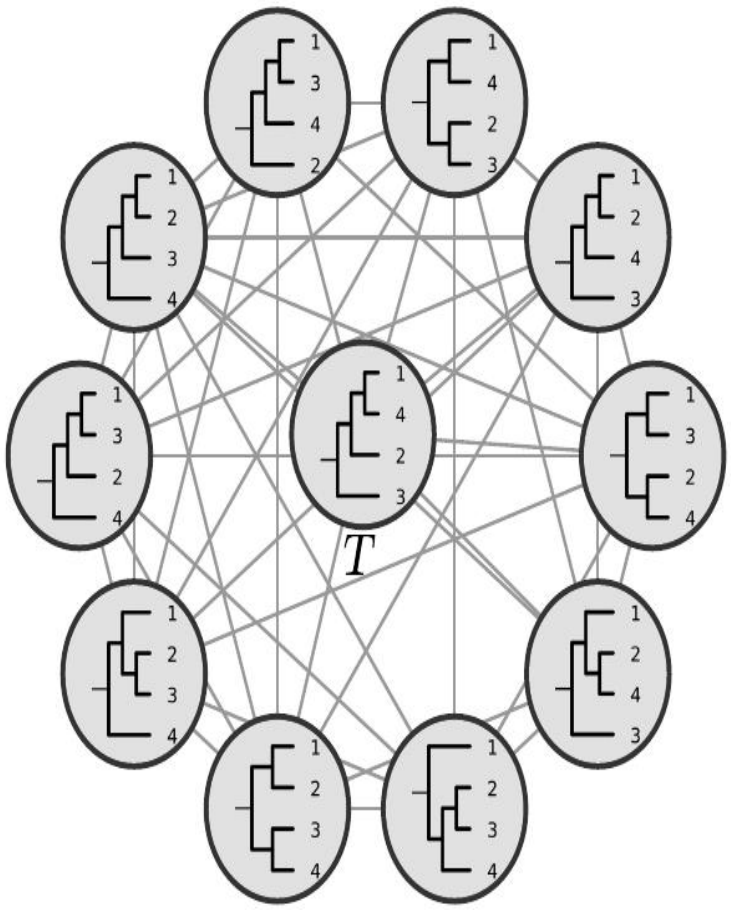






Tree space: a strange land





Metropolis-Hastings algorithm

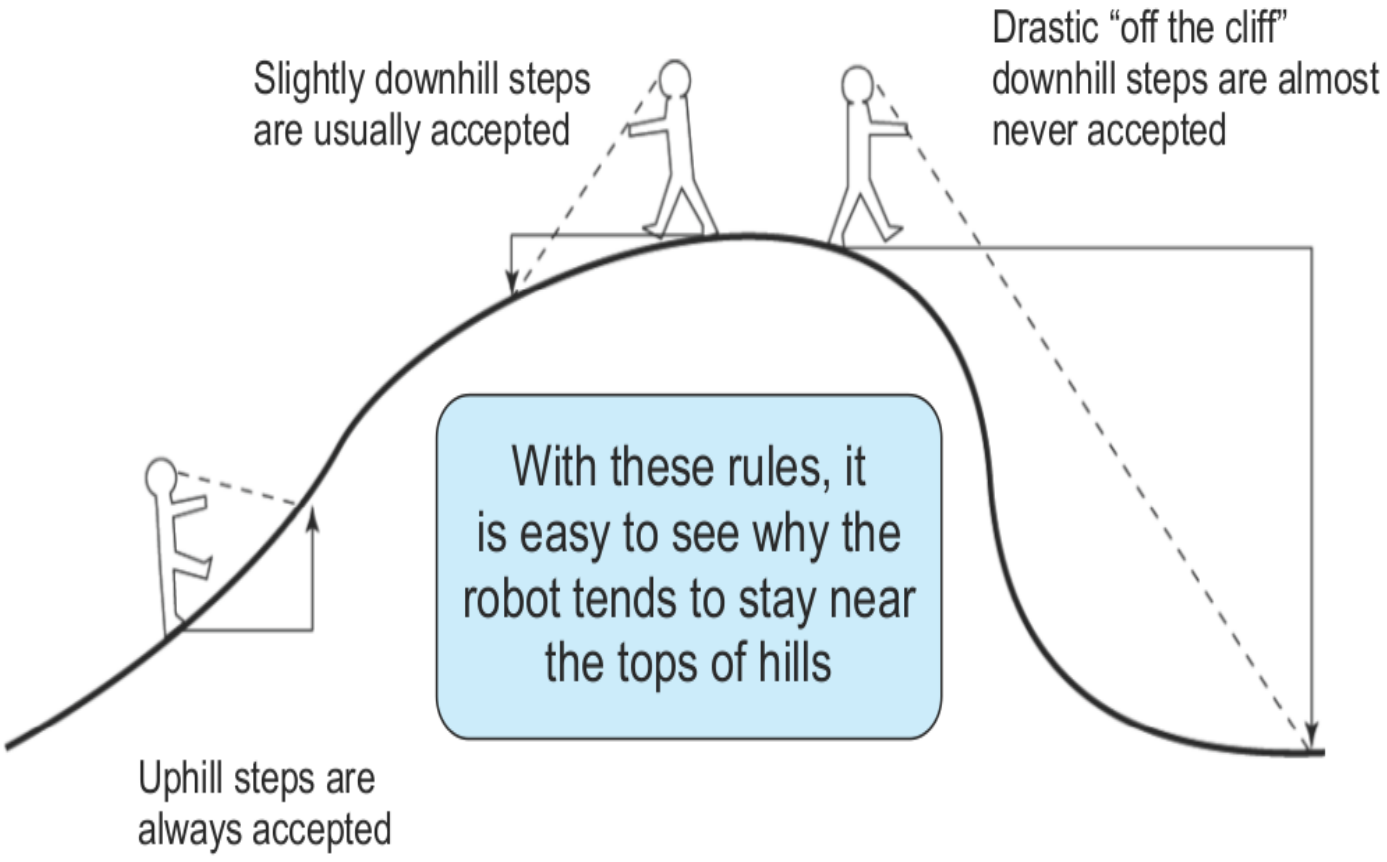
1. Propose new tree: $\lambda \rightarrow \lambda^*$

2. Compute acceptance ratio: $r = \dfrac{P(\text{data}|\lambda^*)\, P(\lambda^*)\, P(\lambda|\lambda^*)}{P(\text{data}|\lambda)\, P(\lambda)\, P(\lambda^*|\lambda)}$

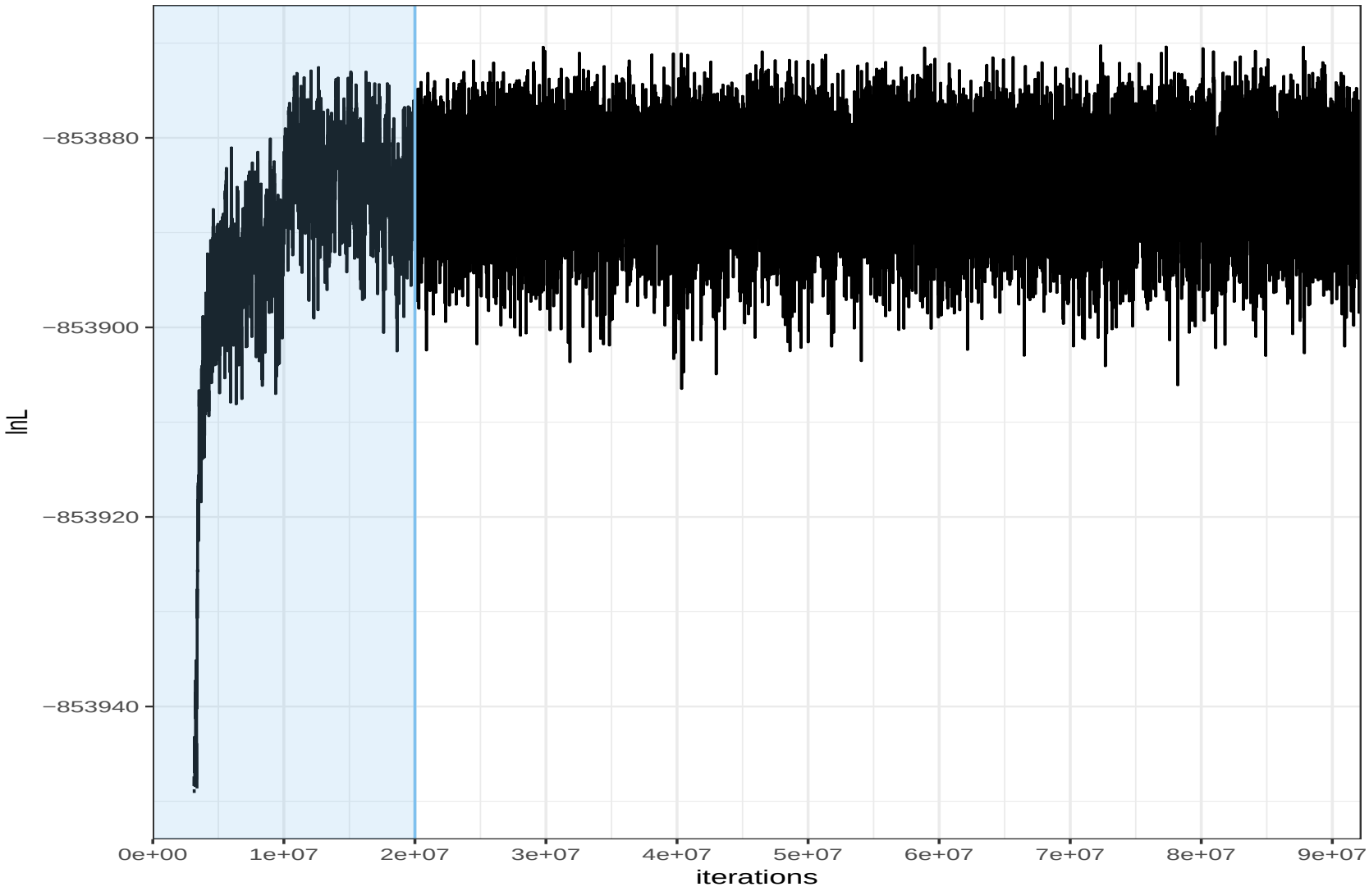
   Likelihood — Prior — Proposal Probability

3. Accept/Reject:
   - if u > r: accept
   - otherwise: reject

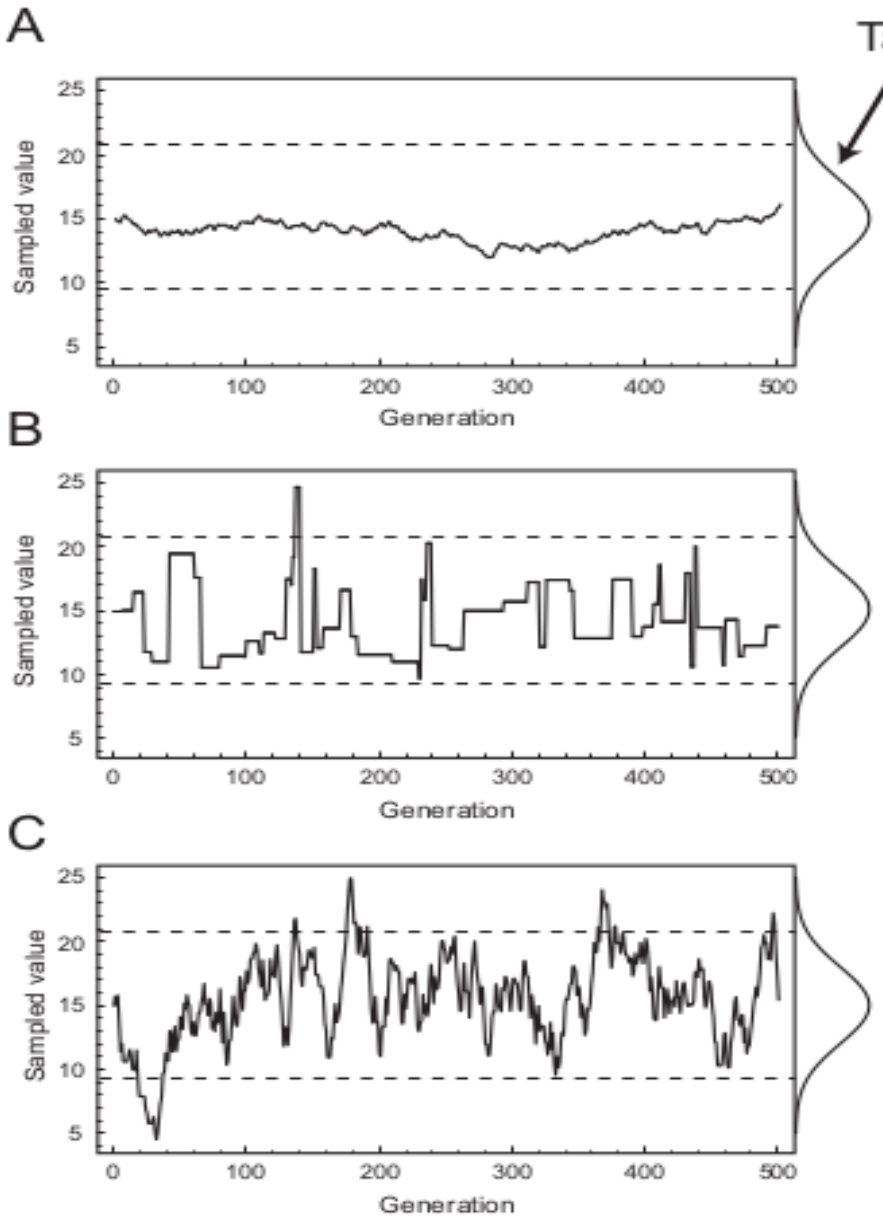
MCMC “robot”


Slightly downhill steps
are usually accepted
Drastic “off the cliff”
downhill steps are almost
never accepted
With these rules, it
is easy to see why the
robot tends to stay near
the tops of hills
Uphill steps are
always accepted

Exploring parameter space: **burn-in**



Exploring parameter space: **mixing**

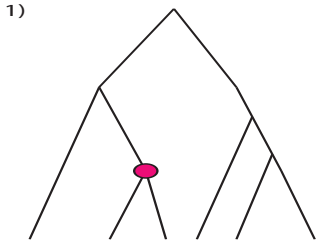


Too modest proposals
Acceptance rate too high
Poor mixing

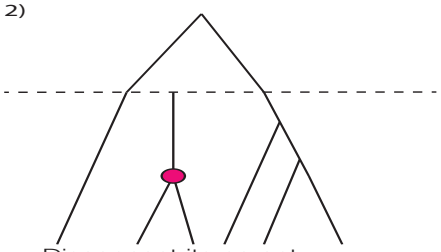
Too bold proposals
Acceptance rate too low
Poor mixing

Moderately bold proposals
Acceptance rate intermediate
Good mixing

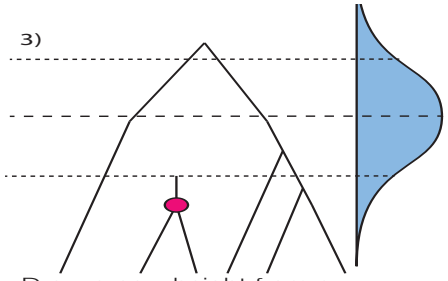
Height-preserving kernels: SubTreeLeap



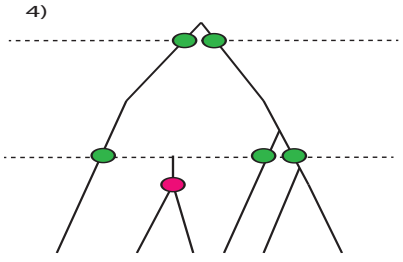
Pick a node



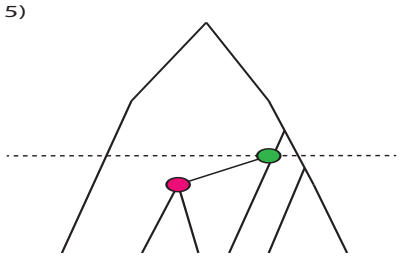
Disconnect its parent



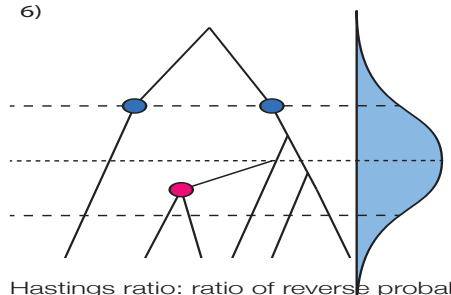
Draw a new height from a normal centred on old height of parent. Also consider the symmetrical height above or below the old height.



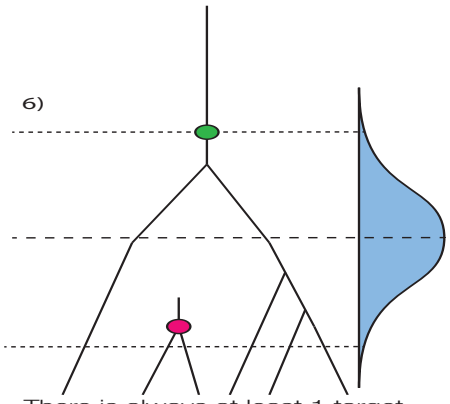
Pick uniformally from branches subtending that height and the symmetrical height above or below (in this case 5).



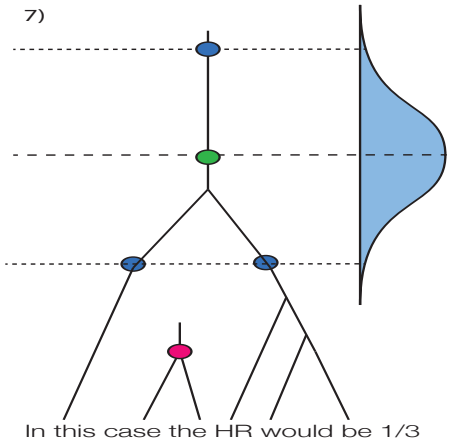
Attach parent to the chosen location.



Hastings ratio: ratio of reverse probability (1 / number of reverse locations, i.e., 1/2) to forwards probability (i.e., 1/5).
Hastings ratio = 5 / 2

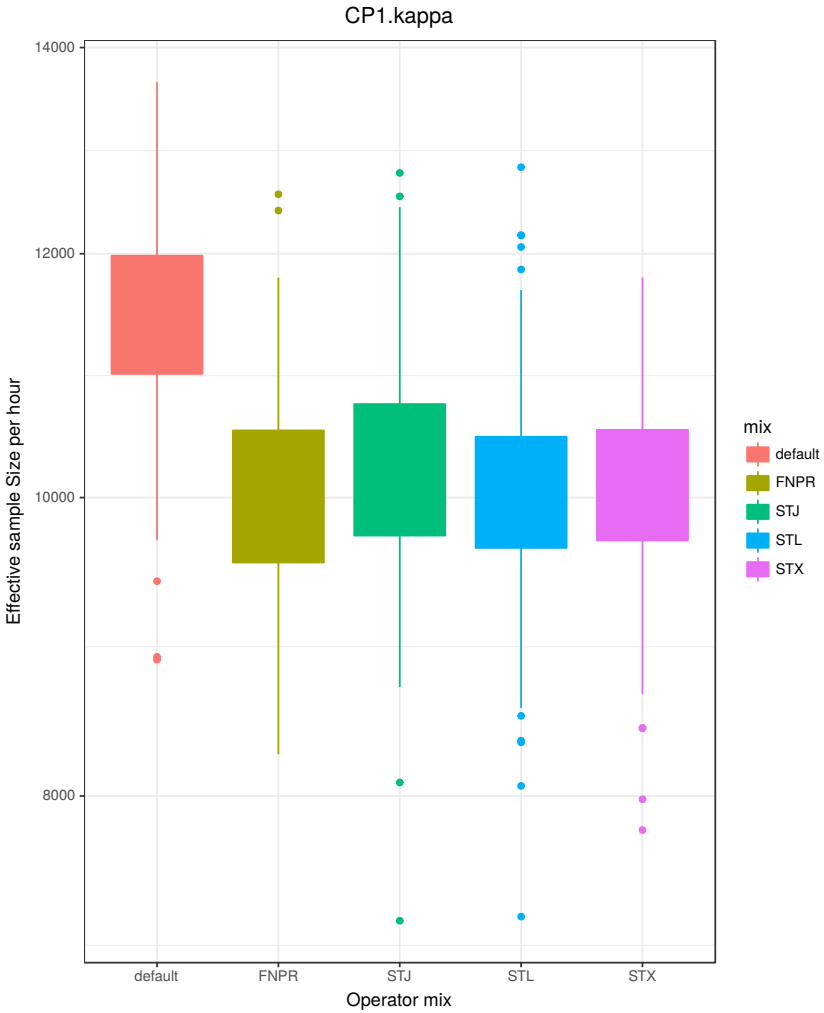


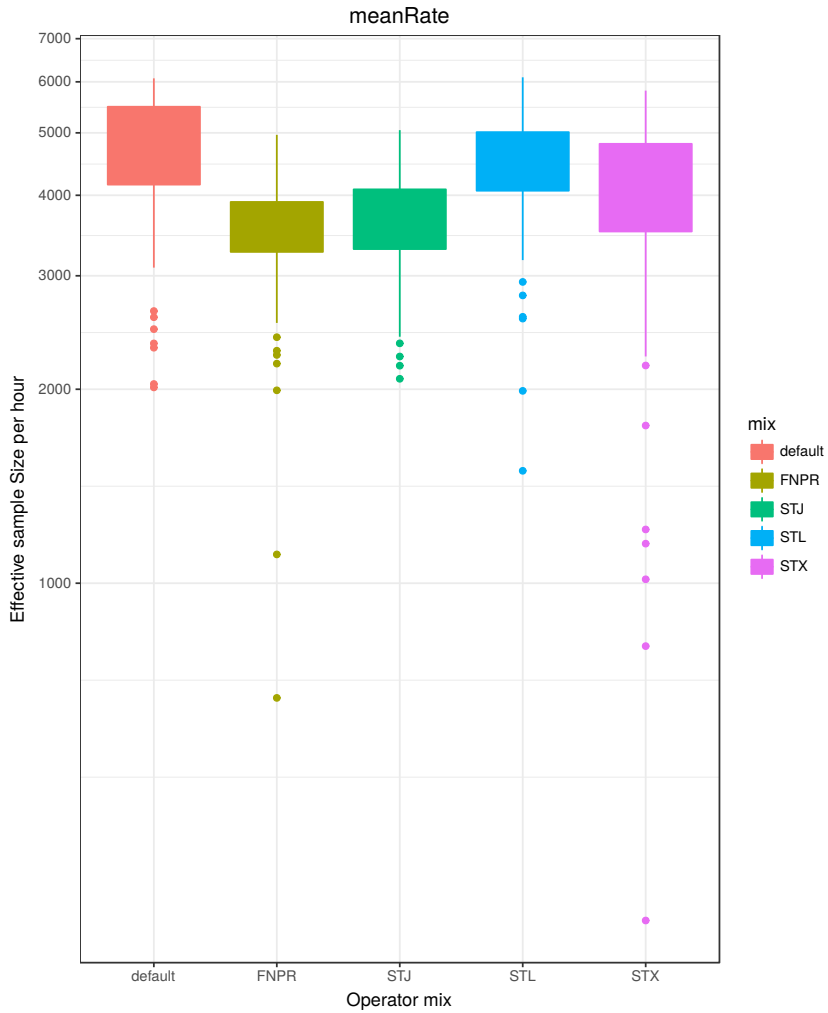
There is always at least 1 target location (above the root).



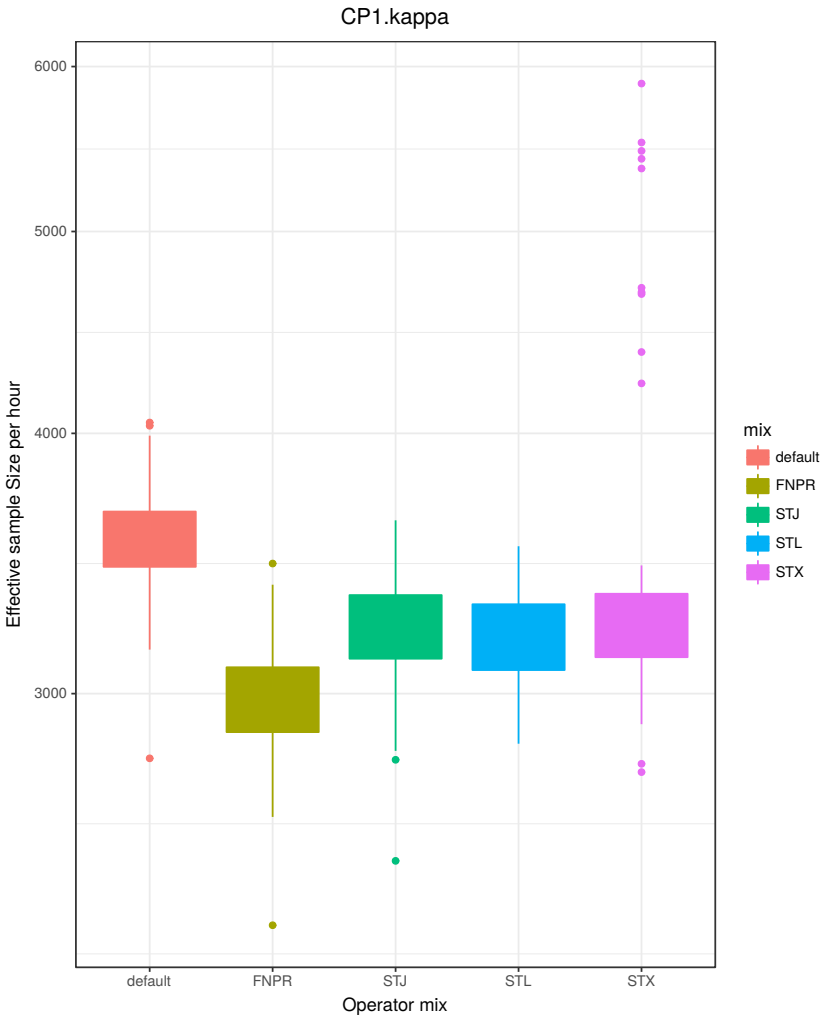
In this case the HR would be 1/3

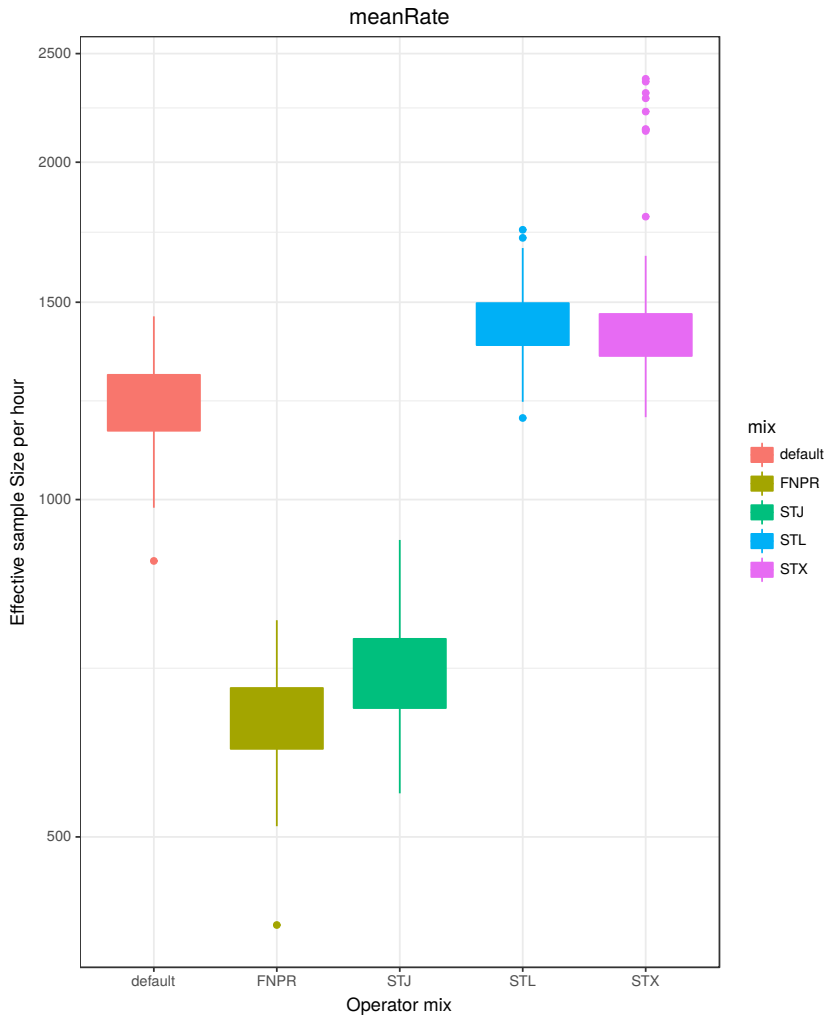
Dengue 4 *env* (17 taxa, 1485 sites)



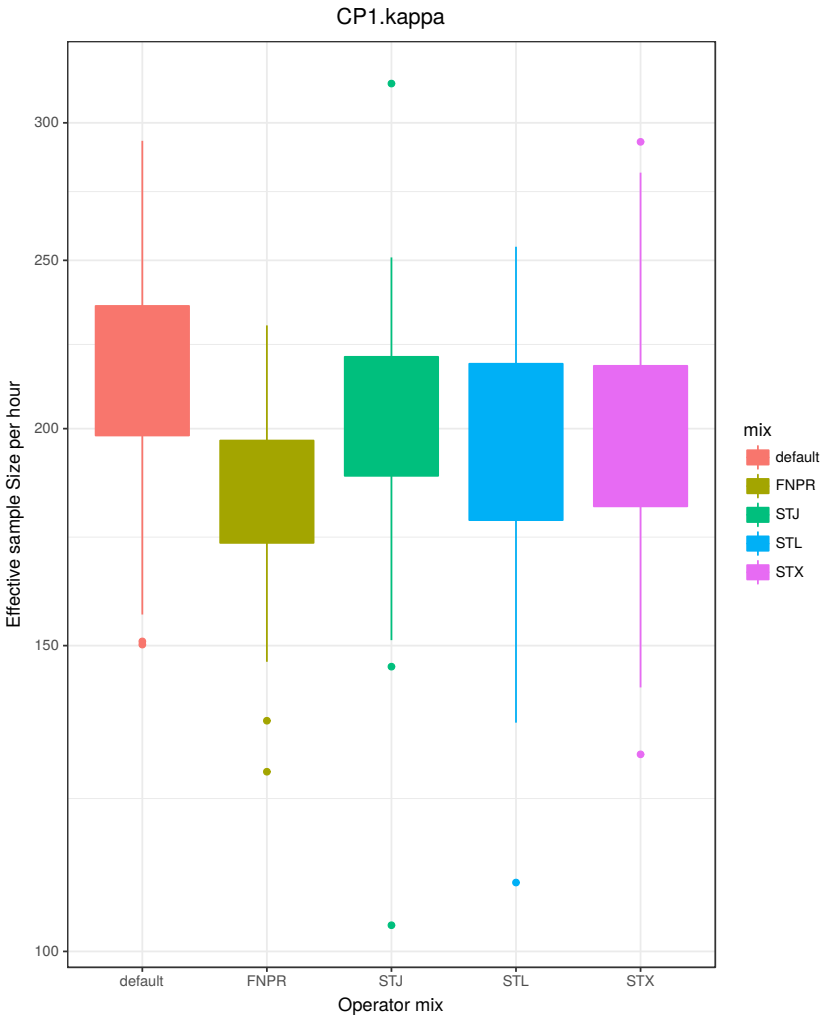


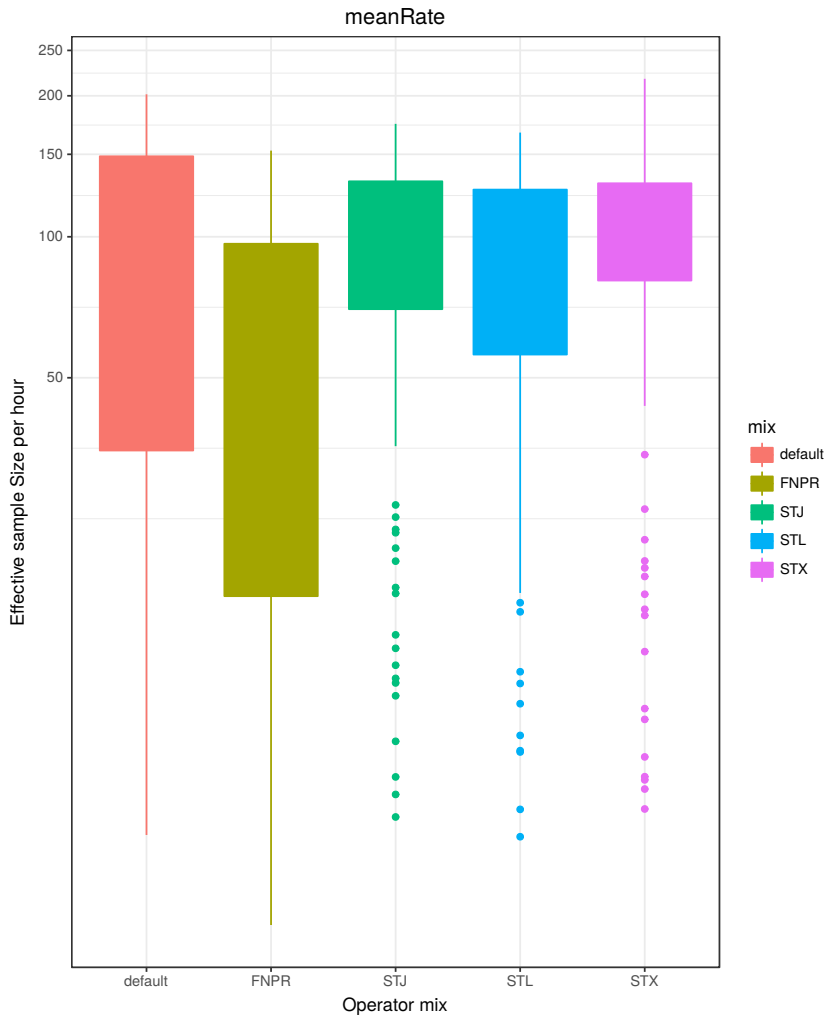
RSVA G protein (35 taxa, 629 sites)



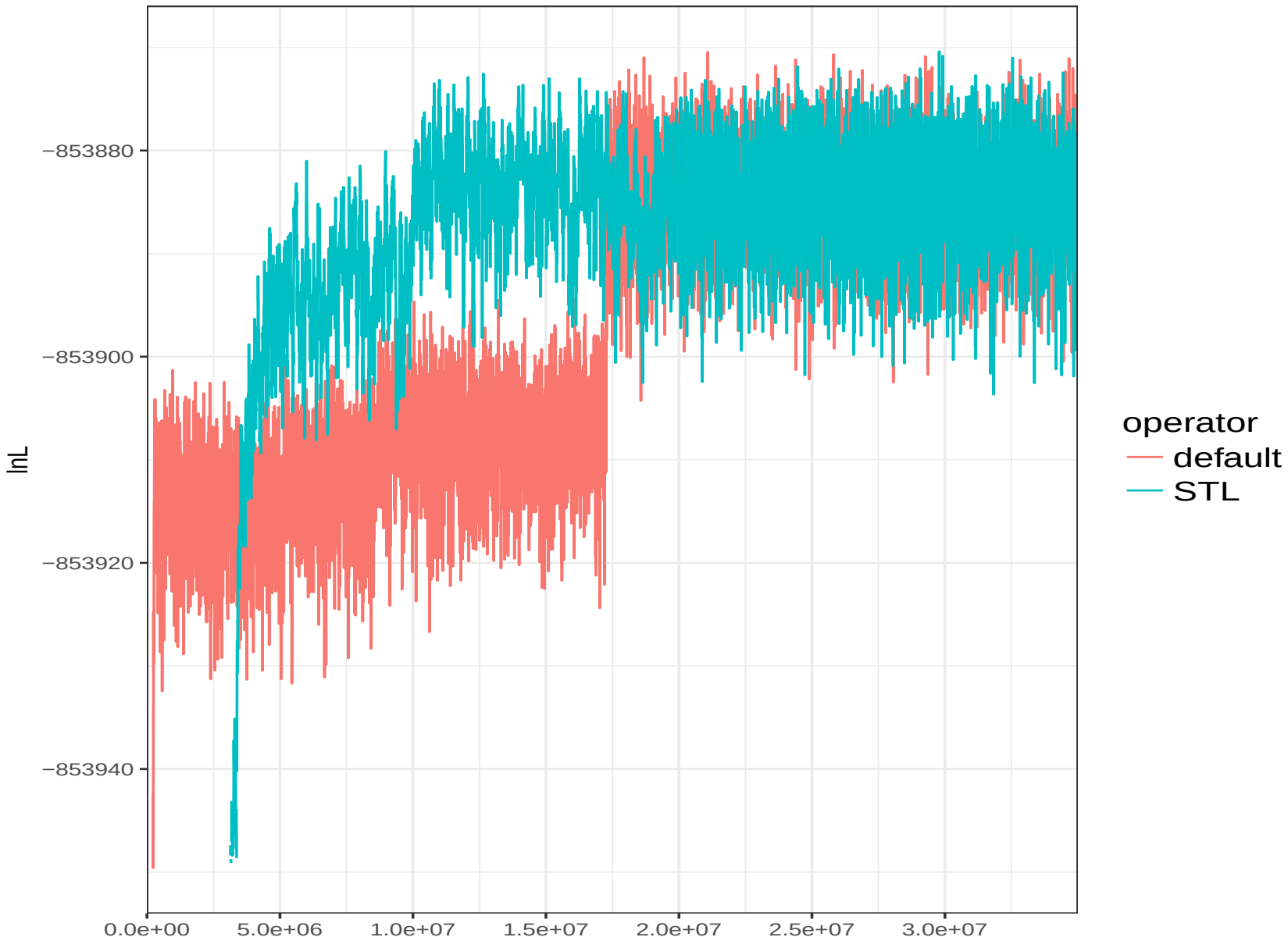


YFV *prM/E* gene (71 taxa, 654 sites)





Metazoans (contemporaneous, 55 taxa, 30257 AA sites)



Take home[1]

Searching trees is hard

Complex, discrete and **HUGE** parameter space

[1] this talk is available online

Take home[1]

Searching trees is hard

Complex, discrete and **HUGE** parameter space

Height-preserving tree rearrangements are good

Use the extra information provided by the tip dates

[1] this talk is available online

Take home[1]

Searching trees is hard

Complex, discrete and **HUGE** parameter space

Height-preserving tree rearrangements are good

Use the extra information provided by the tip dates

Tuneable moves are more efficient

Avoid wasting computing power

[1] this talk is available online

Take home[1]

Searching trees is hard

Complex, discrete and **HUGE** parameter space

Height-preserving tree rearrangements are good

Use the extra information provided by the tip dates

Tuneable moves are more efficient

Avoid wasting computing power

Much more work is needed

We should prepare for an era of plenty

[1] this talk is available online

THE END