Introduction to MrBayes

Fred(rik) Ronquist

Dept. Bioinformatics and Genetics

Swedish Museum of Natural History, Stockholm, Sweden

Installing MrBayes

- Two options:
  - Go to mrbayes.net, click **Download** and follow the instructions to download the current release (3.2.6)
  - OR go to https://github.com/NBISweden/MrBayes/releases to download the prerelease of the next version (3.2.7) (but you have to compile this yourself on your machine)

Running MrBayes

- Use **execute** to bring data in a Nexus file into MrBayes
- Set the model and priors using **lset** and **prset**
- Run the chain using **mcmc**; results in a set of .p and .t files
- The .t files contain tree samples in Nexus format
- The .p files contain tab-delimited samples of the model parameters
- Summarize the parameter samples using **sump**
- Summarize the tree samples using **sumt**

Convergence Diagnostics

- By default, MrBayes performs two independent analyses starting from different random trees (**mcmc nruns=2**)
- Average standard deviation of split frequencies calculated and presented during the run (**mcmc mcmcdiagn=yes diagnfreq=5000**) and written to file (.mcmc). Suggested target < 0.05.
- Standard deviation of each clade frequency and potential scale reduction for branch lengths calculated with **sumt**
- Potential scale reduction calculated for all substitution model parameters with **sump**
- Other tools: **Awty** and **Tracer**

Files output by MrBayes

During an mcmc run

myfile.mcmc	Mcmc run diagnostics
myfile.run1.p	Parameter samples
myfile.run2.p	- ” -
myfile.run1.t	Tree samples
myfile.run2.t	- “ -

After sump

myfile.pstat	Parameter statistics
myfile.lstat	Model likelihood estimates

After sumt

myfile.con.tre	Consensus tree (FigTree fmt by def)
myfile.trprobs	Sampled trees and their probabilities
myfile.parts	Specification of clades or splits
myfile.tstat	Tree statistics
myfile.vstat	Branch length statistics

MrBayes 3.2 additions

- Use combinations of hard, negative, and partial (backbone) constraints on topologies
- Relaxed clocks and dating
- Multi-species coalescent (the BEST model)
- Model jumping across the GTR subspace
- Estimate Bayes factors using stepping-stone sampling
- New tree proposals
- Faster likelihood calculation

Substitution Model Space

GTR model parameters

$\boldsymbol{\pi}$	state frequencies	$\boldsymbol{\pi} = \{\pi_A, \pi_C, \pi_G, \pi_T\}$
$\mathbf{r}$	exchangeability rates	$\mathbf{r} = \{r_{AC}, r_{AG}, r_{AT}, r_{CG}, r_{CT}, r_{GT}\}$

Model	Rate vector	Restr. Growth Fxn	K
GTR	$\mathbf{r} = \{r_{AC}, r_{AG}, r_{AT}, r_{CG}, r_{CT}, r_{GT}\}$	{1,2,3,4,5,6}	6
HKY	$\mathbf{r} = \{r_{tv}, r_{ti}, r_{tv}, r_{tv}, r_{ti}, r_{tv}\}$	{1,2,1,1,2,1}	2
F81	$\mathbf{r} = \{r, r, r, r, r, r\}$	{1,1,1,1,1,1}	1

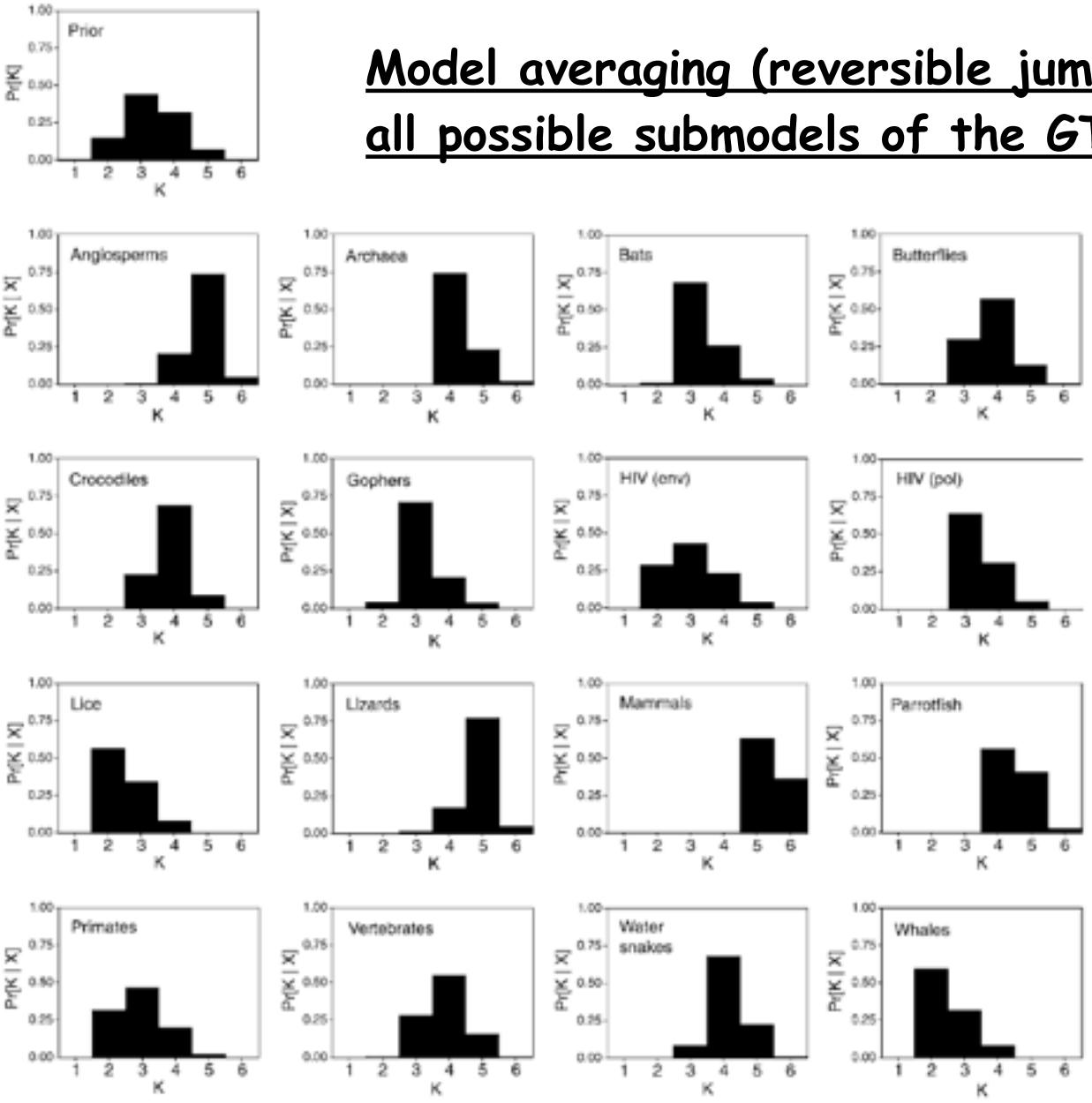
Table 1
All Possible Time-Reversible Models of DNA Substitution

K	Models				
1	$M_1 = 111111$				
2	$M_2 = 122222$	$M_3 = 121111$	$M_4 = 112111$	$M_5 = 111211$	$M_6 = 111121$
	$M_7 = 111112$	$M_8 = 112222$	$M_9 = 121222$	$M_{10} = 122122$	$M_{11} = 122212$
	$M_{12} = 122221$	$M_{13} = 122111$	$M_{14} = 121211$	$M_{15} = 121121$	$M_{16} = 121112$
	$M_{17} = 112211$	$M_{18} = 112121$	$M_{19} = 112112$	$M_{20} = 111221$	$M_{21} = 111212$
	$M_{22} = 111122$	$M_{23} = 111222$	$M_{24} = 112122$	$M_{25} = 112212$	$M_{26} = 112221$
	$M_{27} = 121122$	$M_{28} = 121212$	$M_{29} = 121221$	$M_{30} = 122112$	$M_{31} = 122121$
	$M_{32} = 122211$				
3	$M_{33} = 123333$	$M_{34} = 123222$	$M_{35} = 122322$	$M_{36} = 122232$	$M_{37} = 122223$
	$M_{38} = 123111$	$M_{39} = 121311$	$M_{40} = 121131$	$M_{41} = 121113$	$M_{42} = 112311$
	$M_{43} = 112131$	$M_{44} = 112113$	$M_{45} = 111231$	$M_{46} = 111213$	$M_{47} = 111123$
	$M_{48} = 122333$	$M_{49} = 123233$	$M_{50} = 123323$	$M_{51} = 123332$	$M_{52} = 123322$
	$M_{53} = 123232$	$M_{54} = 123223$	$M_{55} = 122332$	$M_{56} = 122323$	$M_{57} = 122233$
	$M_{58} = 121333$	$M_{59} = 123133$	$M_{60} = 123313$	$M_{61} = 123331$	$M_{62} = 112333$
	$M_{63} = 112322$	$M_{64} = 112232$	$M_{65} = 112223$	$M_{66} = 123122$	$M_{67} = 123212$
	$M_{68} = 123221$	$M_{69} = 121322$	$M_{70} = 121232$	$M_{71} = 121223$	$M_{72} = 122312$
	$M_{73} = 122321$	$M_{74} = 122132$	$M_{75} = 122123$	$M_{76} = 122231$	$M_{77} = 122213$
	$M_{78} = 123311$	$M_{79} = 123131$	$M_{80} = 123113$	$M_{81} = 121331$	$M_{82} = 121313$
	$M_{83} = 121133$	$M_{84} = 123211$	$M_{85} = 123121$	$M_{86} = 123112$	$M_{87} = 122311$
	$M_{88} = 122131$	$M_{89} = 122113$	$M_{90} = 121321$	$M_{91} = 121312$	$M_{92} = 121231$
	$M_{93} = 121213$	$M_{94} = 121132$	$M_{95} = 121123$	$M_{96} = 112331$	$M_{97} = 112313$
	$M_{98} = 112133$	$M_{99} = 112321$	$M_{100} = 112312$	$M_{101} = 112231$	$M_{102} = 112213$
	$M_{103} = 112132$	$M_{104} = 112123$	$M_{105} = 111233$	$M_{106} = 111232$	$M_{107} = 111223$
	$M_{108} = 112233$	$M_{109} = 112323$	$M_{110} = 112332$	$M_{111} = 121233$	$M_{112} = 121323$
	$M_{113} = 121332$	$M_{114} = 122133$	$M_{115} = 122313$	$M_{116} = 122331$	$M_{117} = 123123$
	$M_{118} = 123132$	$M_{119} = 123213$	$M_{120} = 123231$	$M_{121} = 123312$	$M_{122} = 123321$
4	$M_{123} = 123444$	$M_{124} = 123433$	$M_{125} = 123343$	$M_{126} = 123334$	$M_{127} = 123422$
	$M_{128} = 123242$	$M_{129} = 123224$	$M_{130} = 122342$	$M_{131} = 122324$	$M_{132} = 122234$
	$M_{133} = 123411$	$M_{134} = 123141$	$M_{135} = 123114$	$M_{136} = 121341$	$M_{137} = 121314$
	$M_{138} = 121134$	$M_{139} = 112341$	$M_{140} = 112314$	$M_{141} = 112134$	$M_{142} = 111234$
	$M_{143} = 123344$	$M_{144} = 123434$	$M_{145} = 123443$	$M_{146} = 123244$	$M_{147} = 123424$
	$M_{148} = 123442$	$M_{149} = 122344$	$M_{150} = 122343$	$M_{151} = 122334$	$M_{152} = 123423$
	$M_{153} = 123432$	$M_{154} = 123243$	$M_{155} = 123234$	$M_{156} = 123342$	$M_{157} = 123324$
	$M_{158} = 123144$	$M_{159} = 123414$	$M_{160} = 123441$	$M_{161} = 121344$	$M_{162} = 121343$
	$M_{163} = 121334$	$M_{164} = 123413$	$M_{165} = 123431$	$M_{166} = 123143$	$M_{167} = 123134$
	$M_{168} = 123341$	$M_{169} = 123314$	$M_{170} = 112344$	$M_{171} = 112343$	$M_{172} = 112334$
	$M_{173} = 112342$	$M_{174} = 112324$	$M_{175} = 112234$	$M_{176} = 123412$	$M_{177} = 123421$
	$M_{178} = 123142$	$M_{179} = 123124$	$M_{180} = 123241$	$M_{181} = 123214$	$M_{182} = 121342$
	$M_{183} = 121324$	$M_{184} = 121234$	$M_{185} = 122341$	$M_{186} = 122314$	$M_{187} = 122134$
5	$M_{188} = 123455$	$M_{189} = 123454$	$M_{190} = 123445$	$M_{191} = 123453$	$M_{192} = 123435$
	$M_{193} = 123345$	$M_{194} = 123452$	$M_{195} = 123425$	$M_{196} = 123245$	$M_{197} = 122345$
	$M_{198} = 123451$	$M_{199} = 123415$	$M_{200} = 123145$	$M_{201} = 121345$	$M_{202} = 112345$
6	$M_{203} = 123456$				

NOTE.—K is the number of substitution types. The named models are M_1, M_{15}, M_{80}, M_{122}, M_{168}, M_{185}, and M_{203}.

All 203 submodels of GTR

Model averaging (reversible jump MCMC) over all possible submodels of the GTR model



Huelsenbeck et al., 2004, MBE

Fixed number of substitution types

lset nst=1

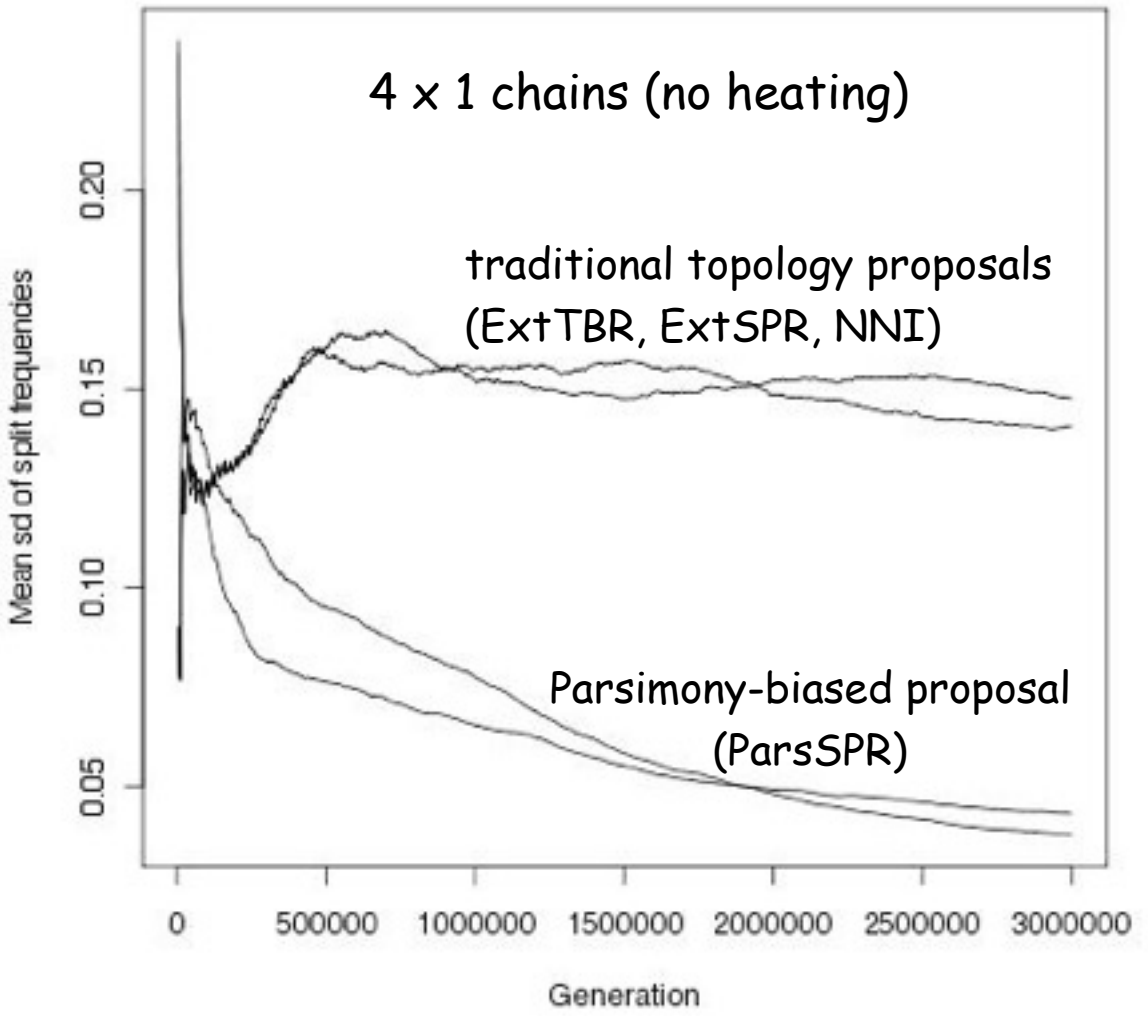
lset nst=2

lset nst=6

Integrate over 203 models

lset nst=mixed

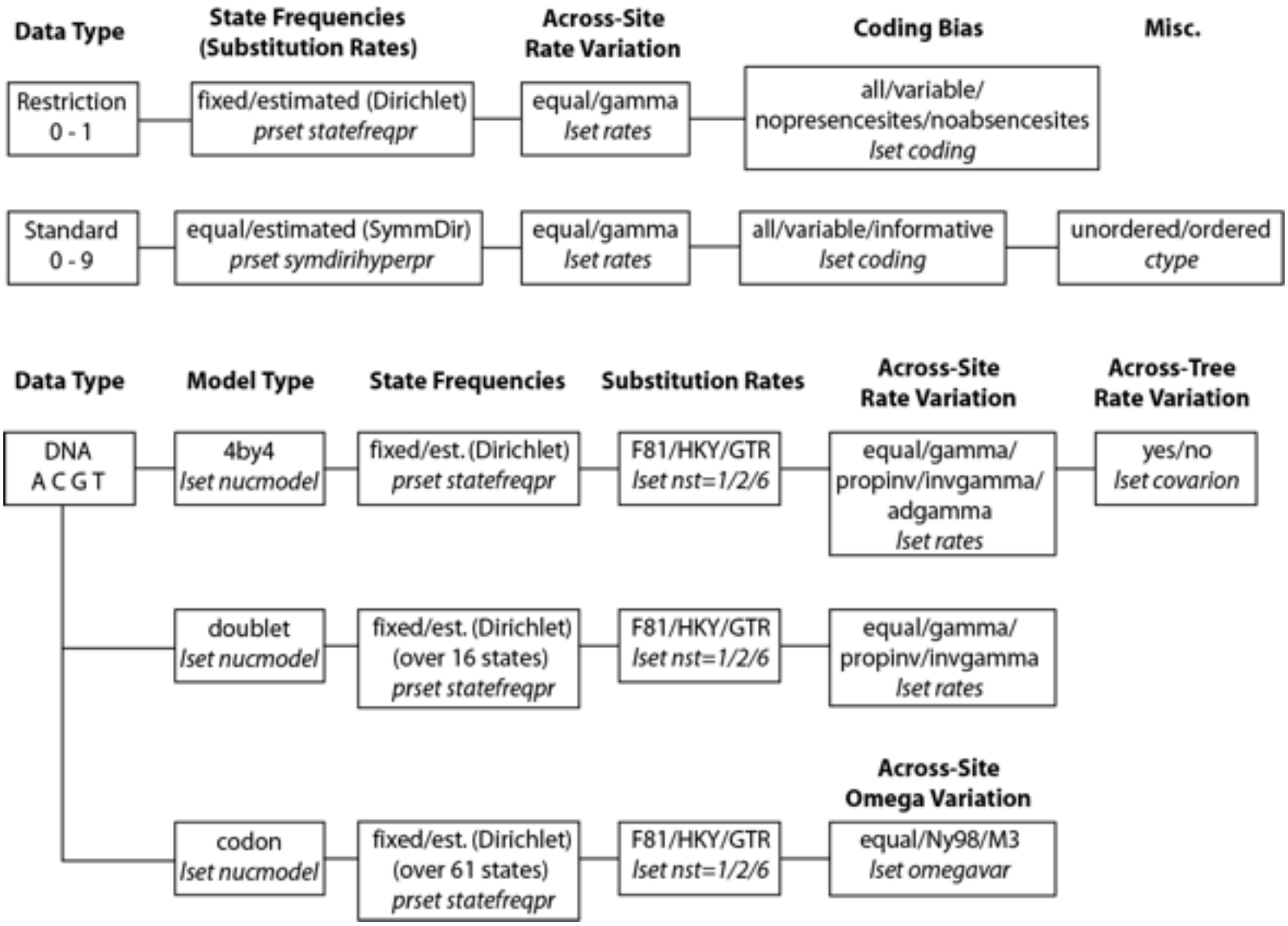
357 taxa, ~3 kb, atpB + rbcL

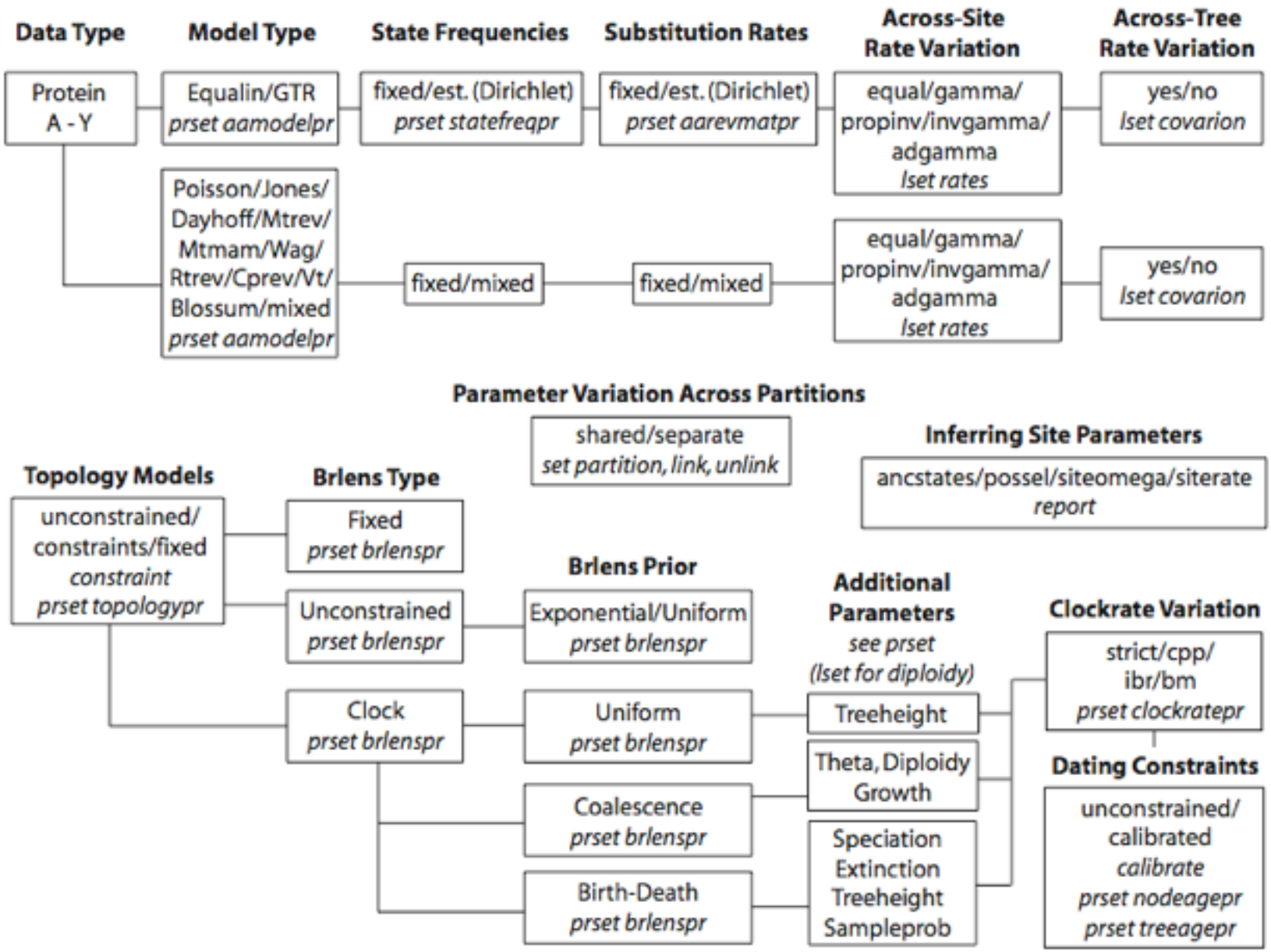


Resources

- MrBayes web site (mrbayes.net)
- Online help in the program (type **help** or **help <command>**)
- MrBayes 3.2 manual with tutorials. The first two tutorials are strongly recommended for beginners: a simple analysis (primates.nex) and an analysis of partitioned data (cynmix.nex).
- Graphical summaries of the MrBayes 3.2 models in the manual (Appendix)
- Books and Papers:
  - Nielsen (ed.) chapter: intro + complex partitioned analysis (kim.nex) (version 3.1)
  - Lemey et al (ed.) chapter: intro + tutorials (version 3.2)
  - Annual Review of Entomology: Review of Bayesian phylogenetics
  - Phylogenetic Trees Made Easy

Models supported by MrBayes 3 (simplified)




Data Type
Model Type
State Frequencies
Substitution Rates
Across-Site Rate Variation
Across-Tree Rate Variation
Protein A - Y
Equalin/GTR
prset aamodelpr
fixed/est. (Dirichlet)
prset statefreqpr
fixed/est. (Dirichlet)
prset aarevmatpr
equal/gamma/ propinv/invgamma/ adgamma
lset rates
yes/no
lset covarion
Poisson/Jones/ Dayhoff/Mtrev/ Mtmam/Wag/ Rtrev/Cprev/Vt/ Blossum/mixed
prset aamodelpr
fixed/mixed
fixed/mixed
equal/gamma/ propinv/invgamma/ adgamma
lset rates
yes/no
lset covarion
Parameter Variation Across Partitions
shared/separate
set partition, link, unlink
Inferring Site Parameters
ancstates/possel/siteomega/siterate
report
Topology Models
unconstrained/ constraints/fixed
constraint
prset topologypr
Brlens Type
Fixed
prset brlenspr
Unconstrained
prset brlenspr
Clock
prset brlenspr
Brlens Prior
Exponential/Uniform
prset brlenspr
Uniform
prset brlenspr
Coalescence
prset brlenspr
Birth-Death
prset brlenspr
Additional Parameters
see prset
(lset for diploidy)
Treeheight
Theta, Diploidy Growth
Speciation Extinction Treeheight Sampleprob
Clockrate Variation
strict/cpp/ ibr/bm
prset clockratepr
Dating Constraints
unconstrained/ calibrated
calibrate
prset nodeagepr
prset treeagepr

Some Advice

- If you use ModelTest or MrModelTest: Do not fix parameters in MrBayes
- Run at least 1,000,000 generations
- Don't worry if average standard deviation of split frequencies (ASDSF) increases in the beginning of the run
- Save time by running the analysis without heating, if it works
- Experiment with MPI, SSE and Beagle code to find the fastest combination on your machine
- If you have difficulties with convergence:
  - Change relative proposal probabilities or tuning parameters
  - If you use heated chains and there are few swaps between chains, try to lower the temperature coefficient
  - Increase the number of heated chains
  - Run the analysis longer
  - Make the model more realistic
  - Start with randomly perturbed good trees