**PART 1**

Q1. Which nucleotide substitution model is the best fit for your data? What are the parameters of this model?

A1.

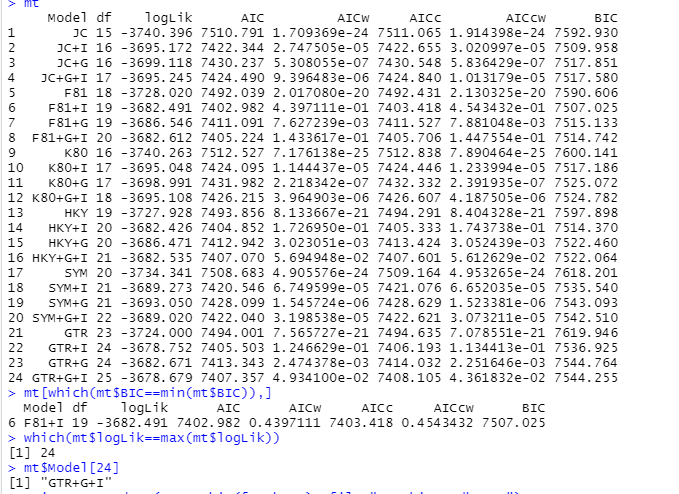


Fig 1: Nucleotide Substitution Models ~ **F81 + I** (our model)

From Fig 1, we can see that the model called Felsenstein 1981 or F81 fits the best to the data provided to us. It has a +I with F81. It contains different equilibrium frequency distributions for each nucleotide but it assumes equal transition - transversion rates. The meaning of +I here means the invariable rates, which means there are some sites which will never mutate (are invariant).

**PART 2**

Q2. Provide a copy of the final model settings/parameters you used in your MrBayes run. You can either copy and paste the text from showmodel, copy the code block from your slurm script, or even provide a screenshot showing one of those things.

A2.

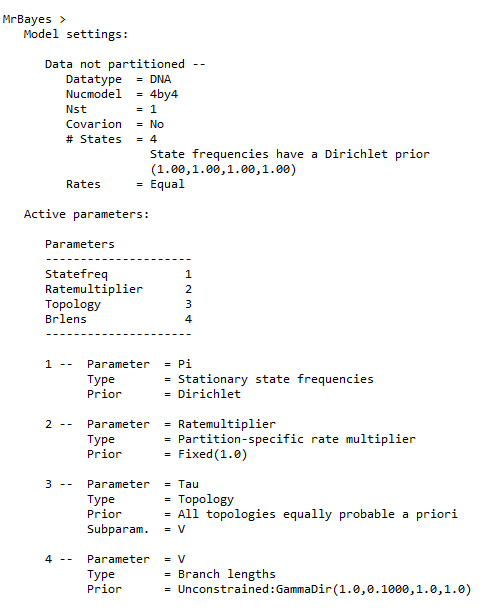


Fig 2A: The default model parameters

Here, we take into account the parameters like state frequency (nstate or nst) and rates as our model has those parameters as peculiarity. So we set the **nst to be 1** (which is default, we don’t need to alter) and the rates as “**propinv**” which means we are considering +I here. To compare the change, Fig 2A and 2B are provided.

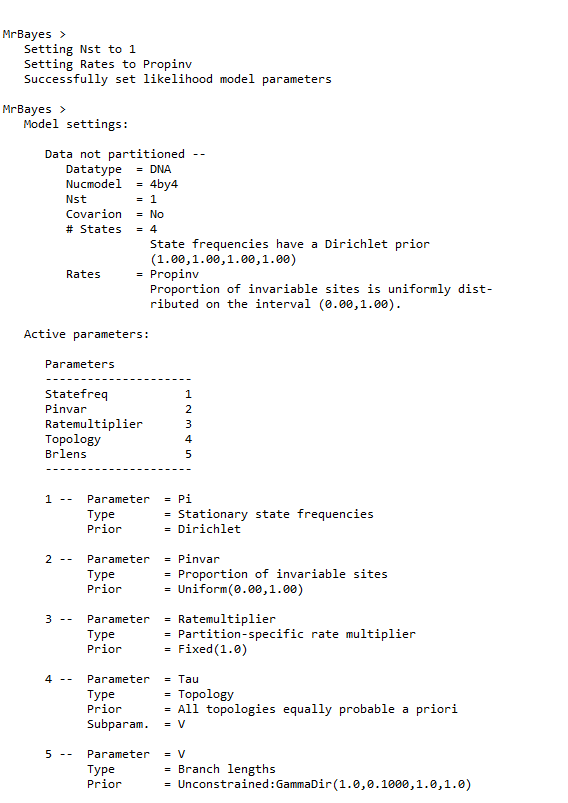


Fig 2B: The changed model parameters as per F81 model

Q3. Do you think that you ran your model long enough to achieve convergence and appropriate levels of mixing? Why or why not? Be sure to provide clear evidence, either in the form of screenshots, text that you have copied and pasted, or plots that you have made in R.

A3.

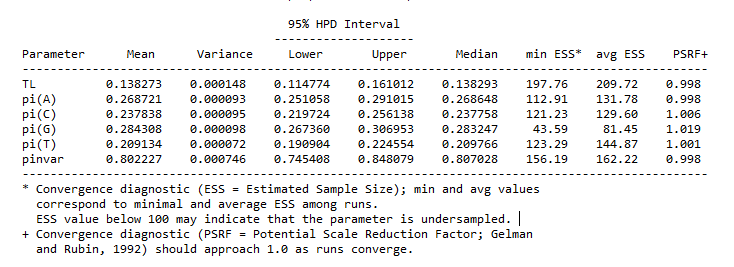


Fig 3A: Convergence diagnostics

We keep in mind 3 things here for convergence: **ESS value, PSRF values and Standard deviation in split frequencies.**

1. We can see here that the PSRF value for all the pi values for all nucleotides is approaching 1. Ideally, it should be less than 1.1.
2. The ESS values for almost all the parameters are above 100, except for the pi(G) value.
3. The standard deviation in split frequencies should be less than 0.01, which we check before stopping the run. For me I did 40,000 iterations and saved every 100 iterations, my value was 0.001 which can be accepted.

For chain mixing: we consider the following image

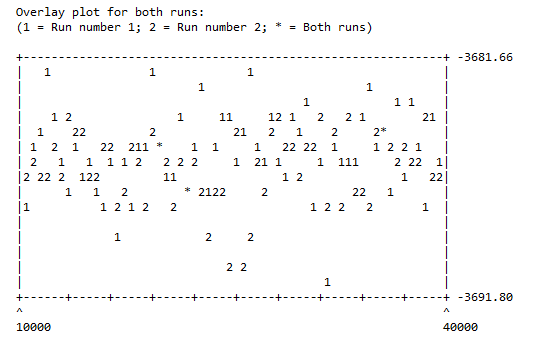


Fig 3B: Chain mixing checking related to convergence.

This data should not have any pattern or trend and it should be **noisy data** to show that values are converged after burnin period and random movement across the space is found.

Hence, we can say that we ran the model long enough to achieve convergence and appropriate levels of mixing.

**PART 3**

Q4. Provide the plot of your tree with support values.

A4.

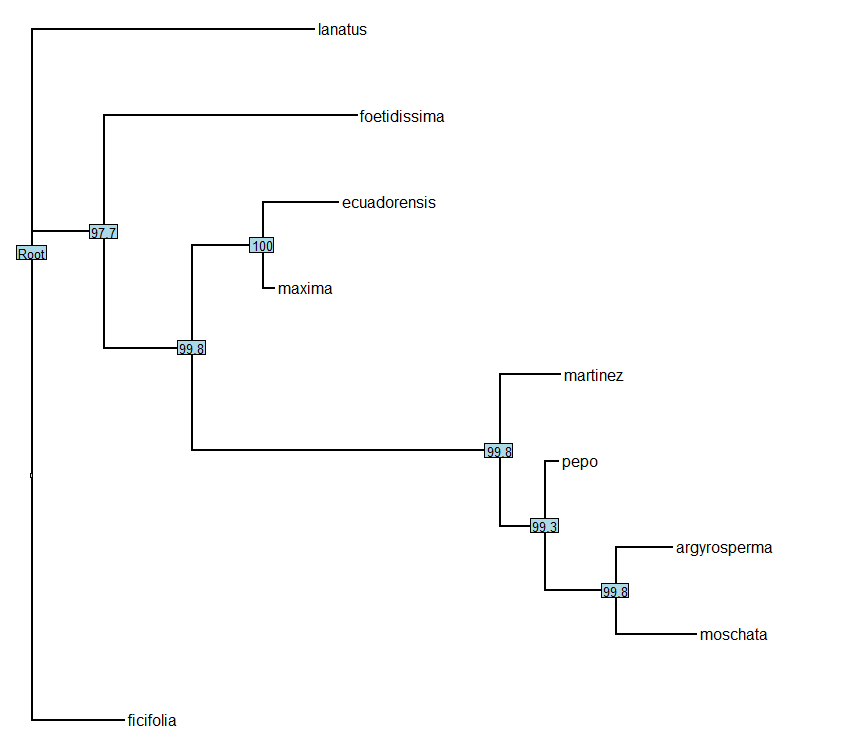


Fig 4: The tree with the support values.

**PART 4**

Q5. Provide the plot of your tree with ancestral states.

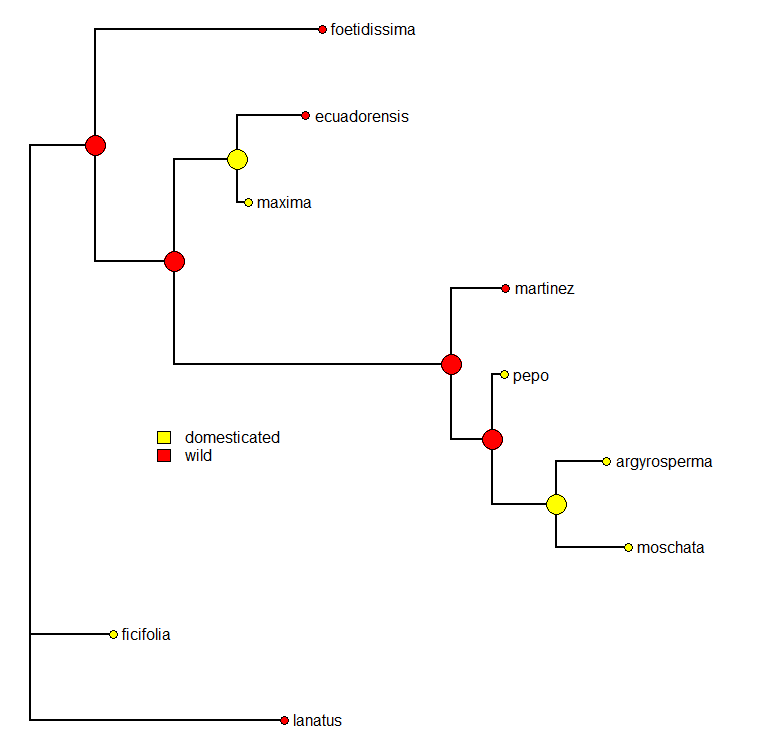
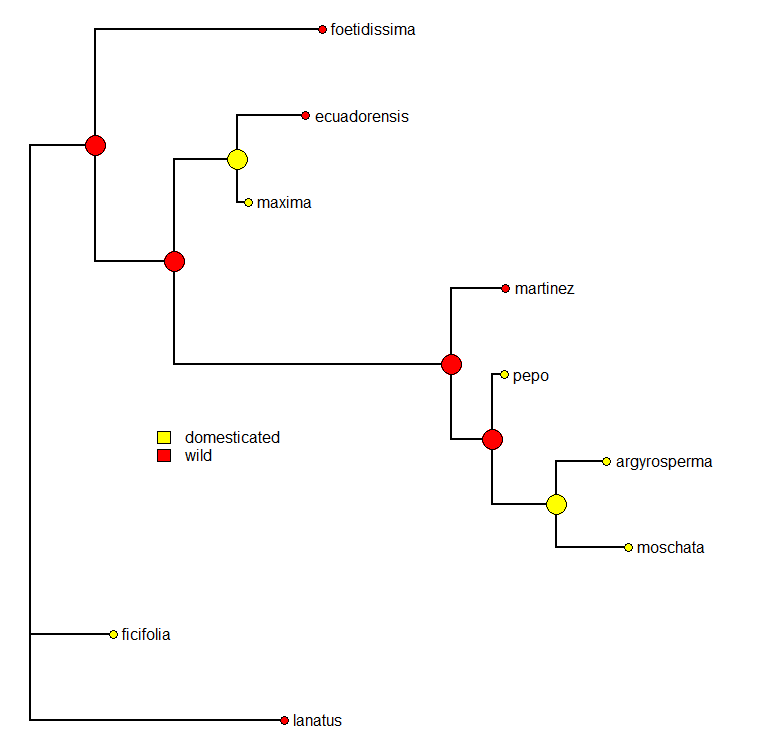
A5. 

Fig 5: Tree with Ancestral States depicted

Q6. How many separate domestication events do you think occurred based on your tree? Be sure to describe which lineages or nodes each domestication event happened on.

A6.



Referring Fig 5 again here, we can see that there are 3 sites from where it is getting domesticated:

1. At the **node of maxima and ecuadorensis**
2. Where it goes to **pepo**
3. At the **node of moschata and argyrosperma**
4. We have an additional one at **ficifolia**

Q7. **Bonus**: Show the plot of your continuous trait evolution tree. Does it look to you like size correlates with evolution, or not really? Be sure to explain your answer.

A7.

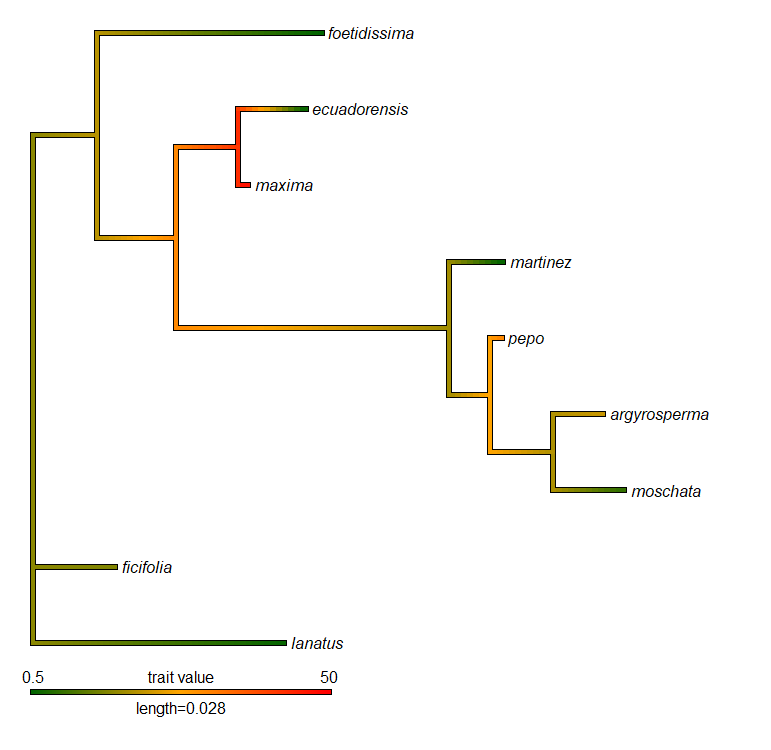


Fig 7: Taking size into account

From the image we understand that the size does vary with time. But when we look closer to the traits of the species, we find that all the domestication occurring doesn’t show increase or decrease in size, ie. no similar trend is being observed. Same goes when we talk about wild species.