

Alternative figures for the divergence timing manuscript

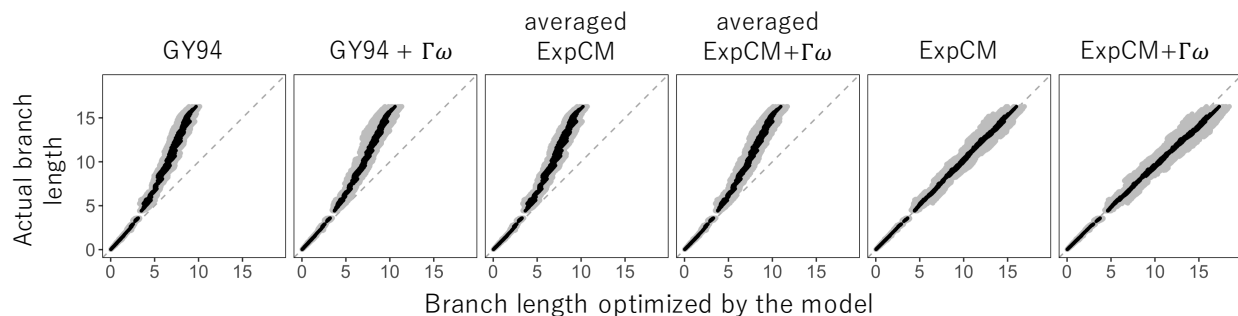
In my In-Prep journal club, I got major comments on three of the figures. I received a lot of minor comments but this summary focuses on the comments related to the ability of the reader to interpret the information rather than small, aesthetic changes. Here are the old figures and the possible new figures based on the suggestions.

Simulations

The major comment on the simulations figure was that the labels didn't make the most intuitive sense. Specifically, having the model labels on the top of the figure and having the true branch length on the y-axis. I made both changes below.

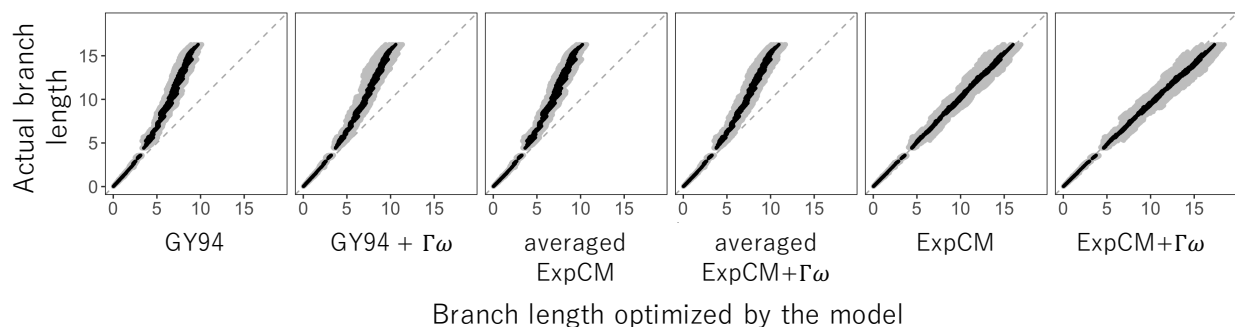
Original figure

Here is the version of the simulations figure I had for In-prep JC:



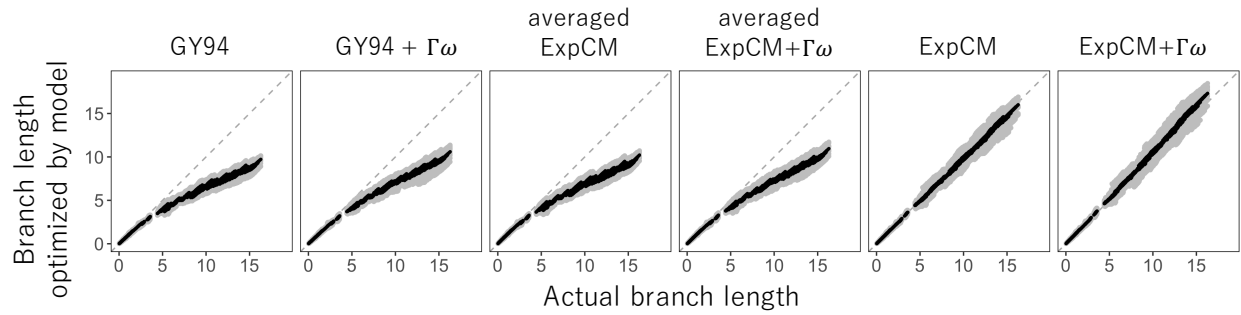
Alt 1

This version keeps the same information on the x/y axes but moving the model labels from above the plots to below the plots:



Alt 2

This version flips the x/y axes so that the “true” branch length is on the x axis and the branch length optimized by the model is on the y axis:



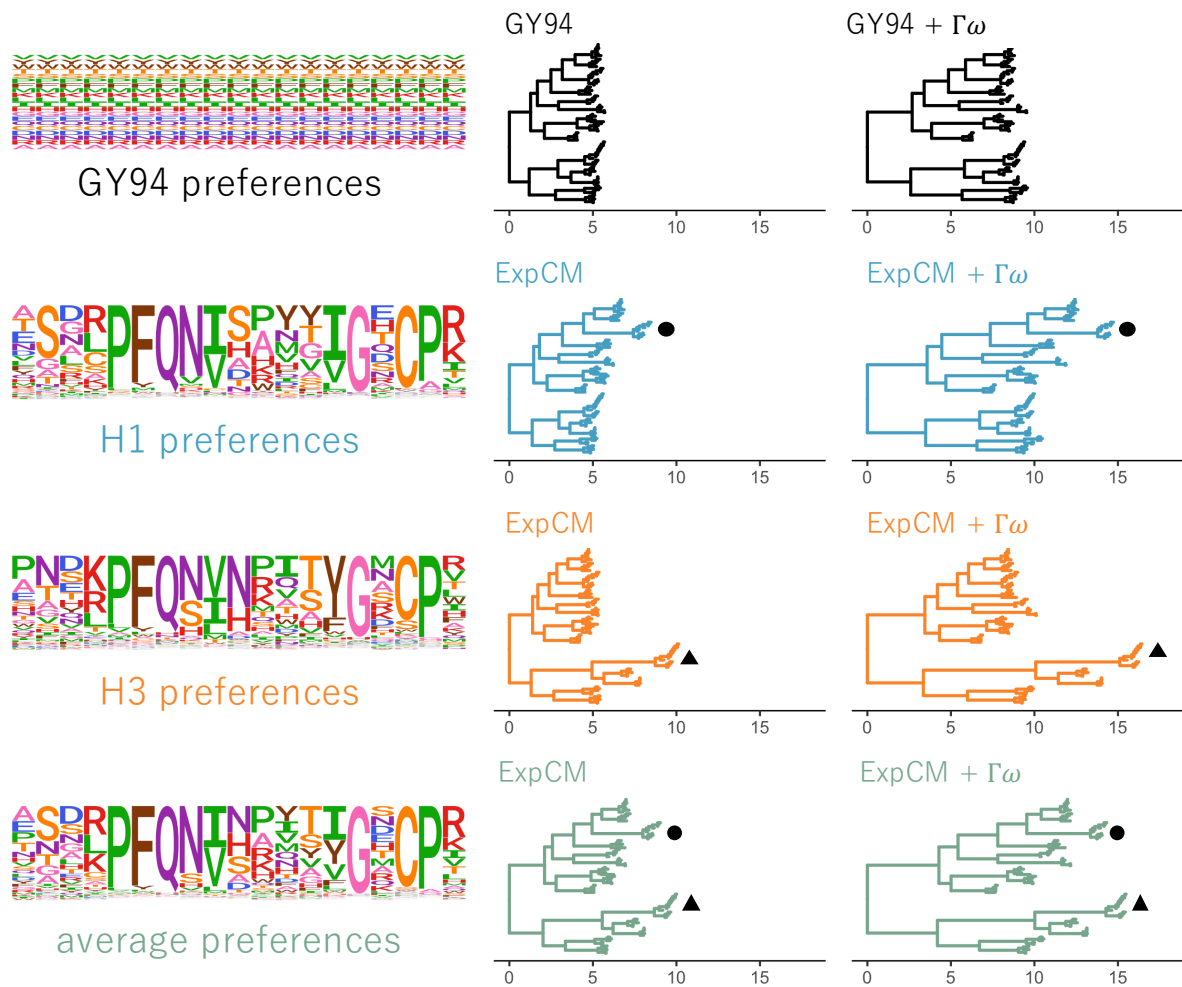
Overall, I see where having the true length on the x-axis is better but I think the labels are more confusing. However, it seems as though any of these versions would be fine as the final version.

Empirical Trees

There were a few minor comments about the empirical trees (should the colors be labeled? Are there too many redundant labels? the labels are not aligned, etc) but the major comment was whether or not it is more appropriate to show mid-point rooted trees post-optimization rather than the mid-point rooted trees pre-optimization

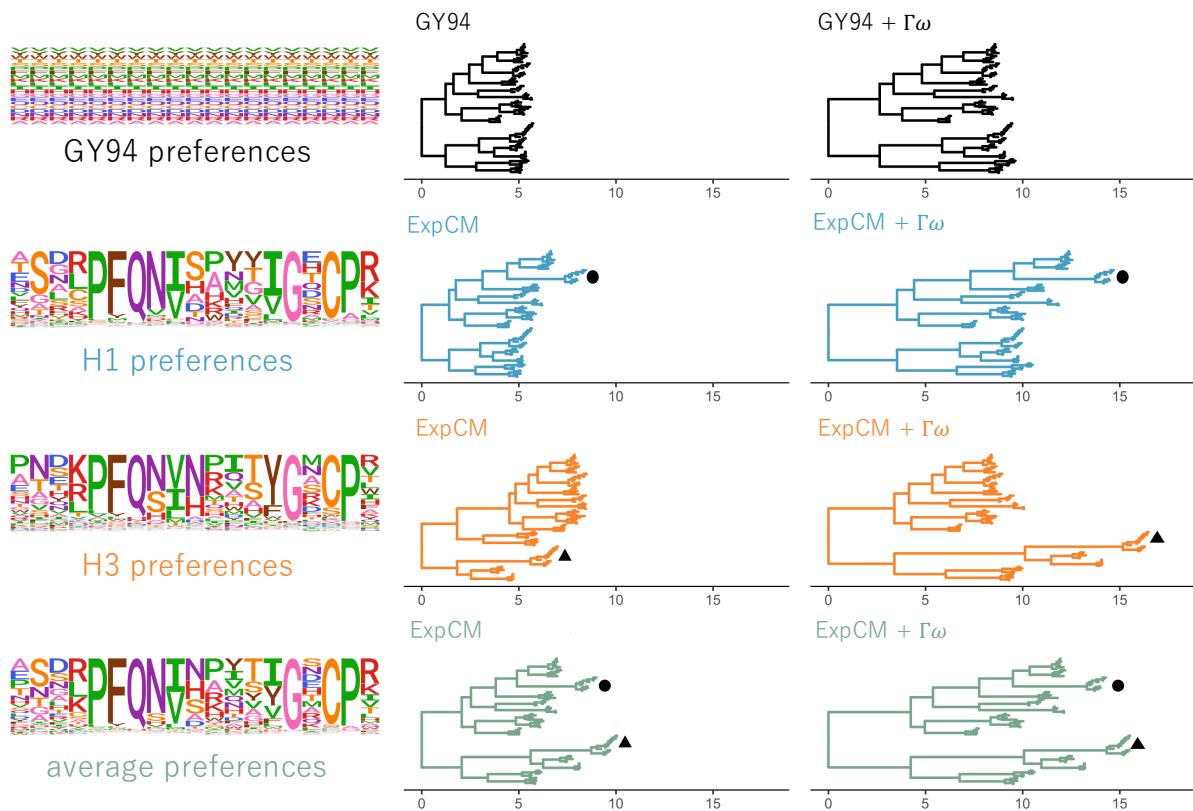
Original figure

Here is the version of the empirical trees figure I had for In-prep JC:



Alt 1

Here is the version where each tree is mid-point rooted:



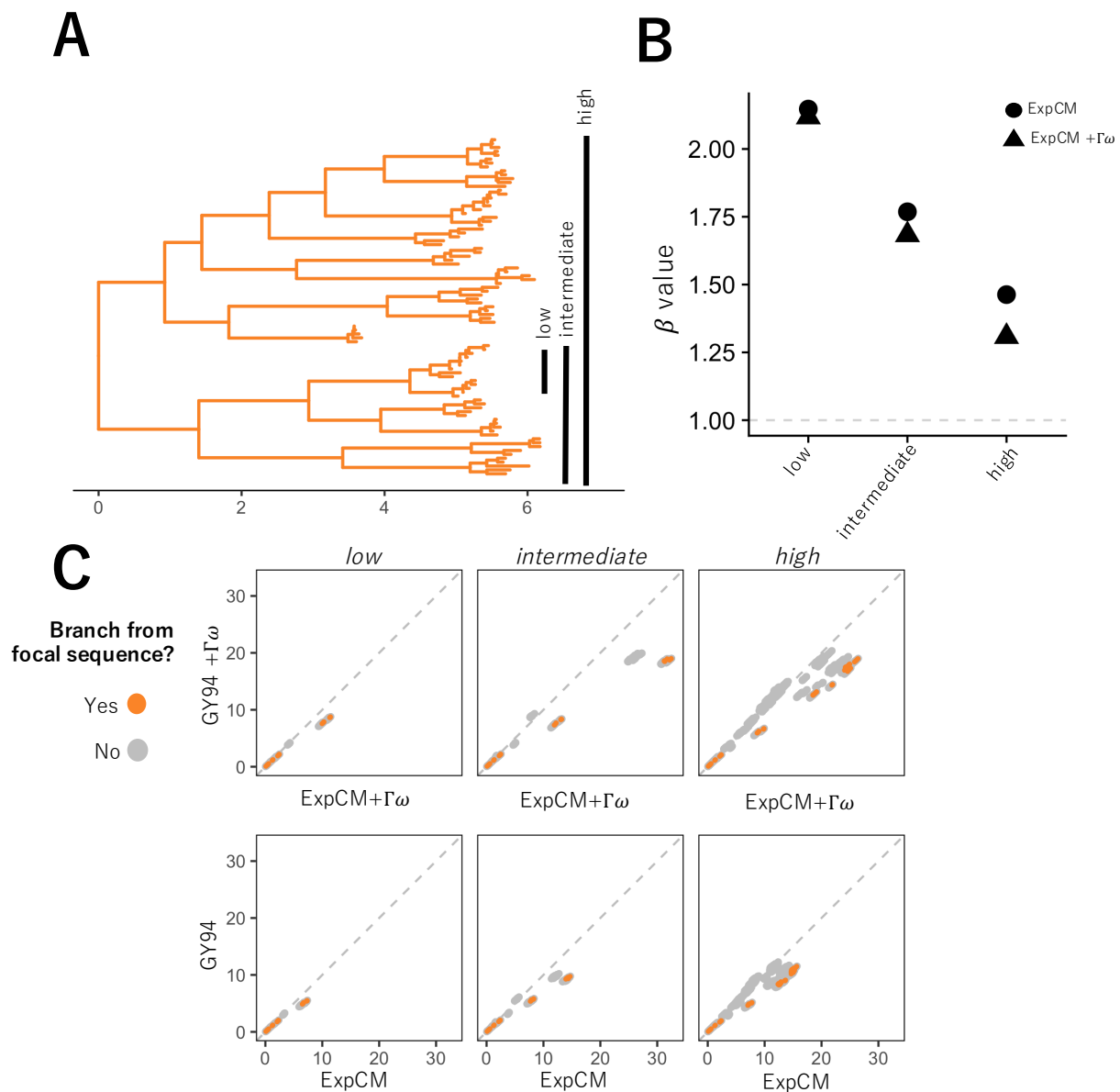
I think the mid-point rooting looks fine except for the ExpCM (H3). If you look closely at the branches it is clear the long branches from Perth have extended in comparison to the YNGKP M0 but I don't think that point is clear from the first impression. In fact, it looks like the Group 1 clade has had more branch length extension than the Group 2 clade. I am guessing this is because there are simply more branches (more overall branch length) in the Group 1 clade compared to the Group 2 clade?

Competing effects

I actually didn't really focus on the competing effects figure in my in-prep JC because we were still going back and forth about it. The point we are trying to make is that the site-specific stationary state has the largest effect on long branches and long branches only occur when there is high divergence but the preferences are least relevant when there is high divergence.

Original figure

This was the original competing effects figure:



When this was the working figure I was thinking of making one for ExpCM (H1) and one for ExpCM (H3). I would show one as a main figure and one as a supp. figure.

Alt figure drafts

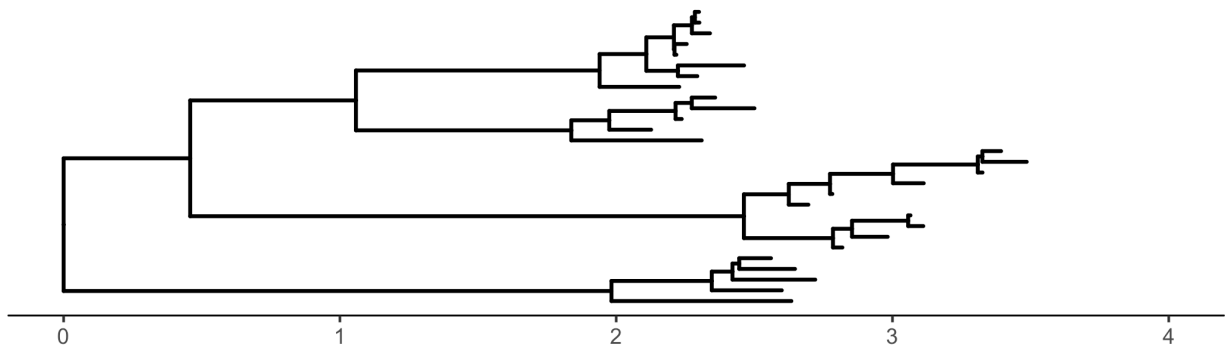
The next idea was to optimize a single tree with an ExpCM, varying only the β value. This would show that as the β value decreased the branch lengths also decreased.

Alt figure 1

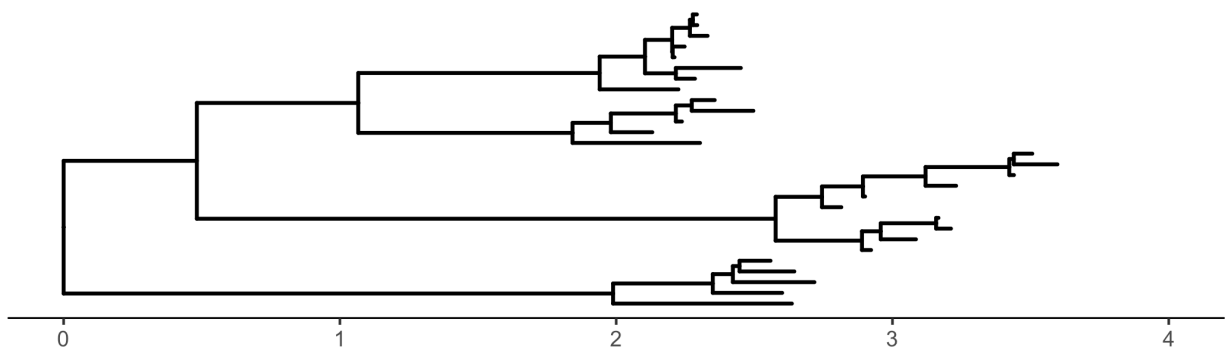
I took the “low divergence” tree and optimized it with the same ExpCM parameters but the β values from

either the “low divergence” (1.56), the “intermediate divergence” (1.34), or the “high divergence” (1.22):

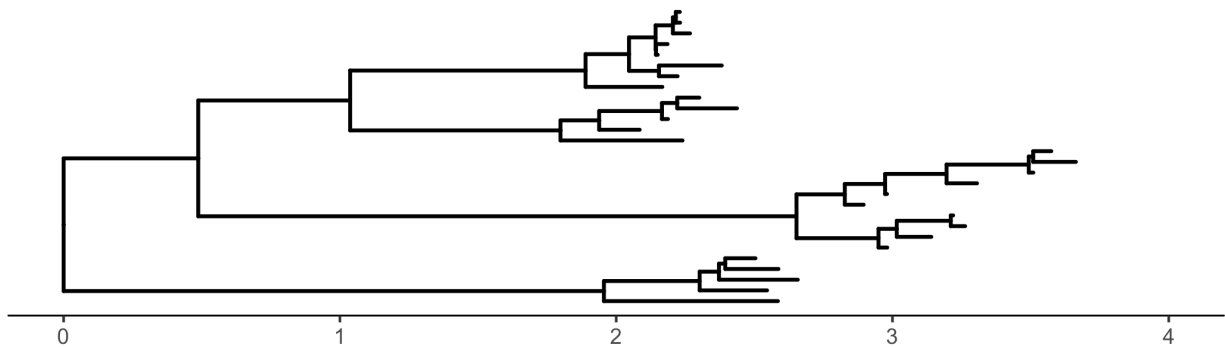
beta_1.22



beta_1.34



beta_1.56

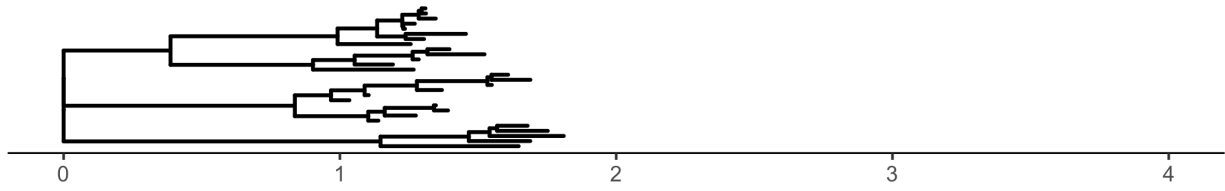


The trees look almost identical. I looked at the raw files and the branch lengths are different, the effect is just too small to see by eye.

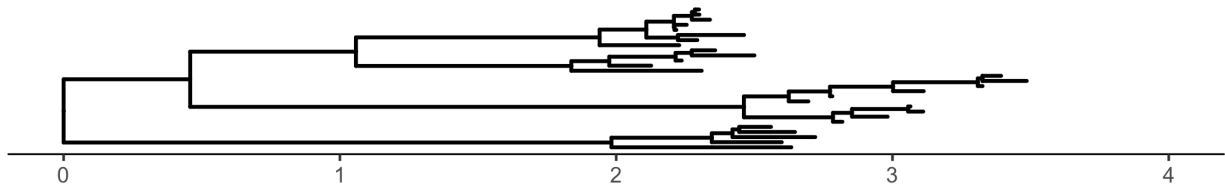
Alt figure 2

I tried two more β values, one on either extreme ($\beta = 0$ and $\beta = 2.5$):

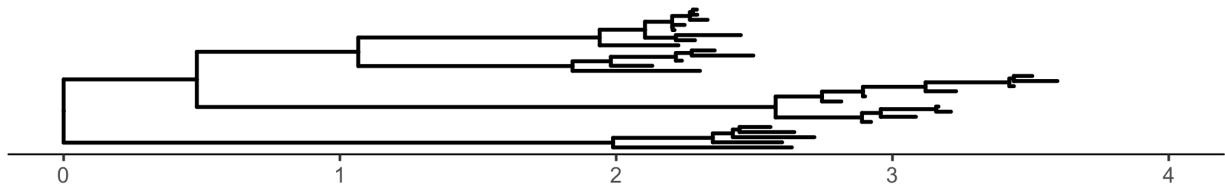
beta_0



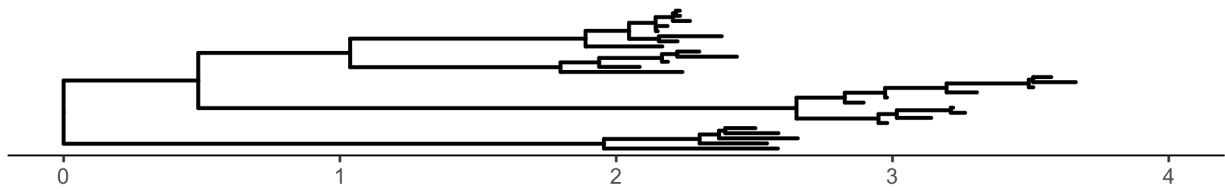
beta_1.22



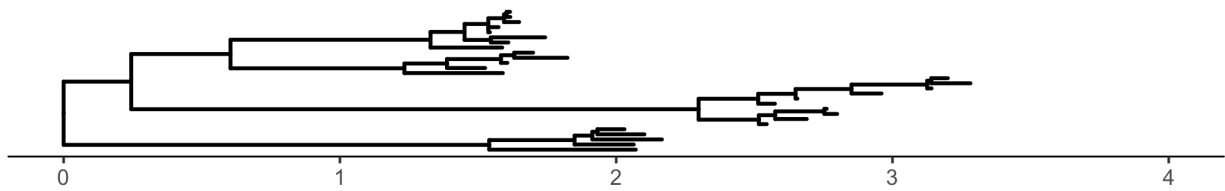
beta_1.34



beta_1.56



beta_2.5



The $\beta = 0$ test worked as expected, the branch lengths are much shorter. However, the $\beta = 2.5$ branches are also shorter. I guess it makes sense that $\beta = 2.5$ doesn't work because $\beta = 1.56$ was the β that was fit originally on this tree.

I came up with a couple of reasons why this might be happening. One might be that the branches are not long enough to see an effect. I could try this on the larger tree but then you run into the problem that the the preferences really do not match. The second is that the optimization is not great when only the branch lengths are being changed.

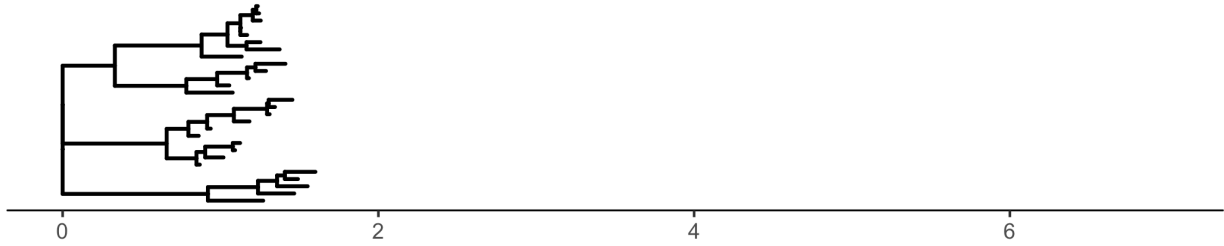
To address to look at these two possibilities I simulated sequences under an ExpCM and then did a similar analysis. The idea being that then we know the preferences really are a good “match” across the entire tree.

Alt figure 3

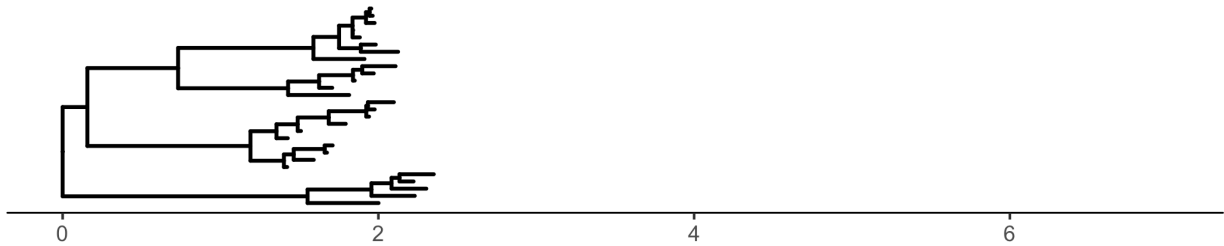
I simulated a sequence using the “low divergence” YNGKP tree and the ExpCM model above except I used

a $\beta = 1.7$. I then re-optimized the branch lengths holding all parameters constant but allowing the branch lengths to be optimized. I used β 1e-05, 1.2, 1.7, and 2.0:

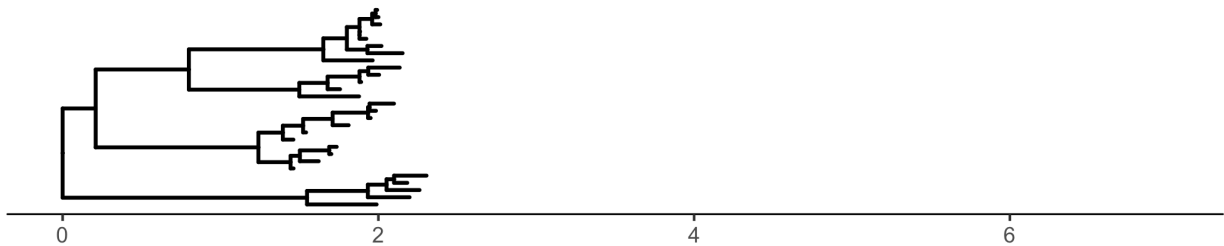
beta_0



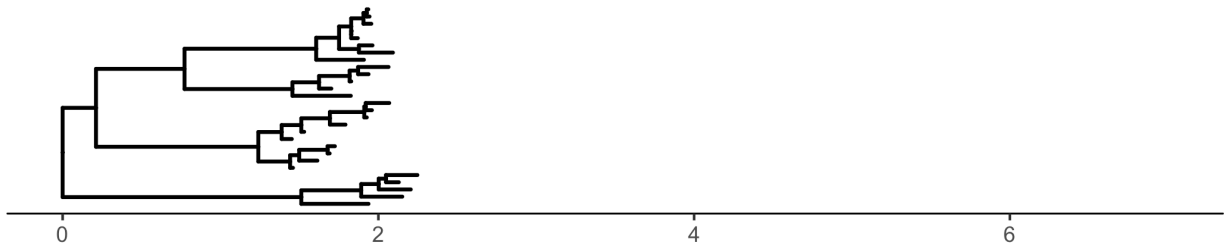
beta_1.2



beta_1.7



beta_2.0



The branch lengths are much shorter when $\beta = 0$ but you can't really tell the other three β s apart.

I don't think that any of these trees do a better job explaining the competing effects of long branches and shifting preferences. It might makes the most sense to stick with the original figure?

Render as pdf

```
pandoc alt_figures.md --latex-engine=xelatex -o alt_figures.pdf -V geometry:margin=1in
```