What I did:

1. Map sequence from a T2T NBPF paralog, to the NHP assemblies, and remove sequences <= 12500 bp in length, removing alignments of exclusively the VNTR sequence.
   1. Paralog used: NBPF25P
   2. Sequence coordinates:
      1. CHM13 V1.0: chr1:144,688,831-144,708,527
      2. [Browser shot:](i.http://genome.ucsc.edu/cgi-bin/hgTracks?db=hub_2898525_t2t-chm13-v1.0&lastVirtModeType=default&lastVirtModeExtraState=&virtModeType=default&virtMode=0&nonVirtPosition=&position=chr1%3A144688831%2D144708527&hgsid=1209310601_hm3QSCffUp0OPzY7OrL28nHYjBz2)
2. Generate a multiple sequence alignment using mafft (mafft/7.453)
   * 1. Command:

mafft --reorder --maxiterate 1000 --thread 16 ${T2T\_NHP\_NBPF\_fa} > ${MSA\_fa}

1. Clean up MSA:
   1. For timing estimates I didn’t need all nonhuman primate copies, so I removed copies that were superfluous. These extra copies didn’t really contribute to the topology of the tree in any way that would affect human paralog dating (complete stratification into one sub-clade, etc. ).
   2. I also removed paralogs which obviously didn’t align well
   3. Lastly, I trimmed the MSA for all bases which had gaps in >=60% of sequences.
2. Generate phylogeny by maximum likelihood:

iqtree2 -nt AUTO -m MFP -s ${trimmed\_MSA\_fasta} \

-o "{macaque\_sequence\_outgroup}" \

--prefix "{NBPF\_timing\_estimate}" --redo-tree -alrt 1000 -bb 1000

1. Bayesian estimate with BEAST2
   1. Using the trimmed MSA I estimated mutation rate at NBPF using two priors: macaque-human divergence of 25MYA +/- 2MYA, and chimp-bonobo divergence of 1.15 MYA +/- 0.3 MYA.
      1. <https://www.annualreviews.org/doi/pdf/10.1146/annurev.genom.9.081307.164420>
      2. https://www.science.org/doi/10.1126/science.aag2602
2. Estimate NBPF expansions given the 95% confidence interval of mutation rate.