

Evolutionary and Quantitative Genetics

Task 2

Daria Plewa

r0976669

Philippe Lemey

24 October 2023

List of contents

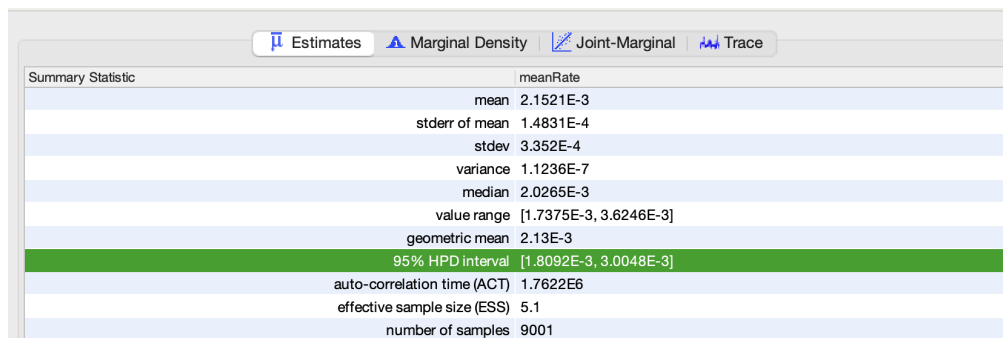
1. Part 1
2. Part 2

Part 1

Perform a BEAST analysis on this data set. For this analysis, specify a GTR substitution model with empirical base frequencies and discrete gamma-distributed rate variation, a relaxed molecular clock with an underlying lognormal distribution, and an exponential growth coalescent prior.

1. Based on the BEAST results, how does your evolutionary rate estimate compare to (i) to the one approximated by TempEst and (ii) to the one reported in Worobey *et al.* (2008)?

The obtained evolutionary rate estimate from BEAST analysis ranges from 1.8092E-3 to 3.0048E-3.



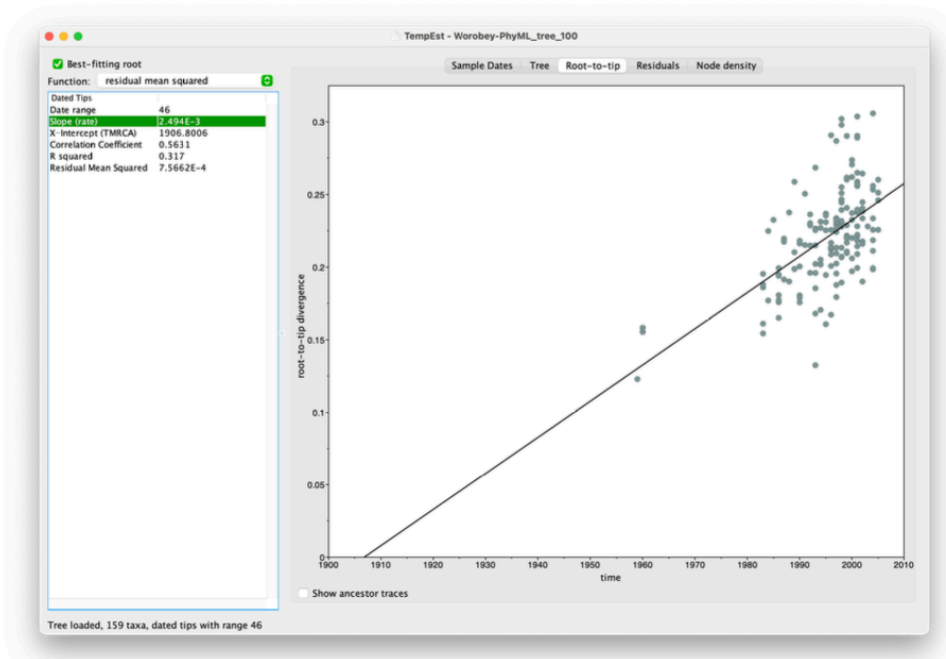
Summary Statistic	meanRate
mean	2.1521E-3
stderr of mean	1.4831E-4
stdev	3.352E-4
variance	1.1236E-7
median	2.0265E-3
value range	[1.7375E-3, 3.6246E-3]
geometric mean	2.13E-3
95% HPD interval	[1.8092E-3, 3.0048E-3]
auto-correlation time (ACT)	1.7622E6
effective sample size (ESS)	5.1
number of samples	9001

The screen of Tracer program with obtained results

The evolutionary rate estimate based on regression analysis from TempEst equals 2.4963E-3.

3. What would be the point estimate of the evolutionary rate based on this regression analysis?

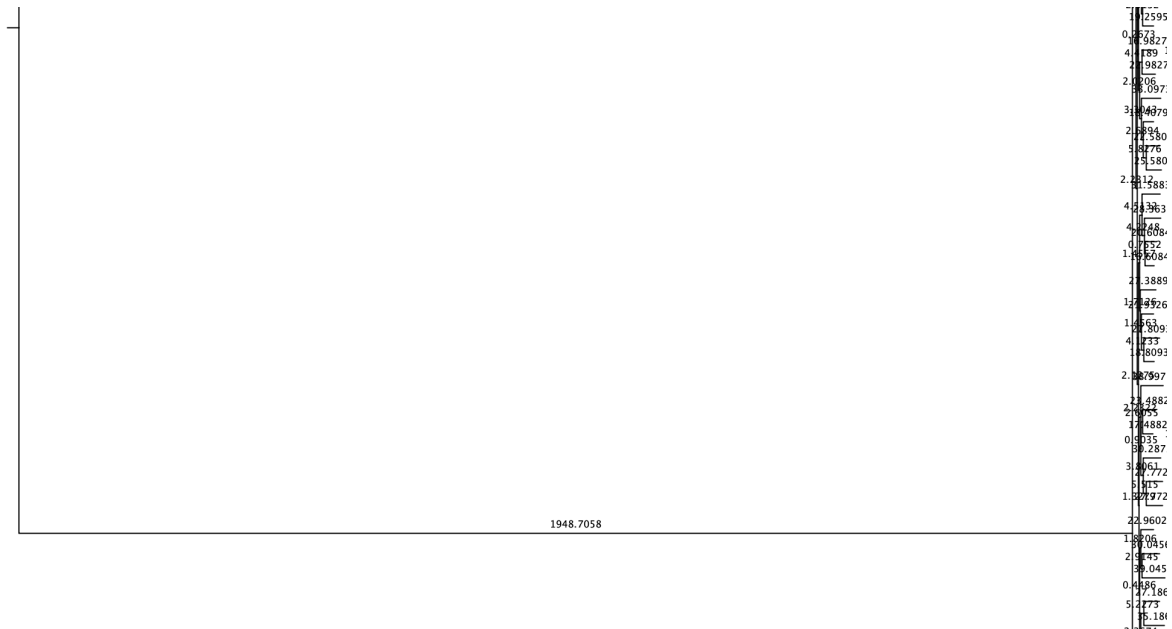
2,4963E-3



The screen of 1st assignment

When the value of the evolutionary rate obtained by Worobey equaled 2.47E-3. All obtained results were quite similar to each other.

2. What is the estimate for the time of the most recent common ancestor of the HIV-1 sequences used in this study? Rank the subtypes according to their mean age estimates (MRCAs).

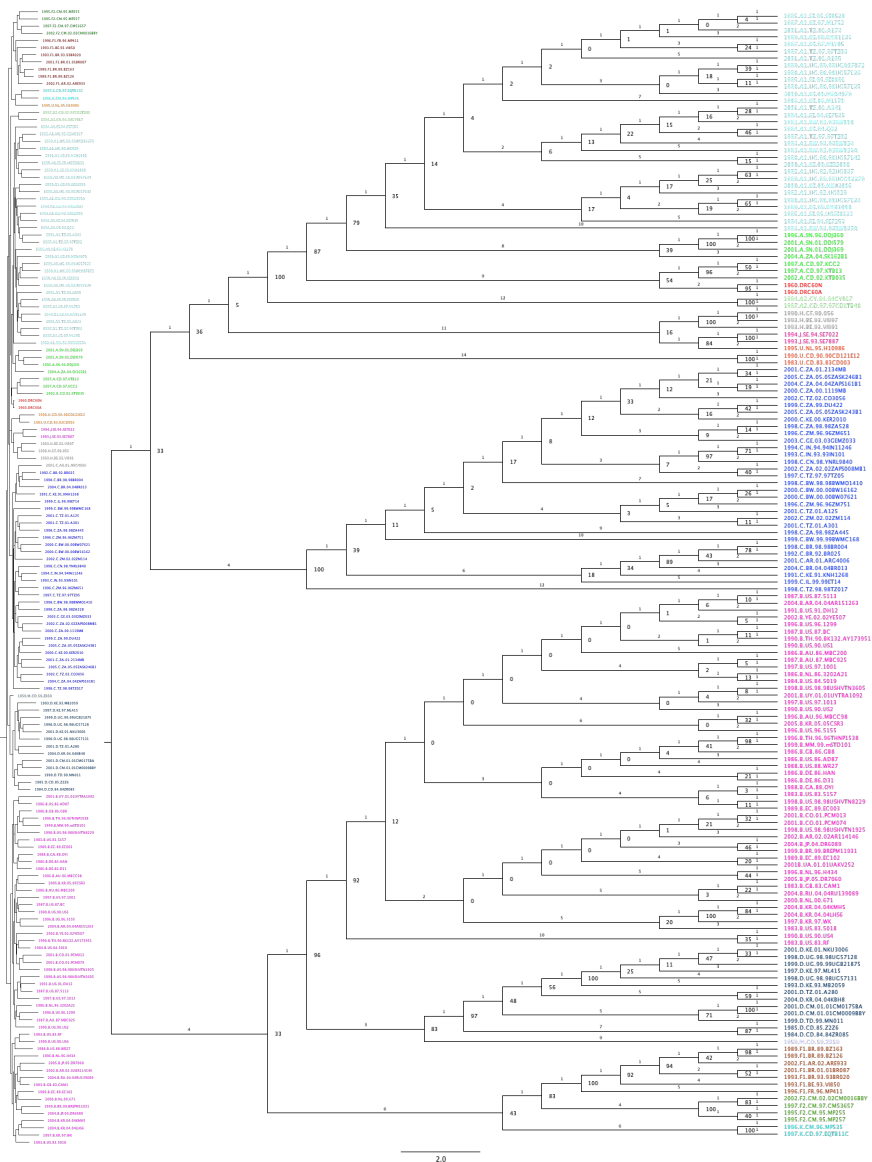


Estimate for the time of MRCA of the HIV-1 sequences

The estimated time for the most recent common ancestor of used HIV-1 sequences was 1948.7058. The subtypes were rank as follows: B, M, D, U, K, F2, F1, C, U, J, H, A (with all subtypes also containing 1960 strains)

3. Is the clustering between the subtypes consistent between the MCC tree and the ML tree?

The clustering differs - it's not consistent. The biggest difference is the placement of the U subtype. In the ML tree, the U subtype has one most recent ancestor, while in the tree MCC it's separated by several other subtypes.



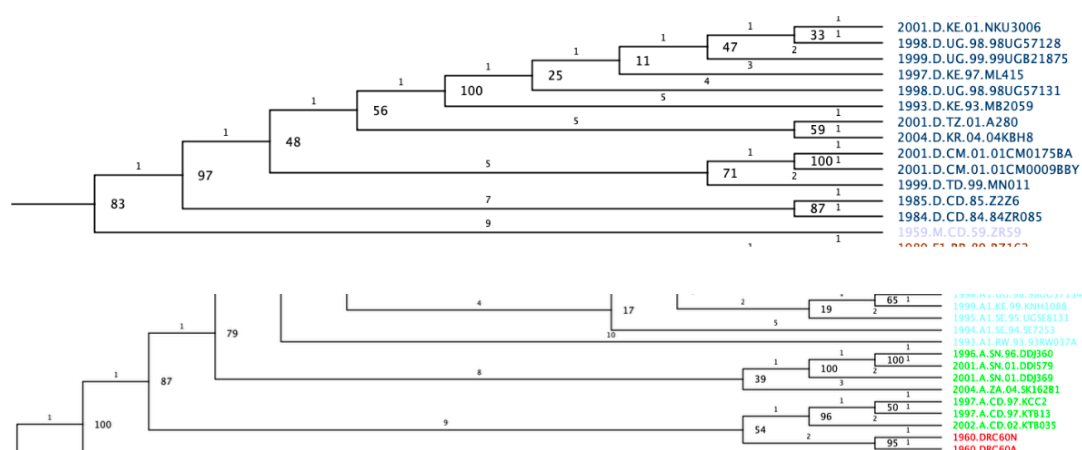
MCC tree

ML tree

Are the answers to question 1d for the ML tree the same for this MCC tree? Do the 1959 and the two 1960 sequences fall in a subtype cluster, If so, which one? If not, to which subtype is/are the sequence(s) most closely related?

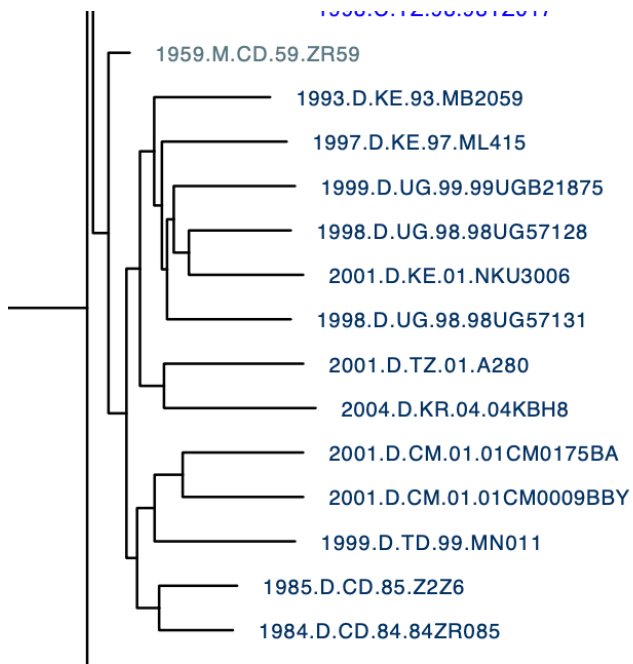
4. Do the 1959 and the two 1960 sequences fall in a subtype cluster. If so, which one? If not, to which subtype is/are the sequence(s) most closely related?

The strains 1959 and 1960 do not fall in the same subtype cluster. Below I present the screen with the evidence of that conclusion. The strain 1959 is the most related to the subtype D - they share their most recent common ancestor, the strain 1960 shares their most recent common ancestor with the subtype A.

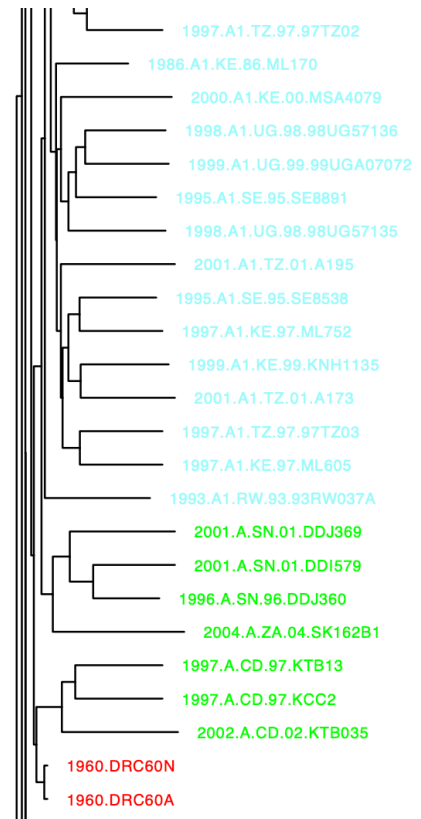


The screen of 1st assignment

The sequences do not fall into the same cluster. Sequence 1960 falls into the A cluster, while 1959 falls into the D subtype. The obtained results are similar to those received from the ML tree.



The cluster with 1959 sequence

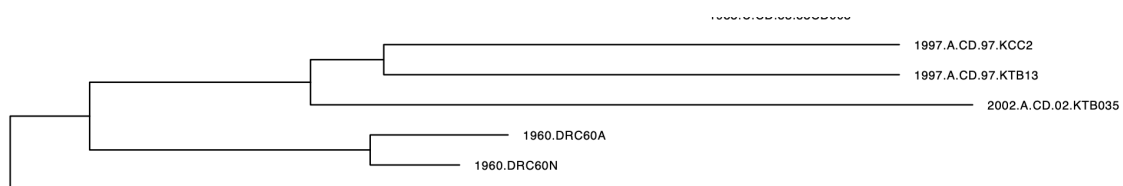


The cluster with 1960 sequence

Part 2

Perform a BEAST analysis with the same settings on this data set, but now estimate the age of both the 1960 sequences and the 1959 sequences (cfr. last tip below). In addition, because only short stretches are obtained for the 1960 sequences, constrain these sequences to be monophyletic with the subtype A sequences. Is this a reasonable assumption?

The 1960 sequences are monophyletic with the subtype A sequences. It is a reasonable assumption because this result I obtained repeatedly through all the analysis.



1960 sequences with the subtype A sequences

1. Does this affect the time estimate for the MRCA of the tree? And the evolutionary rate

It does affect the time estimate for the MRCA of the tree and evolutionary rate.

2. How accurately are the ages for the 1959 and 1960 sequences estimated?

The ages of 1959 and 1960 are not accurately estimated. For 1959 the age is equal to 36.7137 while for 1960 sequences 51,8446 and 55, 151.