Phylogenetic tree of modern human genomes

```bash
{bash}
pwd # /home/matt/ITMO/Omics/2HW

# the first step
cat Human/* > merged_human.fasta

# the second step
mafft --auto merged_human.fasta > merged_human_aligned.fasta

# the third step
trimal -in merged_human_aligned.fasta -out merged_human_aligned_trim.fasta -
keepheader -automated1

# the fourth step
iqtree -s merged_human_aligned_trim.fasta -m MFP -bb 1000 -alrt 1000

# the fifth step
figtree
```
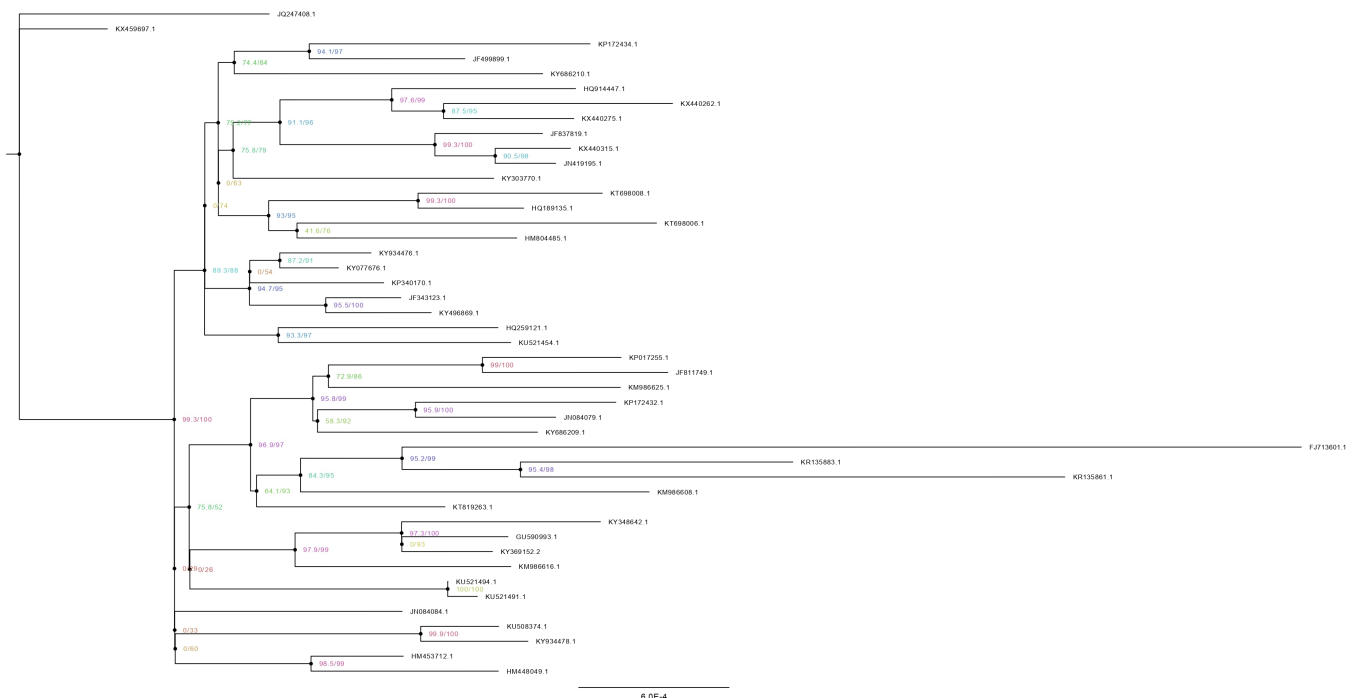

Phylogenetic tree of modern human genomes was obtained with bootstrap support values

Estimation: how old is Mitochondrial Eve within modern human and most recent ancestor of all non-Africans

I used MEGA tool to count real number mutations between samples and found the maximum one using Excel software: 96 mutations(between KX440262.1 Homo sapiens and FJ713601.1 Homo sapiens).

In the atricle (Pedigree derived mutation rate across the entire mitochondrial genome of the Norfolk Island population, https://doi.org/10.1038/s41598-022-10530-3) authors got the mutation rate for the entire mtGenome that is equal to 0.058 mutations/site/Myr. Whereas they provided information about other estimations from different studies but focused on the fact that researchers used different regions in their studies, so i decided to use the mutation rate from this article because they claimed that this rate was calculated for entire mitochondrial genome, so it's a more approriate variant for us as we did alignments of sequences for whole mitochondrial genome.

Using following command:

```bash
{bash}
grep "KX440262.1" * && grep "FJ713601.1" *
```

We got:
J2a2d1_Tunisian.fasta:>KX440262.1 Homo sapiens isolate JT76 haplogroup J2a2d1 mitochondrion, complete genome
L1c1d_Central_African.fasta:>FJ713601.1 Homo sapiens haplogroup L1c1d mitochondrion, complete genome

So we calculate a number of nucleotides in these samples:
J2a2d1_Tunisian.fasta - 16567 bp
L1c1d_Central_African.fasta - 16566 bp

Thus we can calculate:

$$Age = \frac{96 * 1.000.000}{0.058 * 16567} = 99907$$

We got that Mitochondrial Eve for modern human genome samples is approximately 99.907 years.

But Central African and Tunisian samples are related to Africa, so we need to find other variants for **non-Africans samples**.

The maximum number of mutations i found between JF811749.1 Homo sapiens and KT698006.1 Homo sapiens. It is equal to 62.

```{bash}
grep "KT698006.1" * # Serbian
grep "JF811749.1" * # Turkish
```

U2e2a4_Serbian.fasta - 16570 bp
L1c1d_Central_African.fasta - 16570 bp

$$Age = \frac{62 * 1.000.000}{0.058 * 16570} = 64512$$

According to my estimates, the most recent ancestor of all non-Africans was 64,612 years ago.

Estimation: how old is Mitochondrial Eve for modern human population and Neanderthals

For this purpose I did Multiple Sequence Alignment only for modern human and Neanderthals genomes for simplifing finding real number mutations between them.

```{bash}
pwd # /home/matt/ITMO/Omics/2HW

# the first step
cat Human/* > merged_NeanHuman.fasta
cat Neanderthal/* >> merged_NeanHuman.fasta

# the second step
mafft --auto merged_NeanHuman.fasta > merged_NeanHuman_aligned.fasta

# the third step
trimal -in merged_NeanHuman_aligned.fasta -out
merged_NeanHuman_aligned_trim.fasta -keepheader -automated1


# the fourth step

mega
```

I used MEGA tool to count real number mutations between samples and found the maximum one using Excel software: 220 mutations(between KX440262.1 Homo sapiens and KX198086.1 Homo sapiens neanderthalensis, also between KX440262.1 Homo sapiens and KX198085.1 Homo sapiens neanderthalensis).

```bash
{bash}
grep "KX440262.1" * # J2a2d1_Tunisian.fasta
grep "KX198086.1" * # GoyetQ305-7_Neanderthal.fasta
```

J2a2d1_Tunisian.fasta - 16567 bp
GoyetQ305-7_Neanderthal.fasta - 16563 bp
I think we can use the same mutation rate, so we can calculate in the same way:

$$Age = \frac{220 * 1.000.000}{0.058 * 16567} = 229955$$

The most recent common ancestor of modern humans and Neanderthals lived 229955 years ago

Estimation of the time of separation of Neanderthals and Denisovans populations

The code for obtaining a phylogenetic tree of samples of the genome of modern humans, Neanderthals and Denisova:

```bash
{bash}
pwd # /home/matt/ITMO/Omics/2HW

# the first step
cat Human/* > merged_together.fasta
cat Neanderthal/* >> merged_together.fasta
cat Denisovan/* >> merged_together.fasta

# the second step
mafft --auto merged_together.fasta > merged_together_aligned.fasta

# the third step
trimal -in merged_together_aligned.fasta -out
merged_together_aligned_trim.fasta -keepheader -automated1

# the fourth step
```
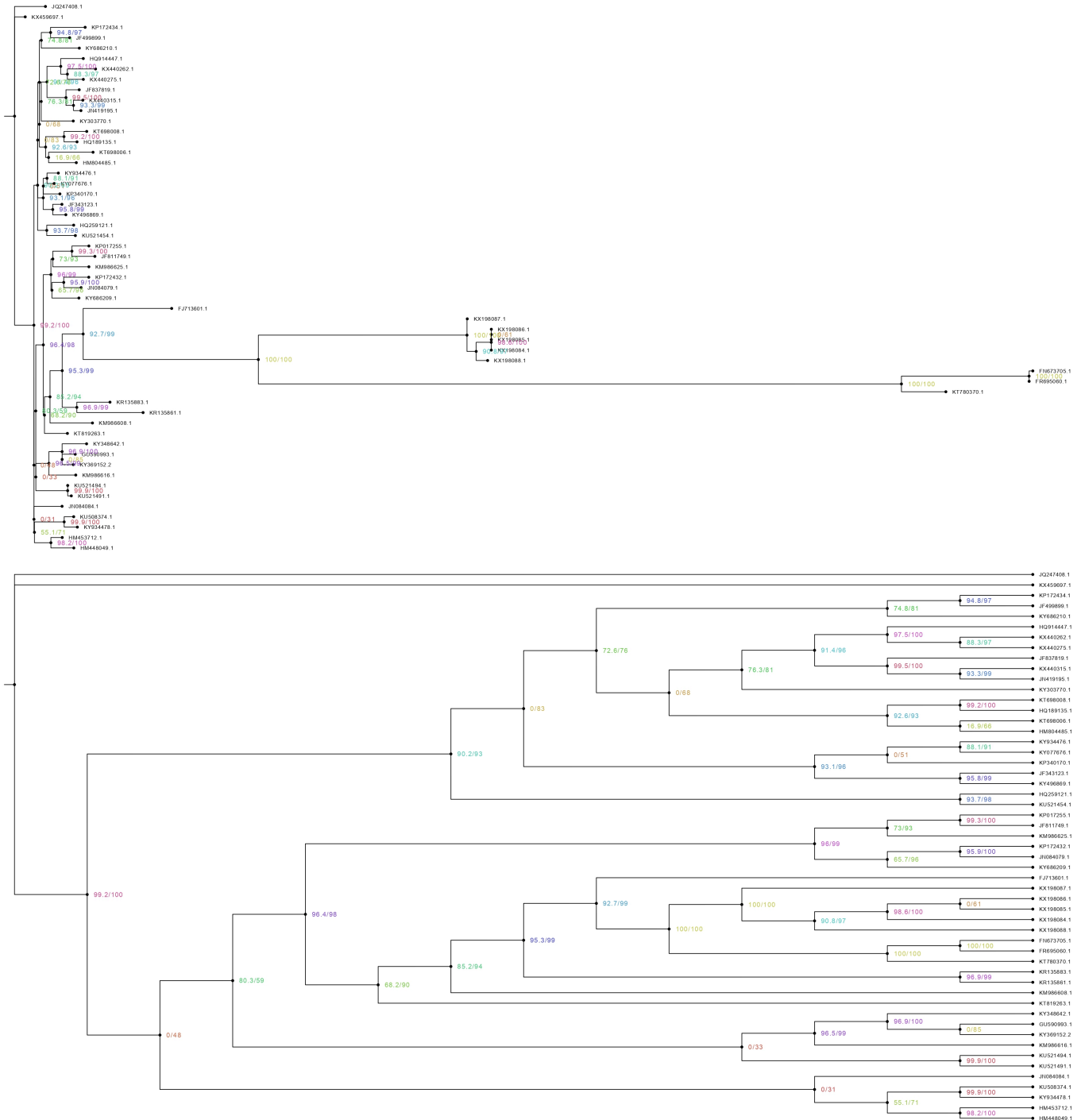
```
iqtree -s merged_together_aligned_trim.fasta -m MFP -bb 1000 -alrt 1000


# the fifth step


figtree
```

I built two types of phylogenetic trees, because in a tree where the length of the branches represents the passage of time, bootstrap values are not visible.

The code for obtaining a phylogenetic tree of samples of the genome of modern humans, Neanderthals, Denisova and Pan mtDNA:

```bash
{bash}
pwd # /home/matt/ITMO/Omics/2HW

# the first step
cat Human/* > merged_final.fasta
cat Neanderthal/* >> merged_final.fasta
cat Denisovan/* >> merged_final.fasta
cat Pan/* >> merged_final.fasta

# the second step
mafft --auto merged_final.fasta > merged_final_aligned.fasta

# the third step
trimal -in merged_final_aligned.fasta -out merged_final_aligned_trim.fasta -keepheader -automated1

# the fourth step
iqtree -s merged_final_aligned_trim.fasta -m MFP -bb 1000 -alrt 1000

# the fifth step

figtree
```
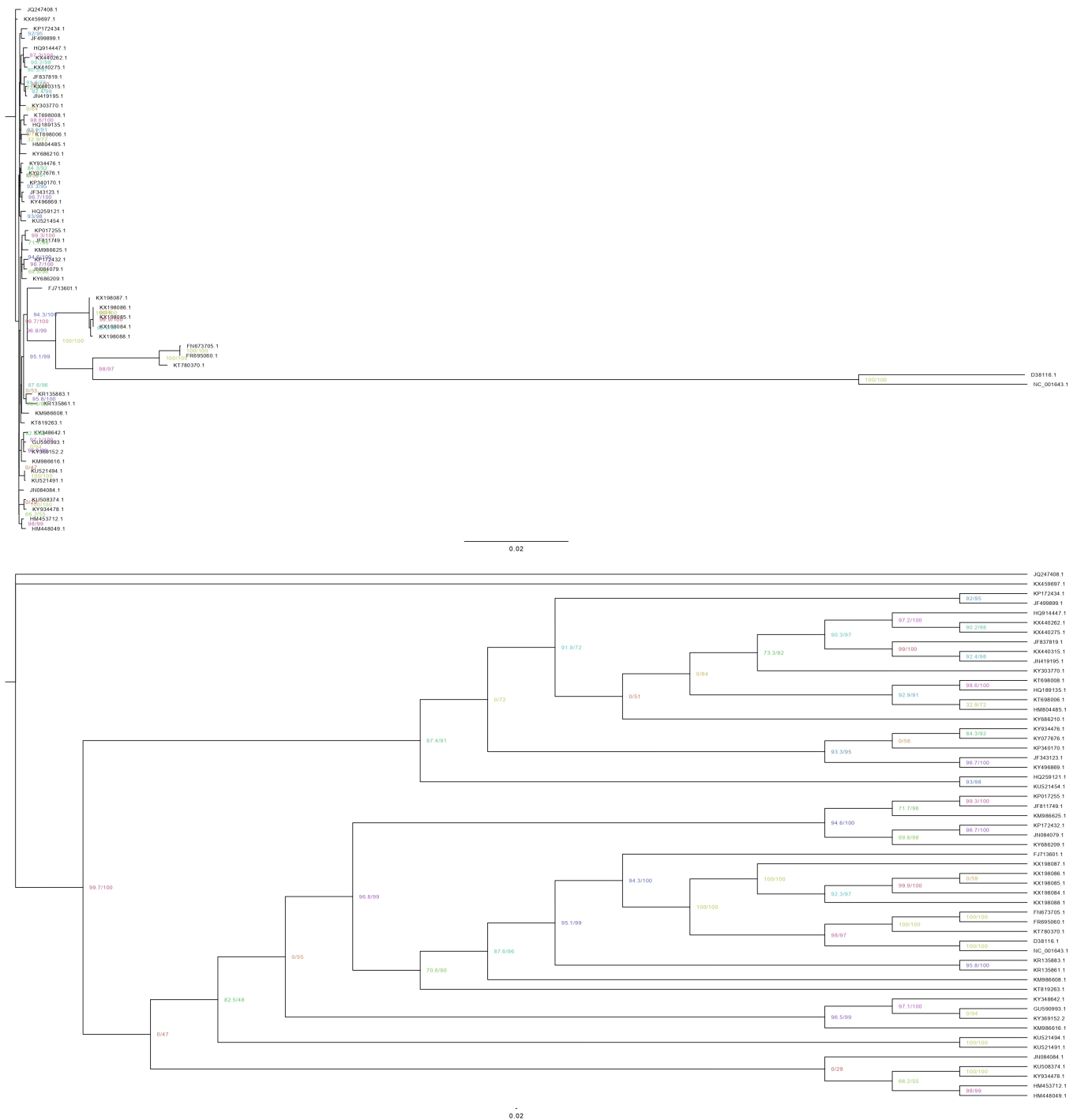
The tree topology haven't changed, because samples from Pan mtDNA have formed a separate branch.

The maximum number of mutation is equal to 382 between KX198085.1 Homo Sapiens neanderthalensis and FN673705.1 Homo sp. Altai.

$$Age = \frac{382 * 1.000.000}{0.058 * 16600} = 396759$$

The time of separation of Neanderthals and Denisovans was 396,759 years ago.