Assignment 3 / Biometry and bioinformatics I / 2014

Topic: Human mitochondrial genome as a tool for tracing human population histories. Human mtDNA-database, http://www.mtdb.igp.uu.se/

During the session 23. 9 there will be a lecture about this topic. You can, however, start collecting data before that.

- Collect ~5 mtDNA genome sequences from each continent, i.e. ~30 sequences in total

- Align the seqs and construct a neighbor joining –tree.

- Interprete the tree by using the map below. The map is a story about human population histories, starting from Africa.

- Inspect your alignment and report what kind of polymorphisms there are and at what genes or other regions of the mitochondrial genome they are located. Note the facility *polymorphic sites* in http://www.mtdb.igp.uu.se/. However, if there are very many polymorphic sites in your dataset, it is not necessary that you report them all. The idea is that you *learn to read information* (by using your eyes, get a *touch with a data and results*) and it is enough that you report some of the polymorphic sites to show that you understand.

Mitochondrial genome of all mammals is a conserved entity: about 16 000bp long, 22 tRNA´s (transfer RNAs), two ribosomal genes (12S and 16S) 13 protein encoding genes, and the non-coding D-loop.

It is circle, but naturally linear as a database-sequence. But: not always cut from the same location. Most database seqs are such that they start from the D-loop and end at two tRNA-genes (tRNA-Thr and tRNA-Pro). Use only this kind of seqs! Otherwise you run into practical problems! This mean that you should check each sequence from its description:

An example. The first seq in the database, EF064321.
After the general description, the start of seq details descrption looks like this:

```
D-loop          join(16024..16569,1..577)
     tRNA            578..648
                     /product="tRNA-Phe"
     rRNA            649..1602
                     /product="12S ribosomal RNA"
     tRNA            1603..1671
                     /product="tRNA-Val"
     rRNA            1672..3229
                     /product="16S ribosomal RNA"
     tRNA            3230..3304
                     /product="tRNA-Leu"
     gene            3307..4263
                     /gene="ND1"
     CDS             3307..4263
                     /gene="ND1"
                     /codon_start=1
                     /transl_table=2
                     /product="NADH dehydrogenase subunit 1"
.
.
```

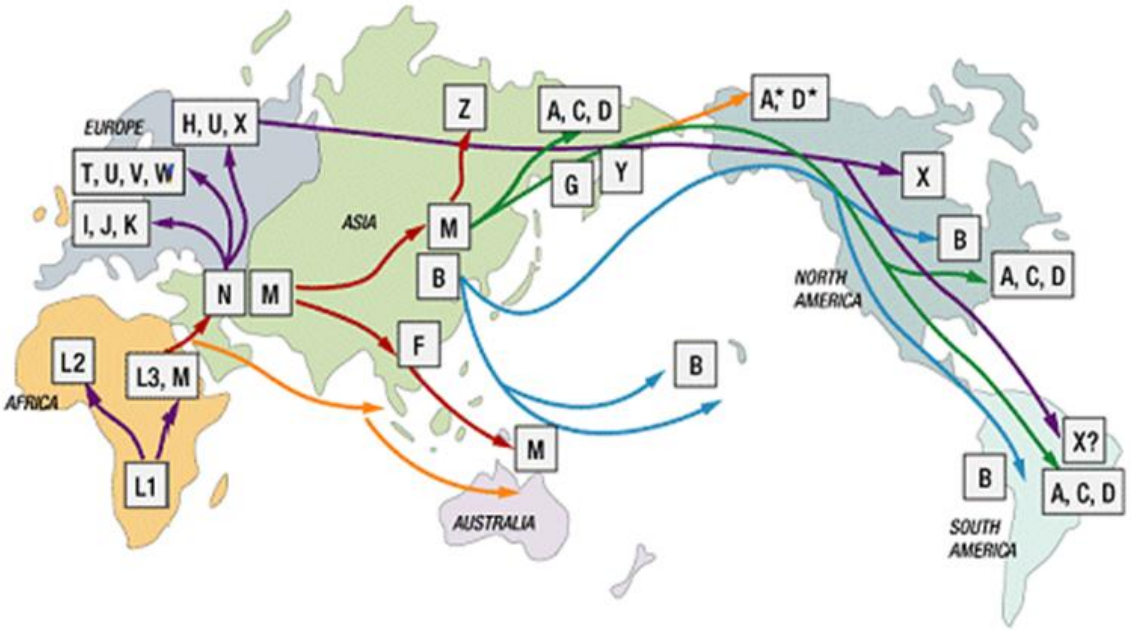
and ends like this:

```
TFHPYYTIKDALGLLLFLLSLMTLTLFSPDLLGDPDNYTLANPLNTPPHIKPEWYFLF

AYTILRSVPNKLGGVLALLLSILILAMIPILHMSKQQSMMFRPLSQSLYWLLAADLLI
                     LTWIGGQPVSYPFTIIGQVASVLYFTTILILMPTISLIENKMLK"
     tRNA            15888..15953
                     /product="tRNA-Thr"
     tRNA            complement(15955..16023)
                     /product="tRNA-Pro"
```

All seqs which you collect should have this configuration.

Alignment is very easy (there are probably only 1-3 gaps) but must be done. Recommendation is that you use the MAFFT-server because it is so fast. The sequences are rather long => Clustal quite slow. (Or try both, that will be a piece of learning useful practical issues.)



EXPANSION TIMES *(years ago)*	
Africa	120,000 - 150,000
Out of Africa	55,000 - 75,000
Asia	40,000 - 70,000
Australia/PNG	40,000 - 60,000
Europe	35,000 - 50,000
Americas	15,000 - 35,000
Na-Dene/Esk/Aleuts	8,000 - 10,000

FamilyTreeDNA

mtDNA Migrations Map