

Genomics credit project



Table of Contents

1. Introduction
2. Comparative analysis of mitochondrial genomes
3. Phylogenetic analysis
4. Sources

Introduction

The Canidae (or Canidea) is a family of terrestrial mammals of the suborder Caniformia within the order Carnivora, including dogs, wolves, foxes, coyotes and jackals, with a total of over 30 living species distributed on all continents except Antarctica. Poland is home to the grey wolf (*Canis lupus*), the red fox (*Vulpes vulpes*) and the Asiatic raccoon (*Nyctereutes procyonoides*) - introduced in 1955.

The domesticated form of the wolf is the domestic dog, which has accompanied humans for thousands of years as a service animal as well as a pet. They are medium-sized mammals. The smallest representative of the canids is the desert fennec and the largest is the grey wolf. Most of the species in the canine family are listed in the Red List of Threatened Species of the International Union for Conservation of Nature and Natural Resources, including the red wolf, Darwin's fox and island fox in the critically endangered (CR) category.

Comparative analysis of mitochondrial genomes

The data was retrieved from the NCBI database. By entering the name of the species with the suffix complete mitochondrial genome into the search engine, we were able to find complete mitochondrial genomes from the canine family. We selected 10 species for our analyses.

Name of species
Canis lupus <- which will be our reference genome
Lycaon pictus
Vulpes vulpes
Nyctereutes procyonoides
Canis lupus familiaris
Vulpes zerda
Chrysocyon brachyurus
Otocyon megalotis
Canis latrans
Cuon alpinus

The screenshot shows the NCBI Genome page for the *Cuon alpinus* complete genome mitochondrial. The main content area displays the genome's name, RefSeq number (NC_013445.1), and length (16,672 bp circular). Below this are links for Sequence Viewer, Primer-BLAST, and Download. To the right, a sidebar titled "Search details" shows the search query used: ("Cuon alpinus"[Organism] OR Cuon alpinus[All Fields]) AND complete[All Fields] AND genome[All Fields] AND mitochondrial[All Fields]. The sidebar also includes a "Recent activity" section listing other genome entries.

```
docker run --rm -v "$(pwd)":/dir -w /dir pstothatd/cgview_comparison_tool
build_blast_atlas.sh -p Projekt build_blast_atlas.sh -p Projekt
```

Once the files had been downloaded, we used the above command to create a Project folder into which we then moved the previously downloaded files with minor changes.

The terminal output shows the creation of a Docker project named 'Projekt'. It lists several genome projects: dna_vs_dna, cds_vs_cds, and others. Below the terminal is a file explorer showing the contents of the 'Projekt' folder. The folder contains subfolders for analysis, cct_projects, comparison_genomes, features, and reference_genome, along with configuration files for project settings.

Name	Date Modified	Size	Kind
> analysis	Yesterday at 23:57	--	Folder
> cct_projects	Yesterday at 23:57	--	Folder
> comparison_genomes	Today at 00:11	--	Folder
> features	Today at 00:09	--	Folder
> maps_for_cds_vs_cds	Yesterday at 23:57	--	Folder
> maps_for_dna_vs_dna	Yesterday at 23:57	--	Folder
project_settings_cds_vs_cds.conf	Yesterday at 23:57	7 KB	Config...tion file
project_settings_dna_vs_dna.conf	Yesterday at 23:57	7 KB	Config...tion file
> reference_genome	Today at 00:10	--	Folder

In the above screenshot we can see the contents of the Project folder that was created with Docker. In the reference_genome folder we place the fasta file for the reference genome (for us it will be *Canis lupus*) with a txt extension, into the features folder we place the *Canis lupus* reference genome with a GFF extension (with the extension changed to txt). In comparison_genomes we place the genomes of other canids also with a txt extension.

reference_genome			
Name	Date Modified	Size	Kind
Canis_lupus.txt	Yesterday at 23:33	17 KB	Plain Text
features			
Name	Date Modified	Size	Kind
Canis_lupus_gff.txt	Yesterday at 23:34	11 KB	Plain Text
comparison_genomes			
Name	Date Modified	Size	Kind
Canis_familiaris.txt	Yesterday at 23:37	17 KB	Plain Text
Canis_latrans.txt	Yesterday at 23:45	17 KB	Plain Text
Chrysocyon_bechyrurus.txt	Yesterday at 23:40	17 KB	Plain Text
Cuon_alpinus.txt	Yesterday at 23:47	17 KB	Plain Text
Lycaon_pictus.txt	Yesterday at 23:24	1 KB	Plain Text
Nyctereutes_procyonoides.txt	Yesterday at 23:36	17 KB	Plain Text
Otocyon_megalotis.txt	Yesterday at 23:42	17 KB	Plain Text
Vulpes_vulpes.txt	Yesterday at 23:36	17 KB	Plain Text
Vulpes_zerda.txt	Yesterday at 23:38	17 KB	Plain Text

We then continue working in the Project folder, this time modifying the maps_for_dna_vs_dna file.

```
#The genetic code to use for translated BLAST searches and for ORF
#translation. The default is the bacterial genetic code (genetic code
#11). See http://bioinformatics.org/sms2/genetic\_code.html for
#descriptions of the different genetic codes. [Integer].
genetic_code = 2
```

Above genetic_code changed from 11 to 2

```
#The query size for BLAST searches, i.e. how much of the reference
#genome is used in each BLAST search. This setting only applies to
#'trans' and 'nucleotide' comparisons (see the query_source option
#below). [Integer].
query_size = 100
```

query_size from 100000 to 100

```
#The start codons to be used when finding ORFs. The default set
#(ttg|ctg|att|atc|ata|atg|gtg) contains the starts for bacterial
#sequences. [Codons separated by '|'].
start_codons = atg
```

start_codon from ttg|ctg|att|atc|ata|atg|gtg to atg

After the above-mentioned commands have been applied, the first command is executed again in the terminal, i.e:

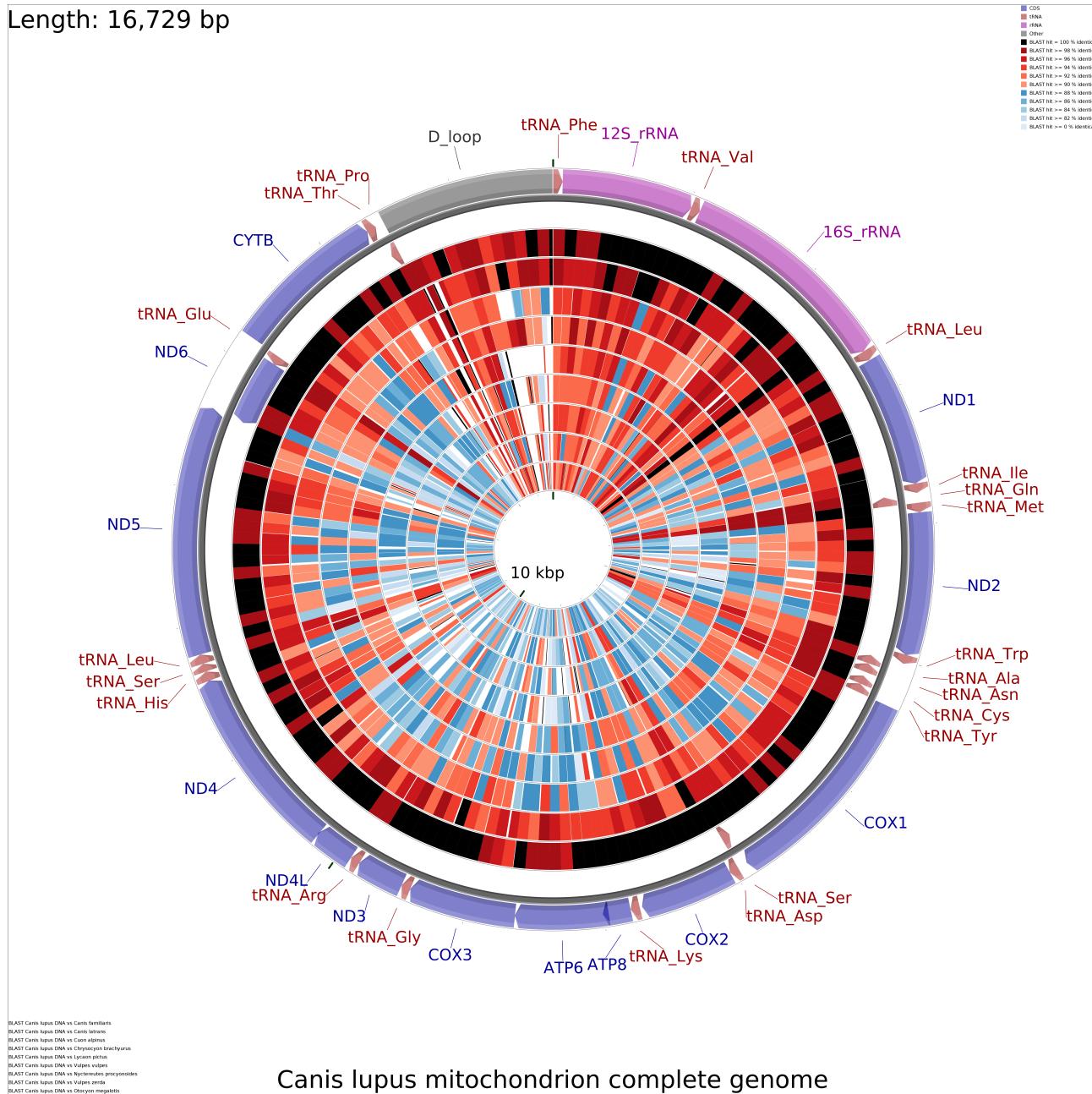
```
docker run --rm -v "$(pwd)":/dir -w /dir pstothard/cgview_comparison_tool  
build_blast_atlas.sh -p Projekt build_blast_atlas.sh -p Projekt
```

We then rework the GFF file from the features folder accordingly, so that it resembles the screenshot below.

A	B	C	D	E	F	G	H	
1	seqname	source	feature	start	end	score	strand	frame
2	ATP6	RefSeq	CDS	7965	8645	.	+	.
3	ATP8	RefSeq	CDS	7804	8007	.	+	.
4	COX1	RefSeq	CDS	5350	6894	.	+	.
5	COX2	RefSeq	CDS	7035	7718	.	+	.
6	COX3	RefSeq	CDS	8645	9428	.	+	.
7	CYTB	RefSeq	CDS	14186	15325	.	+	.
8	D_loop	RefSeq	D_loop	15461	16729	.	+	.
9	ND1	RefSeq	CDS	2748	3704	.	+	.
0	ND2	RefSeq	CDS	3915	4958	.	+	.
1	ND3	RefSeq	CDS	9497	9842	.	+	.
2	ND4	RefSeq	CDS	10204	11581	.	+	.
3	ND4L	RefSeq	CDS	9914	10210	.	+	.
4	ND5	RefSeq	CDS	11781	13601	.	+	.
5	ND6	RefSeq	CDS	13585	14112	.	-	.
6	tRNA_AlA	RefSeq	tRNA	5038	5106	.	-	.
7	tRNA_Arg	RefSeq	tRNA	9843	9911	.	+	.
8	tRNA_Asn	RefSeq	tRNA	5108	5179	.	-	.
9	tRNA_Asp	RefSeq	tRNA	6967	7034	.	+	.
0	tRNA_Cys	RefSeq	tRNA	5213	5280	.	-	.
1	tRNA_Gln	RefSeq	tRNA	3769	3843	.	-	.
2	tRNA_Glu	RefSeq	tRNA	14111	14181	.	-	.
3	tRNA_Gly	RefSeq	tRNA	9429	9496	.	+	.
4	tRNA_His	RefSeq	tRNA	11580	11650	.	+	.
5	tRNA_Ile	RefSeq	tRNA	3704	3772	.	+	.
6	tRNA_Leu	RefSeq	tRNA	2671	2745	.	+	.
7	tRNA_Leu	RefSeq	tRNA	11711	11780	.	+	.
8	tRNA_Lys	RefSeq	tRNA	7736	7802	.	+	.
9	tRNA_Met	RefSeq	tRNA	3845	3914	.	+	.
0	tRNA_Phe	RefSeq	tRNA	1	69	.	+	.
1	tRNA_Pro	RefSeq	tRNA	15395	15460	.	-	.
2	tRNA_Ser	RefSeq	tRNA	6892	6962	.	-	.
3	tRNA_Ser	RefSeq	tRNA	11651	11710	.	+	.
4	tRNA_Thr	RefSeq	tRNA	15326	15395	.	+	.
5	tRNA_Trp	RefSeq	tRNA	4957	5024	.	+	.
6	tRNA_Tyr	RefSeq	tRNA	5281	5348	.	-	.

The next step will be to modify the dna_vs_dna_large.xml file. After modifying it, we generate the map with the changes, using the command:

```
docker run --rm -v "$(pwd)":/dir -w /dir pstothard/cgview_comparison_tool  
redraw_maps.sh -p Projekt
```



Comparative genome map analysis

As presented in the analysis above, *Canis lupus* and *Canis lupus familiaris* have the most genes in common, indicating that our domestic dogs are closely related to grey wagons living in the wild. As the differences in the occurrence of the selected canids increase, as well as the phenotypic differences, one can see a decrease in the number of shared genes in the genotypes of the selected species. For the comparison of canine as well as wolf genotypes, we could also add the Dingo (*Canis dingo*, which was previously called *Canis lupus dingo*) - but we were unable to find its sequence in the NCBI database.

The most common stretches for our genomes are found on the 16S_RNA and 12S_RNA segments. The differences in genes for canids are evidence of their high genetic diversity, which also impinge on their adaptation to the environment in which they live. In our analyses, we compared the grey wolf, red fox, desert fennec or Australian raccoon, among others, which, in addition to visual differences, live in different ecological niches, which imply different adaptations to their ecological conditions. In the case of similar sequences, we can expect accidental cross-breeding between species, which negatively affects population purity. (As is the case with crosses between dogs and dingoes, or dogs and wolves). Similar sequences may also mean that species may have descended from a common ancestor but, through environmental or breeding factors, interbred with individuals of similar or specific physical characteristics resulting in specialisation.

Analiza filogenetyczna

Phylogenetic analysis was performed on 80 individuals of the canine family, the selected gene for analysis was Cytochrome B.

1. Preparation of the sequence

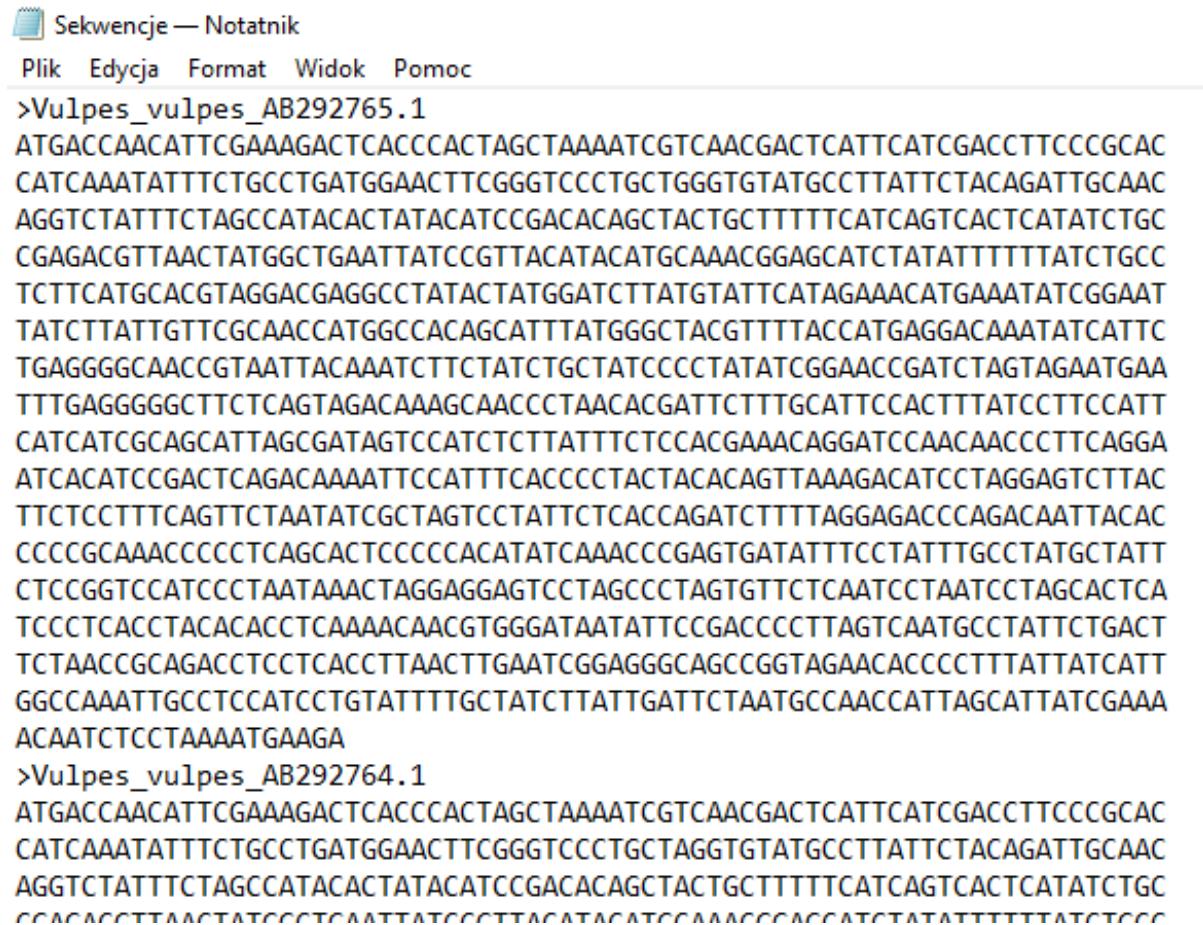
All the sequences we used are from the GenBank database, 80 of them are from the canid family and the other 4 are from another order (scales), and were placed to nest the tree.

Animal species analysed:

Name of species	Number of sequences used
<i>Vulpes vulpes</i>	10
<i>Vulpes Rueppellii</i>	6
<i>Vulpes lagopus</i>	10
<i>Canis lupus</i>	10
<i>Canis lupus familiaris</i>	11
<i>Otocyon megalotis</i>	3
<i>Canis aureus</i>	10
<i>Urocyon cinereoargenteus</i>	5
<i>Nyctereutes procyonoides</i>	5
<i>Cuon alpinus</i>	3
<i>Lycaon pictus</i>	3
<i>Canis anthus</i>	1
<i>Canis lupaster</i>	1
<i>Lycalopex</i>	1
<i>Canis simensis</i>	1
Species of a different order	4

Their irregular number is due to their availability in the database used.

2. We then put all the sequences, in a single file, in FASTA format.



Sekwencje — Notatnik

Plik Edycja Format Widok Pomoc

>Vulpes_vulpes_AB292765.1

ATGACCAACATTGAAAGACTACCCACTAGCTAAAATCGTCAACGACTCATTCATCGACCTCCGCAC
CATCAAATATTCTGCCGTATGGAACCTCGGGTCCCTGCTGGGTGTATGCCTTATTCTACAGATTGCAAC
AGGTCTATTCTAGCCATACACTATAACATCCGACACAGCTACTGCTTTTACATCAGTCACTCATATCTGC
CGAGACGTTAACATATGGCTGAATTATCGTTACATACATGCAAACGGAGCATCTATATTTTATCTGCC
TCTTCATGCACGTAGGACGAGGCCTATACTATGGATCTTATGTATTGATAGAAACATGAAATATCGGAAT
TATCTTATTGTTCGCAACCATTGGCACAGCATTATGGGCTACGTTTACCATGAGGACAAATATCATT
TGAGGGGCAACCGTAATTACAAATCTTCTATCTGCTATCCCCTATATCGGAACCGATCTAGTAGAATGAA
TTTGAGGGGGCTTCTCAGTAGACAAAGCAACCTAACACGATTCTTGCAATTCCACTTATCCTTCATT
CATCATCGCAGCATTAGCGATAGTCCATCTTCTTCCACGAAACAGGATCCAACAACCCCTCAGGA
ATCACATCCGACTCAGACAAAATTCCATTCAACCTACTACACAGTTAAAGACATCCTAGGAGTCTTAC
TTCTCCTTCAGTTCTAATATCGCTAGTCTTCTACCAAGATCTTCTAGGAGACCCAGACAATTACAC
CCCCGCAAACCCCTCAGCACTCCCCACATATCAAACCCGAGTGATATTCCATTGCCTATGCTATT
CTCCGGTCCATCCCTAATAAAACTAGGAGGAGTCTAGCCCTAGTGTCTCAATCCTAATCCTAGCACTCA
TCCCTCACCTACACACCTCAAAACACGTTGGGATAATATTCCGACCCCTTAGTCATGCCTATTCTGACT
TCTAACCGCAGACCTCCTCACCTTAACCTGAATGGAGGGCAGCCGGTAGAACACCCCTTATTATCATT
GCCCAAATTGCCTCATCCTGTATTTGCTATCTTATTGATTCTAATGCCAACATTAGCATTATCGAAA
ACAATCTCTAAAATGAAGA

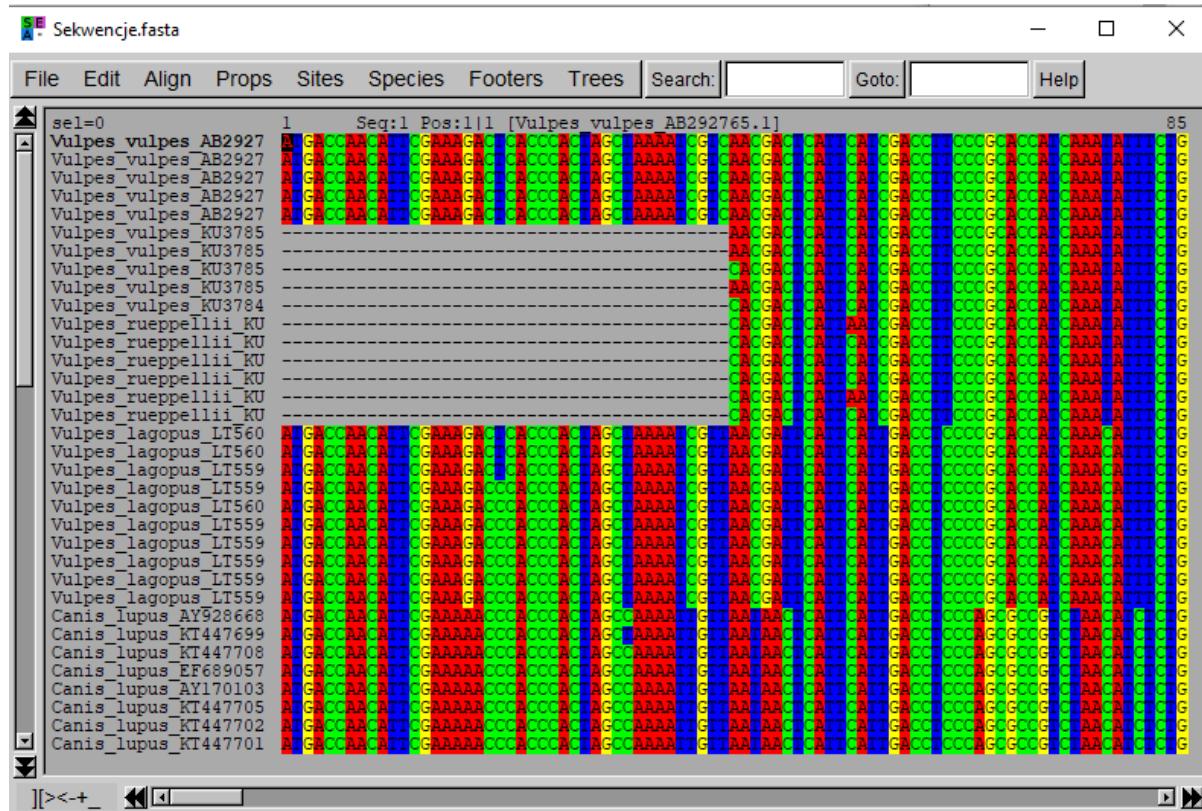
>Vulpes_vulpes_AB292764.1

ATGACCAACATTGAAAGACTACCCACTAGCTAAAATCGTCAACGACTCATTCATCGACCTCCGCAC
CATCAAATATTCTGCCGTATGGAACCTCGGGTCCCTGCTAGGTGTATGCCTTATTCTACAGATTGCAAC
AGGTCTATTCTAGCCATACACTATAACATCCGACACAGCTACTGCTTTTACATCAGTCACTCATATCTGC
CGACACCTTAACCTACACCTGATTTCTACATACATCCGACACAGCTACTGCTTTTACATCAGTCACTCATATCTGC

3. In this step, we will use the SeaView programme. First, we load the file, next we align the sequences as follows:

align --> alignment options --> muscle --> align all

The alignment looks as follows:



4. In SeaView, we highlight the nucleotide positions in the codons (separately for the first position, the second position and the third position). We then save each set in a separate file.

sites --> create set --> First codon pos.(dla 1 poz)-->ok--> file --> save selection

We still save the selections themselves in NEXUS format, this will be useful later when creating the phylogenetic tree.

5. From now on, we will use jModelTest to determine the optimal models for our sequences. First, we load our alignment file.

```
jModelTest 2.1.10 v20160303
(c) 2011-onwards D. Darriba, G.L. Taboada, R. Doallo and D. Posada,
(1) Department of Biochemistry, Genetics and Immunology
    University of Vigo, 36310 Vigo, Spain.
(2) Department of Electronics and Systems
    University of A Coruna, 15071 A Coruna, Spain.
e-mail: ddarriba@udc.es, dposada@uvigo.es

-----
Mon Jan 10 00:56:25 CET 2022
Windows 8.1 6.3, arch: amd64, bits: 64, numcores: 4

jModelTest 2.1.10 v20160303
Copyright (C) 2011 D. Darriba, G.L. Taboada, R. Doallo and D. Posada
This program comes with ABSOLUTELY NO WARRANTY
This is free software, and you are welcome to redistribute it under certain
conditions

Notice: This program may contain errors. Please inspect results carefully.

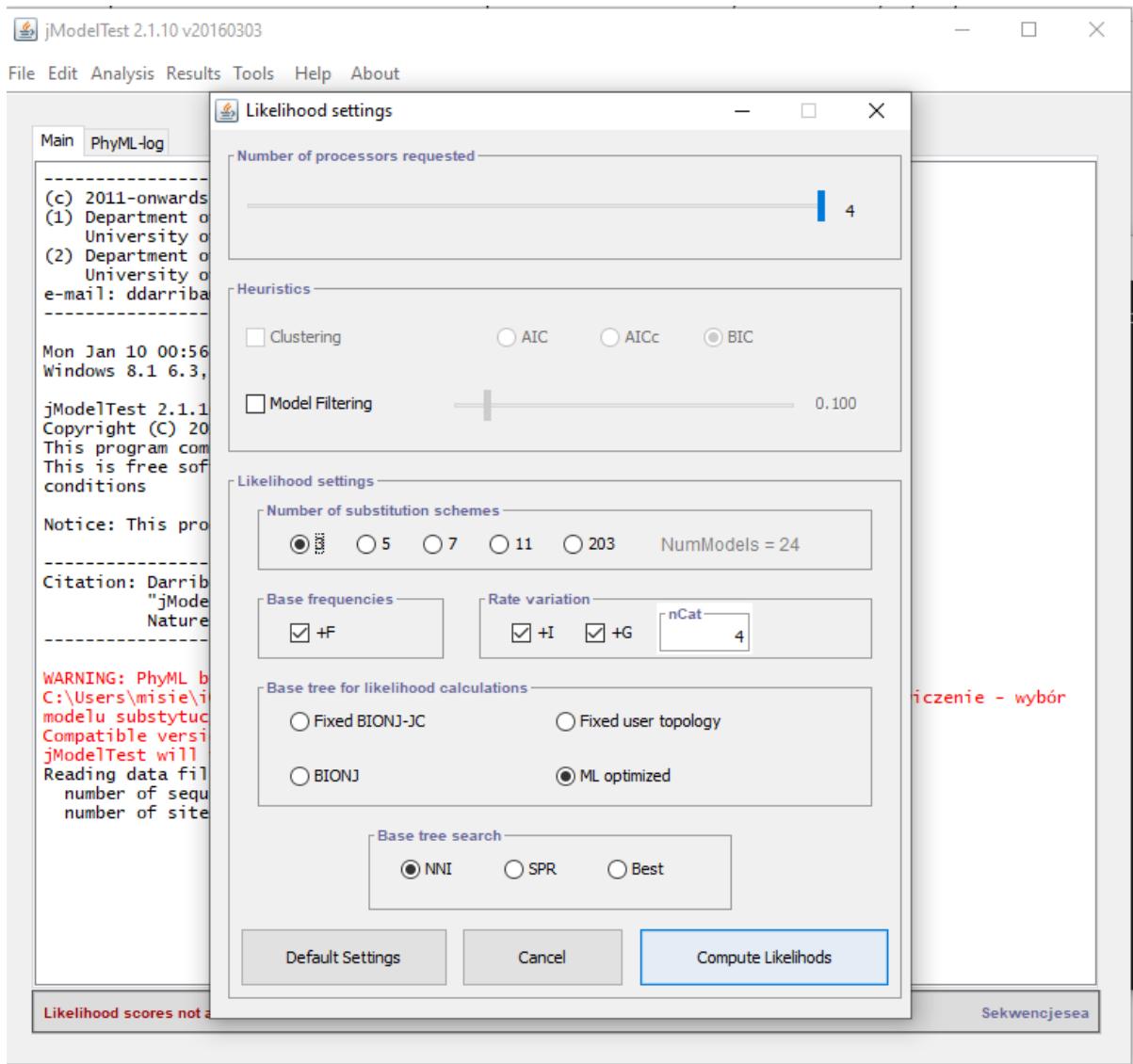
-----
Citation: Darriba D, Taboada GL, Doallo R and Posada D. 2012.
"jModelTest 2: more models, new heuristics and parallel computing".
Nature Methods 9(8), 772.

-----
WARNING: PhyML binary is not in the list of compatibility:
C:\Users\misie\iCloudDrive\Desktop\Studia\Rok 3 Semestr 5\Genomika i proteomika\4 Cwiczenie - wybór
modelu substytucji\jmodeltest-2.1.10\exe\phyml\PhyML_3.0_win32.exe v3.0
Compatible versions: 20130103 20131022 20141009 20141029 20150501 20151222
jModelTest will try to continue execution anyway, but it might fail.

Reading data file "Sekwencje sea"... OK.
    number of sequences: 84
    number of sites: 1140

-----
Likelihood scores not available
Sekwencje sea
```

Analysis --> compute likelihood scores --> Number of substitution schemes zmieniamy na 3 -->Compute likelihoods



6. The next step is to determine the values of the AIC, AICc and BIC statistics, based on which we will select our models.

Analysis --> Do AIC calculation --> Do AIC calculation
Analysis --> Do AIC calculation --> Zaznaczamy Use AICc correction --> Do AICc calculation
Analysis --> Do BIC calculation --> Do BIC calculation
Results --> Show results table

The scoreboard looks as follows:

ID	Name	Partition	-lnL	p	AICc	deltaAICc	weight	cumWeight	uDelta
7	F81+G	000000	8863.1657	170	18126.3314	1765.9283	0.0		1-
8	F81+I+G	000000	8846.5679	171	18095.9044	1735.5013	0.0		1-
9	K80	010010	9003.1459	167	18398.0202	2037.617	0.0		1-
10	K80+I	010010	8328.6639	168	17051.8077	691.4046	0.0		1-
11	K80+G	010010	8297.8518	168	16990.1834	629.7803	0.0		1-
12	K80+I+G	010010	8281.7748	169	16960.7867	600.3835	0.0		1-
13	HKY	010010	8862.3108	170	18124.6215	1764.2184	0.0		1-
14	HKY+I	010010	8054.9675	171	16512.7037	152.3005	0.0		1-
15	HKY+G	010010	8015.3069	171	16433.3823	72.9792	0.0		1-
16	HKY+I+G	010010	7979.5037	172	16364.5504	4.1472	0.1117		1-
17	SYM	012345	8792.855	171	17988.4787	1628.0755	0.0		1-
18	SYM+I	012345	8157.1674	172	16719.8778	359.4746	0.0		1-
19	SYM+G	012345	8148.5285	172	16702.5999	342.1967	0.0		1-
20	SYM+I+G	012345	8129.3916	173	16667.1062	306.7031	0.0		1-
21	GTR	012345	8849.6187	174	18110.3463	1749.9431	0.0		1-
22	GTR+I	012345	8040.5842	175	16495.0689	134.6657	0.0		1-
23	GTR+G	012345	8009.6561	175	16433.2126	72.8094	0.0		1-
24	GTR+I+G	012345	7971.8527	176	16360.4032	0.0	0.8883	0,888-	

Decimal numbers are rounded. Click on column headers to sort data in ascending or descending order (+Shift)
10 styczeń 2022

In the table we look for the model that has the best fit, it is highlighted in red.

We perform all the above steps for whole genes, for the first, second and third codons, and collect all the data in a table:

	AIC	AICc	BIC
Cały	GTR+I+G	GTR+I+G	HKY+I+G
1st	GTR+G	HKY+G	HKY+G
2nd	GTR+I+G	HKY+I	HKY+I
3rd	GTR+G	HKY+G	HKY+G

For further analysis, we have selected models designated by AICc statistics.

7. We use the MrBayes programme to determine our phylogenetic tree. First, we load our file.

```
MrBayes > execute "C:\Users\misie\iCloudDrive\Desktop\Studia\Rok 3 Sem  
Executing file "C:\Users\misie\iCloudDrive\Desktop\Studia\Rok 3 Sem  
DOS line termination  
Longest line length = 62  
Parsing file  
Expecting NEXUS formatted file  
Reading data block  
    Allocated taxon set  
    Allocated matrix  
    Defining new matrix with 84 taxa and 1140 characters  
    Data is Dna  
    Gaps coded as -  
    Taxon 1 -> Vulpes_vulpes_AB292765  
    Taxon 2 -> Vulpes_vulpes_AB292764  
    Taxon 3 -> Vulpes_vulpes_AB292763  
    Taxon 4 -> Vulpes_vulpes_AB292762  
    Taxon 5 -> Vulpes_vulpes_AB292761  
    Taxon 6 -> Vulpes_vulpes_KU378543  
    Taxon 7 -> Vulpes_vulpes_KU378531  
    Taxon 8 -> Vulpes_vulpes_KU378558  
    Taxon 9 -> Vulpes_vulpes_KU378553  
    Taxon 10 -> Vulpes_vulpes_KU378493  
    Taxon 11 -> Vulpes_rueppellii_KU378369  
    Taxon 12 -> Vulpes_rueppellii_KU378372  
    Taxon 13 -> Vulpes_rueppellii_KU378368  
    Taxon 14 -> Vulpes_rueppellii_KU378370  
    Taxon 15 -> Vulpes_rueppellii_KU378371  
    Taxon 16 -> Vulpes_rueppellii_KU378372
```

Next, we create sets (charsets) with a sequence length of 1140. Later, we create partitions named "codons"

```
>charset 1st=1-1140/3  
>charset 2nd=2-1140/3  
>charset 3rd=3-1140/3  
>partition kodony=3:1st,2nd,3rd  
>set partition=kodony
```

Next, for each charset, we assign the appropriate substitution models. We then run the program, specifying the number of steps in the Markov chains.

```
lset applyto=(1,3) nst=2 rates=gamma
lset applyto=(2) nst=2 rates=propinv
prset applyto=(3) pinvarpr=fixed(0)
mcmc ngen=100000
```

```
Using a relative burnin of 25.0 % for diagnostics
Chain results (100000 generations requested):
4495000 -- [-30428.491] / [-38222.301] / [-38252.529] / [-38159.258] * [-38257.438] / [-38195.333] / [-81349.419] / [-82259.325] -- 0.00:40
Average standard deviation of split frequencies: 0.006128

4496000 -- (-8252.741) (-8213.494) [-8174.448] (-8203.882) * (-8236.750) [-8191.563] (-8227.540) (-8216.733) -- 0:00:38
4497000 -- (-8245.124) (-8217.271) (-8197.809) [-8195.769] * (-8235.186) (-8178.150) (-8228.121) [-8187.538] -- 0:00:28
4498000 -- (-8228.765) (-8240.666) (-8196.198) [-8199.065] * (-8220.540) (-8188.640) (-8226.572) [-8190.792] -- 0:00:19
4499000 -- (-8222.085) (-8226.958) (-8184.157) [-8178.650] * (-8220.396) [-8184.615] (-8222.675) (-8210.275) -- 0:00:09
4500000 -- (-8218.367) (-8222.632) [-8185.429] (-8211.067) * (-8226.962) [-8186.965] (-8253.111) (-8226.618) -- 0:00:00

Average standard deviation of split frequencies: 0.006143

Continue with analysis? (yes/no):
Additional number of generations:
Average standard deviation of split frequencies: 0.106625

11000 -- (-8225.409) (-8262.637) (-8234.354) [-8219.142] * [-8207.294] (-8218.188) (-8243.420) (-8191.386) -- 0:16:10
12000 -- (-8219.232) (-8250.466) [-8213.939] (-8207.411) * [-8170.114] (-8211.677) (-8236.854) (-8203.967) -- 0:16:00
13000 -- [-8206.156] (-8228.839) (-8209.522) (-8208.083) * (-8192.853) (-8229.990) (-8235.802) [-8188.916] -- 0:15:50
14000 -- (-8215.961) (-8219.947) [-8207.303] (-8228.053) * (-8198.347) (-8208.280) (-8254.100) [-8184.851] -- 0:15:39
15000 -- (-8223.298) [-8208.204] (-8223.394) (-8239.676) * (-8190.435) (-8238.776) (-8252.060) [-8181.963] -- 0:15:40

Average standard deviation of split frequencies: 0.088651
```

The number of steps in the Markov chains, we increase until the "Avarage standard deviation of split frequencies", does not fall below 0.01. When this happens, we perform three more times the previous steps. Only after these calculations have been performed do we get a consensus tree.

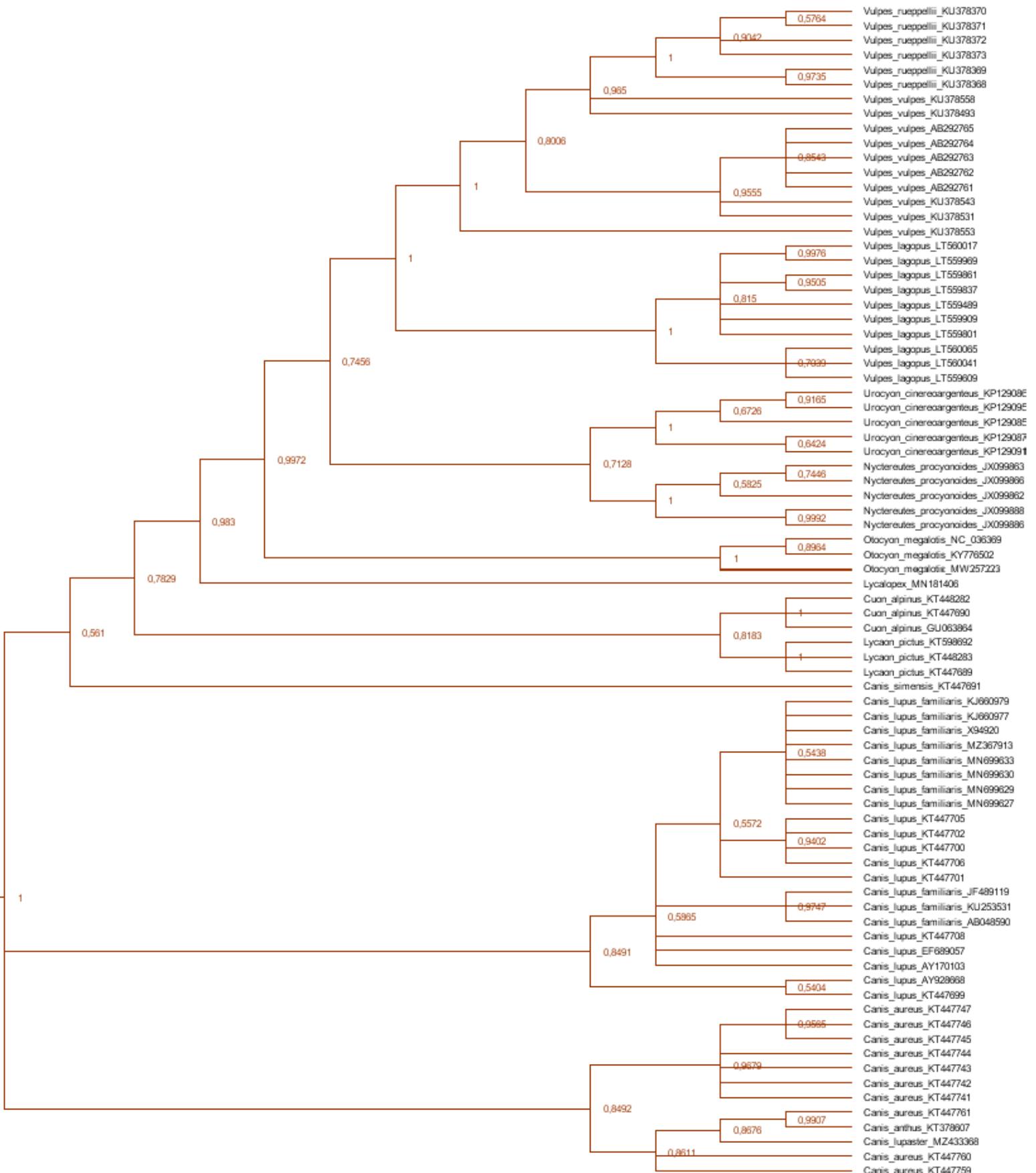
In our case, the value fell below 0.01 after about 1.1 million steps. In total, the number of steps taken was 4.5 million.

```
4495000 -- (-8251.187) (-8222.040) [-8182.854] (-8191.598) * (-8247.708) (-8195.505) [-8190.749] (-8228.598) -- 0.00:40
Average standard deviation of split frequencies: 0.006128

4496000 -- (-8252.741) (-8213.494) [-8174.448] (-8203.882) * (-8236.750) [-8191.563] (-8227.540) (-8216.733) -- 0:00:38
4497000 -- (-8245.124) (-8217.271) (-8197.809) [-8195.769] * (-8235.186) (-8178.150) (-8228.121) [-8187.538] -- 0:00:28
4498000 -- (-8228.765) (-8240.666) (-8196.198) [-8199.065] * (-8220.540) (-8188.640) (-8226.572) [-8190.792] -- 0:00:19
4499000 -- (-8222.085) (-8226.958) (-8184.157) [-8178.650] * (-8220.396) [-8184.615] (-8222.675) (-8210.275) -- 0:00:09
4500000 -- (-8218.367) (-8222.632) [-8185.429] (-8211.067) * (-8226.962) [-8186.965] (-8253.111) (-8226.618) -- 0:00:00

Average standard deviation of split frequencies: 0.006143

Continue with analysis? (yes/no):
Additional number of generations:
```



Analiza drzewa filogenetycznego

The tree has been rooted to a sequence of non-psidal organisms. Starting from the root, we can see two branches:

- Lower, which includes almost all wolf species (*Canis*), the only exception being *Canis Simensis*, which is genetically distant from the others, but should be included in this branch.
- Upper, where the remaining.

Another branch, is the one in which *Canis Simensis* erroneously appeared, this was probably due to the quality of the sequence used from the GenBank database. The low prob value, indicates the randomness of this match.

Another branching out is the separation of the outer group of laikons (*Lycaon Pictus*) and cyons (*Coua Alpinus*), prob values, both of which are high, this is a good fit.

In the next branching, with a high prob value, an external group of one species, Nibylis (*Lycalopex*), separates.

The next branching is the separation of the Otocions (*Otocyon Megalotis*). Here, too, the prob value is very high.

The next branching is crucial:

- Lower down, there are specimens of *Urocyon cinereoargenteus* and raccoons (*Nyctereutes procyonoides*)
- Only fox species (*Vulpes vulpes*, *Vulpes lagopus*, *Vulpes ruppelli*) were already higher.

Pro values, again, were high.

- Further differences were beyond these species themselves, *Vulpes lagopus* detached first, then *Vulpes vulpes*, at the very end *Vulpes ruppelli*. The differences between the last two are so negligible that two sequences of *Vulpes vulpes* were assigned to the branch on which *Vulpes ruppeli* is found. The remaining differences are only within the species concerned.

Comparison with literature

A molecular phylogeny of the Canidae based on six nuclear loci

C. Bardeleben et al. / Molecular Phylogenetics and Evolution 37 (2005) 815-831

Turning to the analysis of the phylogenetic tree, a comparison of our tree with the literature shows some deviations. Only the differences will be described.

Urocyon cinereoargenteus in our tree is found in one branch together with *Nyctereutes procyonoides*, but according to the literature this species separated much earlier.

The species *Cuon alpinus* and *Lycaon pictus*, according to our tree, do not belong to the branch on which *Canis* species are found; according to the literature, they belong to it but have quickly become detached from it.

Vulpes lagopus in our tree detached before *Vulpes vulpes*. According to the literature, however, *Vulpes lagopus* detached last.

DATA SET INCONGRUENCE, MISLEADING CHARACTERS, AND INSIGHTS FROM THE FOSSIL RECORD: THE CANID PHYLOGENY - Caldasia 33(2):637-658. 2011

- Again comparing with the literature, our species *Cuon* and *Lycaon*, were placed in the wrong place. (They were placed there with a prob of 0.561, which is a low value).

All the abnormalities that were shown in our analysis can be explained. The analysis we carried out referred only to the variation in the CytochromeB gene, not the whole genome. Furthermore, the literature trees to which we compared ours cannot be taken as certain and ideal. They were created using similar techniques, only on the basis of more extensive data.

Sources

<https://ocdn.eu/pulscms-transforms/1/xlwk9kpTURBXy8xYjVhYTE5NmY1MmNlODgzZDZhYmVjYWExZjNlOGJlYi5qcGeTIQMBzQE4zQ-gzQjKkwXNAxTNAbyTCaYwNTI4YzkGgaEwBQ/wilki-szare.webp>

https://upload.wikimedia.org/wikipedia/commons/5/59/Vulpes_vulpes_at_shipwreck.jpg

<https://zoo.wroclaw.pl/wp-content/uploads/2021/11/Fenek.webp>

<https://pl.wikipedia.org/wiki/Psowate>