

Projekt zaliczeniowy z Genomiki



Spis Treści

1. Wstęp
2. Analiza porównawcza genomów mitochondrialnych
3. Analiza filogenetyczna
4. Źródła

Wstęp

Psowate (inaczej Canidea) jest to rodzina lądowych ssaków z podrzędu psokształtnych (Caniformia) w obrębie rzędu drapieżnych (Carnivora) obejmująca między innymi psy, wilki, lisy, kojoty i szakale - łącznie ponad 30 współcześnie żyjących gatunków rozprzestrzenionych na wszystkich kontynentach poza Antarktydą. W Polsce występuje wilk szary (*Canis lupus*), lis rudy (*Vulpes vulpes*) i jenot azjatycki (*Nyctereutes procyonoides*) - introdukowany w 1955.

Udomowioną formą wilka jest pies domowy, który od tysięcy lat towarzyszy ludziom jako zwierzę użytkowe, a także domowe. Są to ssaki średniej wielkości. Najmniejszym przedstawicielem psowatych jest fenek pustynny, a największym - wilk szary. Większość gatunków z rodziny psowatych została wpisana do Czerwonej Księgi Gatunków Zagrożonych Międzynarodowej Unii Ochrony Przyrody i Jej Zasobów, w tym Wilk rudy, Lis Darwina i Lis wyspowy w kategorii krytycznie zagrożonych wyginięciem (CR).

Analiza porównawcza genomów mitochondrialnych

Dane zostały pobrane z bazy NCBI. Wpisując z wyszukiwarkę nazwę gatunku wraz z dopiskiem complete genome mitochondrial byliśmy w stanie odnaleźć kompletne genomy mitochondrialne z rodziny psowatych. Do naszych analiz wybraliśmy 10 gatunków.

Nazwa gatunku
Canis lupus <- który będzie naszym genomem referencyjnym
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 browser interface. At the top, it displays "Cuon alpinus complete genome mitochondrial". Below this, there's a "COVID-19 Information" banner with links to CDC, NIH, NCBP, and HHS. The main content area shows the "Dhole mitochondrial reference genome" (Cuon alpinus, RefSeq: NC_013445.1, Length: 16,672 bp circular). It includes a "Sequence Viewer", "Primer-BLAST", and "Download" button. To the right, there's a "Search details" panel with the search query: ("Cuon alpinus"[Organism] OR Cuon alpinus[All Fields]) AND complete[All Fields] AND genome[All Fields] AND mitochondrial[All Fields]. Below it is a "Recent activity" section listing previous searches for Cuon alpinus, Cerdocyon thous, and Cerdcoyon thous genomes.

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

Po pobraniu plików, za pomocą powyższego polecenia stworzyliśmy folder Projekt do którego następnie przenieśliśmy wcześniejszepobrane pliki z małymi zmianami.

The terminal window shows the creation of a Docker project named 'Projekt'. It includes commands to create projects for dna_vs_dna and cds_vs_cds, and a configuration file written to 'Projekt'. Below this, instructions for creating maps are provided, followed by a list of files and their details.

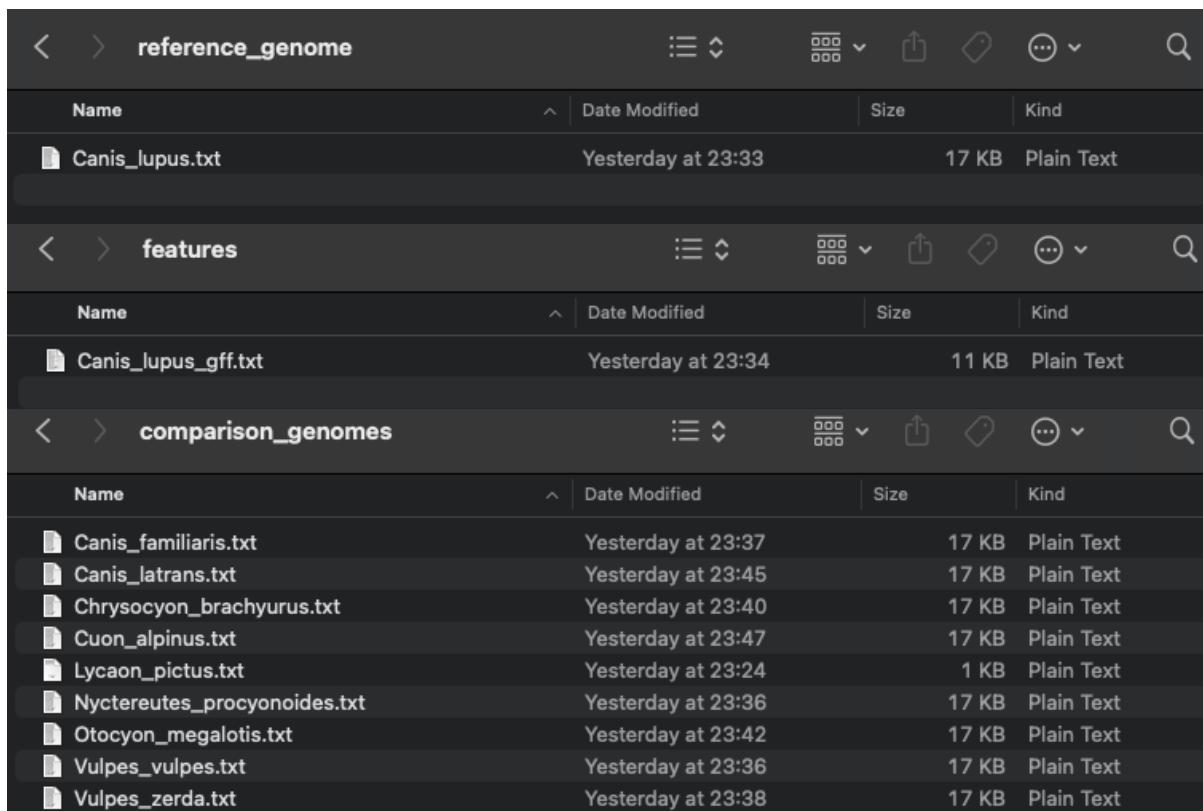
```
(base) daria@Darias-MBP ~ % docker run --rm -v "$(pwd)":/dir -w /dir pstothatd/cgview_comparison_tool build_blast_atlas.sh -p Projekt
Creating new project in 'Projekt'
Creating project for dna_vs_dna
[Saturday January 22 22:57:55 2022] [Notice] A new project has been created in Projekt/cct_projects/dna_vs_dna.
Creating project for cds_vs_cds
[Saturday January 22 22:57:55 2022] [Notice] A new project has been created in Projekt/cct_projects/cds_vs_cds.
Configuration files written to 'Projekt'.

TO CREATE MAPS:
- Place a reference genome into the 'Projekt/reference_genome' directory.
- Place comparison genomes into the 'Projekt/comparison_genomes' directory.
- Edit configurations file (Optional).
- Add analysis and/or features data files (Optional).
- Rerun this script.
```

The file explorer shows the contents of the 'Projekt' directory:

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...ption file
project_settings_dna_vs_dna.conf	Yesterday at 23:57	7 KB	Config...ption file
reference_genome	Today at 00:10	--	Folder

Na powyższym screenie możemy zobaczyć zawartość folderu Projekt który został utworzony za pomocą Dockera. W folderze reference_genom umieszczałyśmy plik fasta dla genomu referencyjnego (dla nas będzie to *Canis lupus*) z rozszerzeniem txt, do folderu features umieszczałyśmy genom referencyjny *Canis lupus* z rozszerzeniem GFF (ze zmienionym rozszerzeniem na txt). W comparison_genomes umieszczałyśmy genomy innych psowatych również z rozszerzeniem txt.



Następnie dalej pracujemy w folderze Projekt, tym razem modyfikując plik maps_for_dna_vs_dna plik.

```
#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
```

Powyżej gentic_code z 11 zmieniliśmy na 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 z 100000 na 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 z ttg|ctg|att|atc|ata|atg|gtg na atg

Po zastosowaniu wyżej wymienionych poleceń ponownie puszczamy w terminalu pierwsze polecenie czyli:

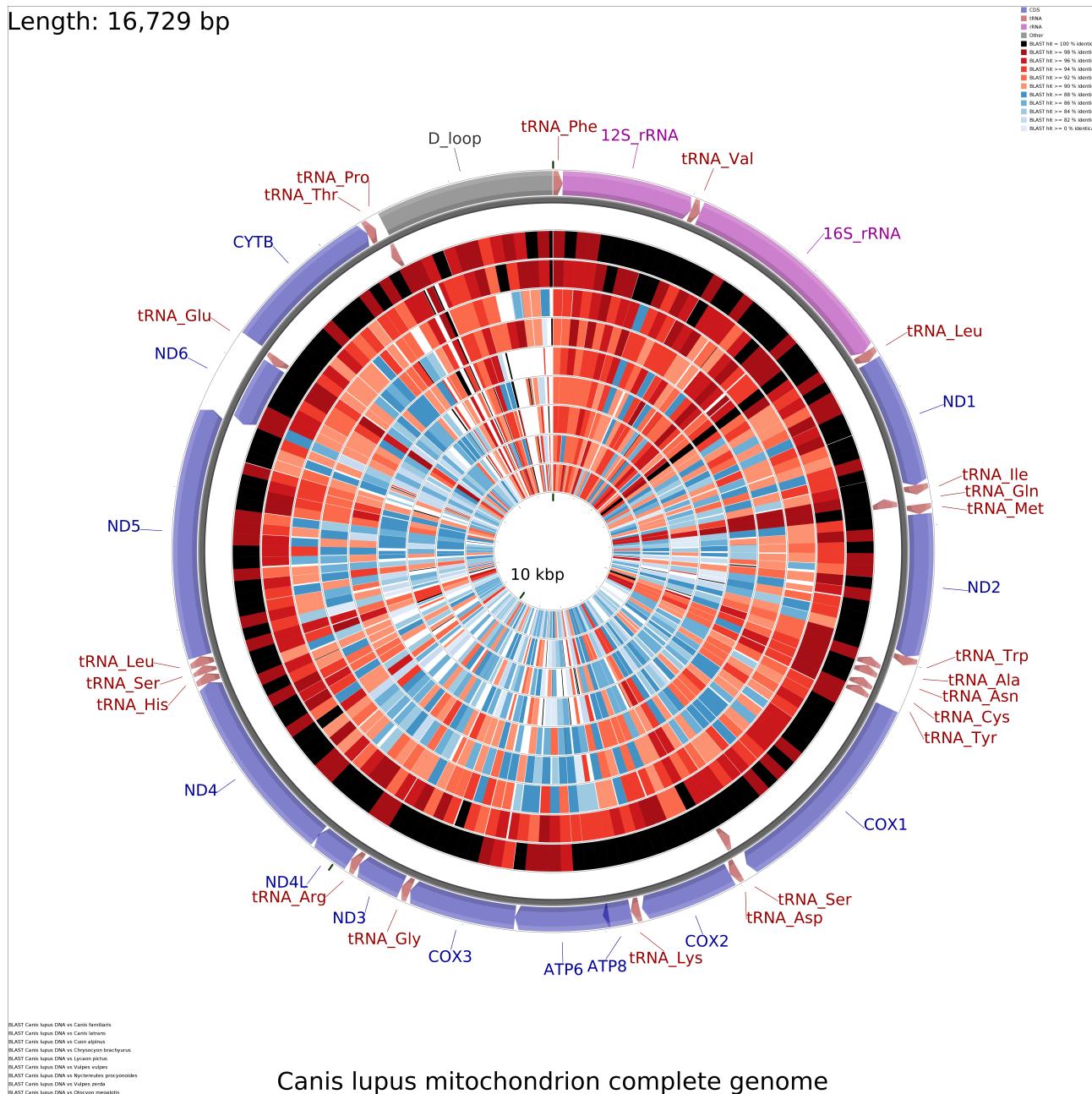
```
docker run --rm -v "$(pwd)":/dir -w /dir pstothard/cgview_comparison_tool  
build_blast_atlas.sh -p Projekt build_blast_atlas.sh -p Projekt
```

Następnie odpowiednio przerabiamy plik GFF z folderu features, tak aby przypominał poniższy zrzut ekranu.

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

Kolejnym krokiem będzie zmodyfikowanie pliku dna_vs_dna_large.xml. Po jego modyfikacji generujemy mapę ze zmianami, za pomocą komendy:

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



Analiza mapy porównawczej genomów

Jak prezentuje powyższa analiza najwięcej wspólnych genów posiadają *Canis lupus* oraz *Canis lupus familiaris*, co świadczy o bliskości pokrewieństwa naszych psów domowych z Wilkami szarymi żyjącymi na wolności. Wraz z zwiększeniem się różnic występowania wybranych psowatych jak i różnic fenotypowych można dostrzec zmniejszanie się ilości wspólnych genów w genotypach wybranych gatunków. Dla porównania genotypów psa jak i wilka moglibyśmy dodać również Dingo (*Canis dingo*, który wcześniej nazywał się *Canis lupus dingo*) - jednak jego sekwencji nie udało nam się odnaleźć w bazie NCBI.

Najwięcej wspólnych odcinków dla naszych genomów znajduje się na odcinku 16S_RNA i 12S_RNA. Różnice w genach dla psowatych świadczą o ich dużej różnorodności genetycznej, które też rzutują na ich przystosowanie do środowiska w którym żyją. W naszych analizach porównywaliśmy między innymi Wilka szarego, Lisa rudego, Fenka pustynnego czy Jenota australijskiego, które oprócz różnic wizualnych żyją w odmiennych niszach ekologicznych, co rzutuje na ich odmienne przystosowania do warunków ekologicznych. W przypadku podobnych sekwencji możemy spodziewać się przypadkowych krzyżówek gatunków, które negatywnie wpływają na czystość populacji. (Tak jak jest to w przypadku krzyżówek psów i dingo, czy psów i wilków). Podobne sekwencje mogą również oznaczać, że gatunki mogły pochodzić od wspólnego przodka jednak przez czynniki środowiskowe czy też hodowlane krzyżowały się z osobnikami podobnymi czy o określonych cechach fizycznych co spowodowało specjacje.

Analiza filogenetyczna

Analiza filogenetyczna została przeprowadzona na 80 osobnikach rodziny psowatych, wybranym genem do analizy został Cytochrom B.

1. Przygotowanie sekwencji

Wszystkie wykorzystane przez nas sekwencje pochodzą z bazy GenBank, 80 z nich pochodzi z rodziny psowatych, a 4 pozostałe z innego rzędu (łuskowce), zostały umieszczone celem zagnieźdżenia drzewa.

Gatunki zwierząt poddanych analizie:

Nazwa gatunku	Ilość wykorzystanych sekwencji
Vulpes vulpes	10
Vulpes Rueppellii	6
Vulpes lagopus	10
Canis lupus	10
Canis lupus familiaris	11
Otocyon megalotis	3
Canis aureus	10
Urocyon cinereoargen	5
Nyctereutes procyonoides	5
Cuon alpinus	3
Lycaon pictus	3
Canis anthus	1
Canis lupaster	1

Nazwa gatunku	Ilość wykorzystanych sekwencji
Lycalopes	1
Canis simensis	1
Gatunki z innego rzędu	4

Ich nieregularna ilość wynika ze względu na ich dostępność w wykorzystywanej bazie danych.

2. Następnie wszystkie sekwencje, umieszczamy w jednym pliku, w formacie FASTA.

Sekwencje — Notatnik

Plik Edycja Format Widok Pomoc

>Vulpes_vulpes_AB292765.1

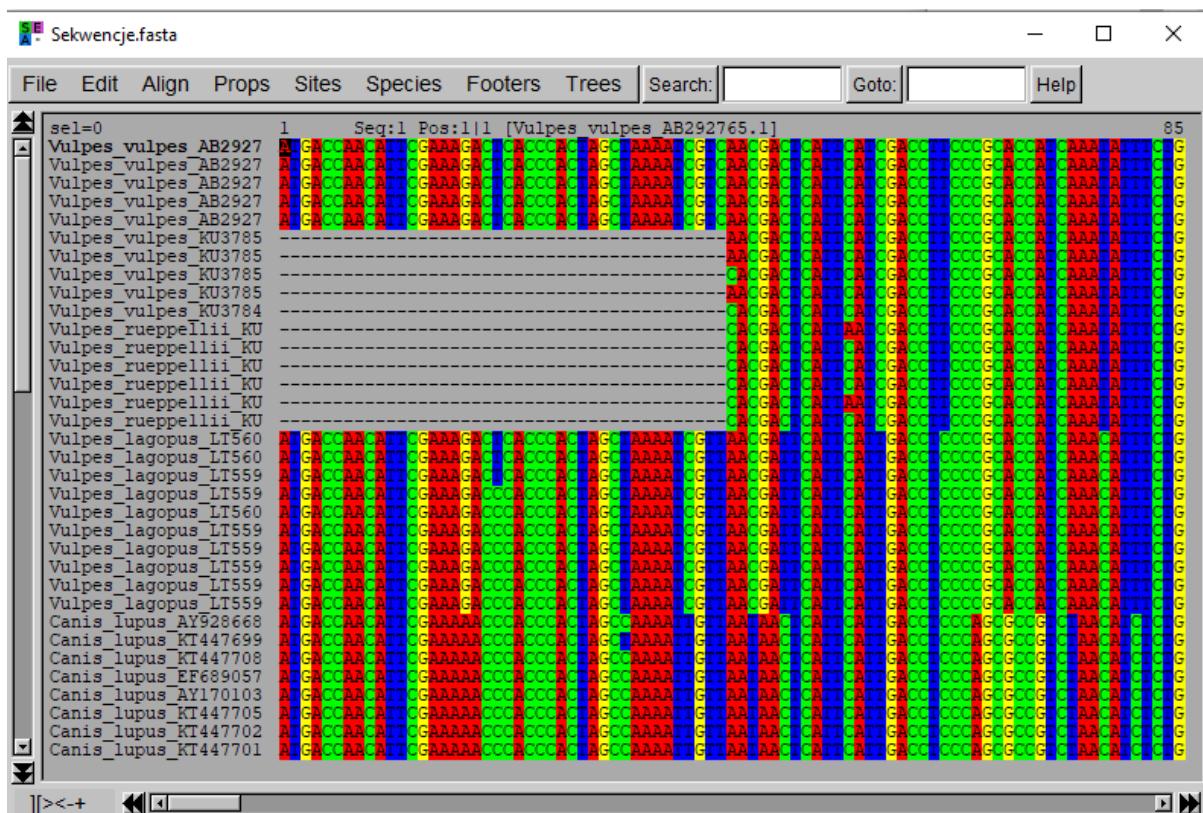
ATGACCAACATTGAAAGACTCACCCACTAGCTAAAATGTCAACGACTCATTGACCTTCCGCAC
CATCAAATATTCTGCTGATGGAACCTCGGGTCCCTGCTGGGTGTATGCCTTATTCTACAGATTGCAAC
AGGTCTATTCTAGCCATACACTATACTACATCCGACACAGCTACTGCTTTTCTACAGTCACTCATATCTGC
CGAGACGTTAACATATGGCTGAATTATCGTTACATACATGCAAACGGAGCATCTATATTCTATCTGCC
TCTTCATGCACTAGGACGAGGCCTATACTATGGATCTTATGTATTCTAGAAACATGAAATATCGGAAT
TATCTTATTGTCGCAACCATTGCCACAGCATTATGGGCTACGTTTACCATGAGGACAAATATCATT
TGAGGGGCAACCGTAATTACAATCTTCTATCTGCTATCCCTATATCGGAACCGATCTAGTAGAATGAA
TTTGAGGGGGCTTCTCAGTAGACAAAGCAACCTAACACGATTCTTGATTCCACTTATCCTTCCATT
CATCATCGCAGCATTAGCGATAGTCCATCTTATTCTCAGGAAACAGGATCCAACAAACCCCTCAGGA
ATCACATCCGACTCAGACAAAATTCCATTCCACCCCTACTACACAGTTAAAGACATCCTAGGAGTCTTAC
TTCTCCTTTCAGTTCTAAATCGCTAGTCTTCTCAGGAGATCTTCTAGGAGACCCAGACAATTACAC
CCCCGAAACCCCCCTCAGCACTCCCCCACATATCAAACCCGAGTGATATTCTTCTAGGAGTCTTAC
CTCCGGTCCATCCCTAATAAACTAGGAGGAGTCTTAGCCCTAGTGTCTCAATCTAACACTCA
TCCCTCACCTACACACCTCAAAACAAACGTGGGATAATATTCCGACCCCTAGTCAATGCTTACT
TCTAACCGCAGACCTCCTCACCTAACCTGAATCGGAGGGCAGCCGGTAGAACACCCCCTTTATTATCATT
GGCAAATTGCCTCCATCCTGATTTGCTATCTTATTGATTCTAAATGCCAACCTAGCATTATCGAAA
ACAATCTCCTAAAATGAAGA

>Vulpes_vulpes_AB292764.1

ATGACCAACATTGAAAGACTCACCCACTAGCTAAAATGTCAACGACTCATTGACCTTCCGCAC
CATCAAATATTCTGCTGATGGAACCTCGGGTCCCTGCTAGGTGTATGCCTTATTCTACAGATTGCAAC
AGGTCTATTCTAGCCATACACTATACTACATCCGACACAGCTACTGCTTTTCTACAGTCACTCATATCTGC
CGACACCTAACATCCCTAACATACATGCAAACCCACCATCTATATTCTTATCTGCC

3. W tym kroku wykorzystamy program SeaView. W pierwszej kolejności, ładujemy plik, następnie przyrównujemy sekwencje w następujący sposób:

Aligment wygląda w następujący sposób:

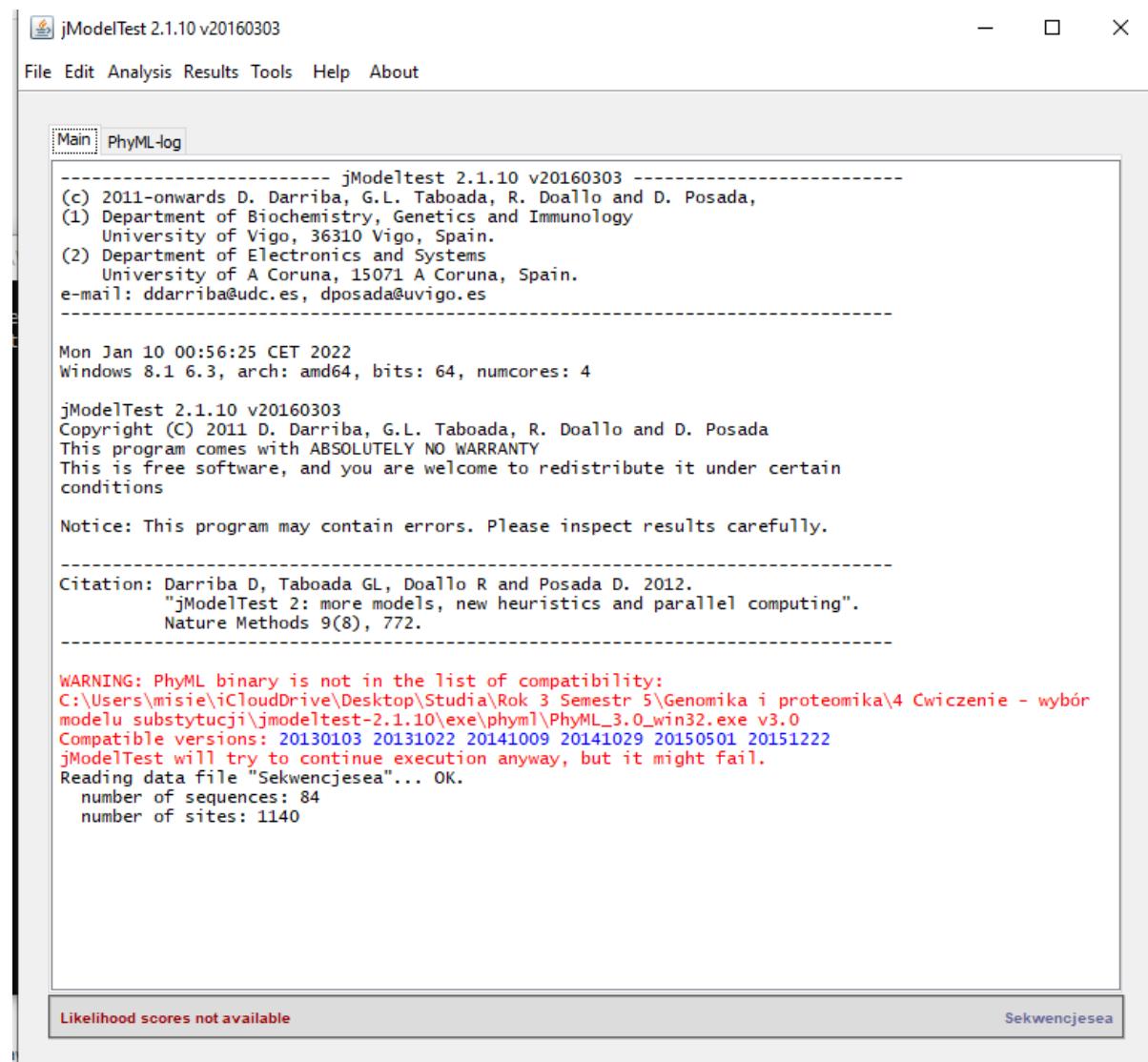


4. W programie SeaView, zaznaczamy pozycje nukleotydów w kodonach (osobno dla pierwszej pozycji, drugiej i trzeciej). Następnie każdy set zapisujemy w osobnym pliku.

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

Same selekcje zapisujemy jeszcze w formacie NEXUS, przyda się to później podczas tworzenia drzewa filogenetycznego.

5. Od teraz będziemy wykorzystywać program jModelTest, celem wyznaczenia optymalnych modeli dla naszych sekwencji. Najpierw ładujemy nasz plik z aligmentem.



The screenshot shows the jModelTest 2.1.10 v20160303 application window. The menu bar includes File, Edit, Analysis, Results, Tools, Help, and About. The main window has two tabs: Main (selected) and PhyML-log. The PhyML-log tab displays the following text:

```
----- 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.

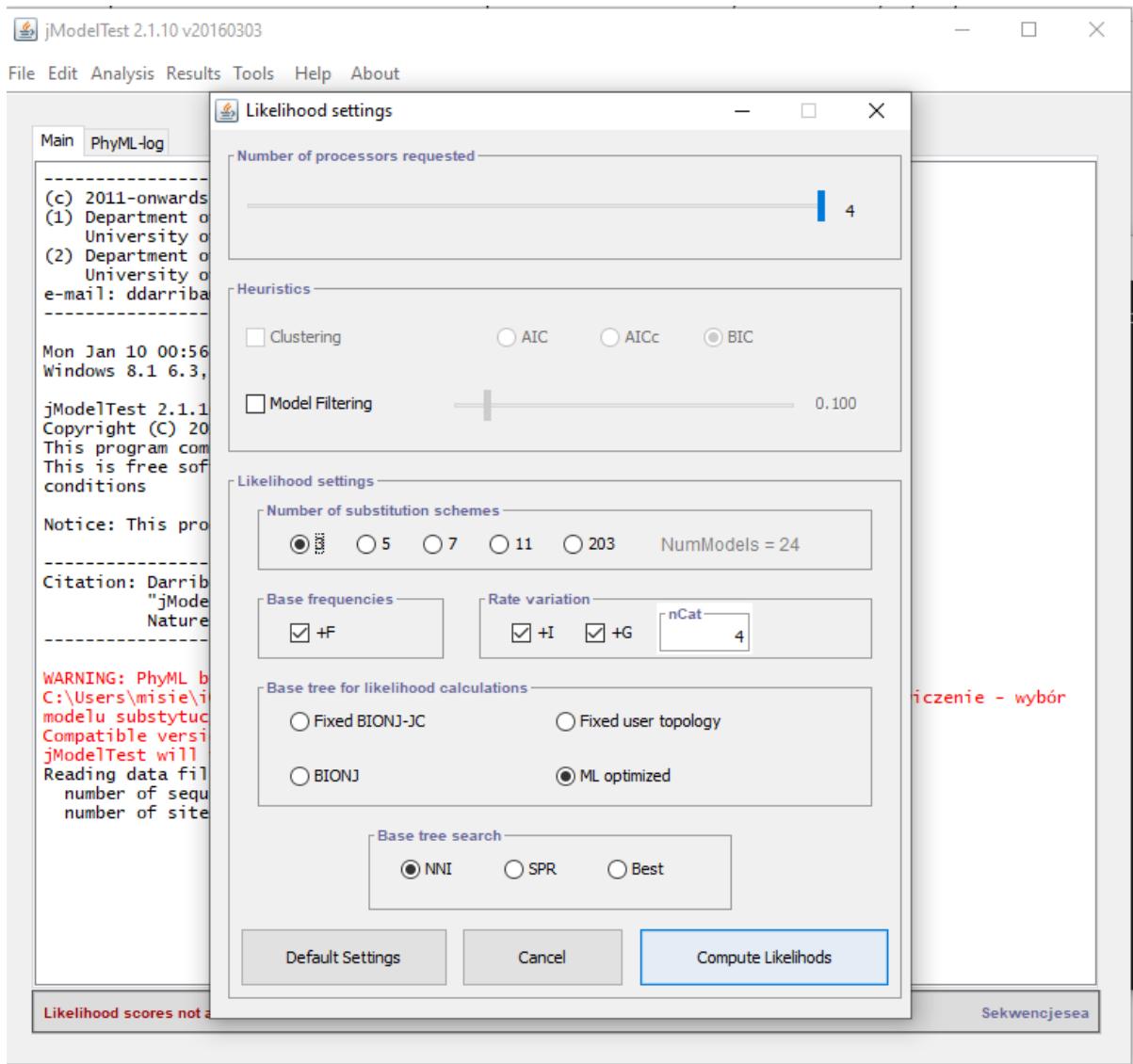
-----
Citation: Darriba D, Taboada GL, Doallo R and Posada D. 2012.
    "jModelTest 2: more models, new heuristics and parallel computing".
    Nature Methods 9(8), 772.

-----
WARNING: PhyML binary is not in the list of compatibility:
C:\Users\misie\CloudDrive\Desktop\Studio\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 "Sekwencjesea"... OK.
    number of sequences: 84
    number of sites: 1140
```

At the bottom of the log window, there are two status messages: "Likelihood scores not available" and "Sekwencjesea".

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



6. Kolejnym krokiem jest wyznaczenie wartości statystyk AIC, AICc i BIC, na podstawie których wybierzemy nasze modele.

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

Tabela wyników wygląda następująco:

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

W tabeli szukamy modelu który ma najlepsze dopasowanie, jest zaznaczony czerwonym kolorem.

Wszystkie powyższe kroki wykonujemy dla całych genów, dla pierwszego, drugiego oraz trzeciego kodonu, wszystkie dane zbieramy w tabelce:

	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

Do dalszej analizy wybraliśmy modele wyznaczone przez statystyke AICc.

7. Wykorzystujemy program MrBayes, celem wyznaczenia naszego drzewa filogenetycznego. Na początku ładowujemy nasz plik.

```
MrBayes > execute "C:\Users\misie\iCloudDrive\Desktop\Studia\Rok 3 Sem 1\Wojciech\Kodony.nex"

Executing file "C:\Users\misie\iCloudDrive\Desktop\Studia\Rok 3 Sem 1\Wojciech\Kodony.nex"
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
```

Następnie tworzymy zbiory (charsety) o długości sekwencji 1140. Później tworzymy partycje o nazwie "kodony"

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

Następnie dla poszczególnych charsetów, przypisujemy odpowiednie modele substytucji. Następnie uruchamiamy program, określając liczbę kroków w łańcuchach Markova

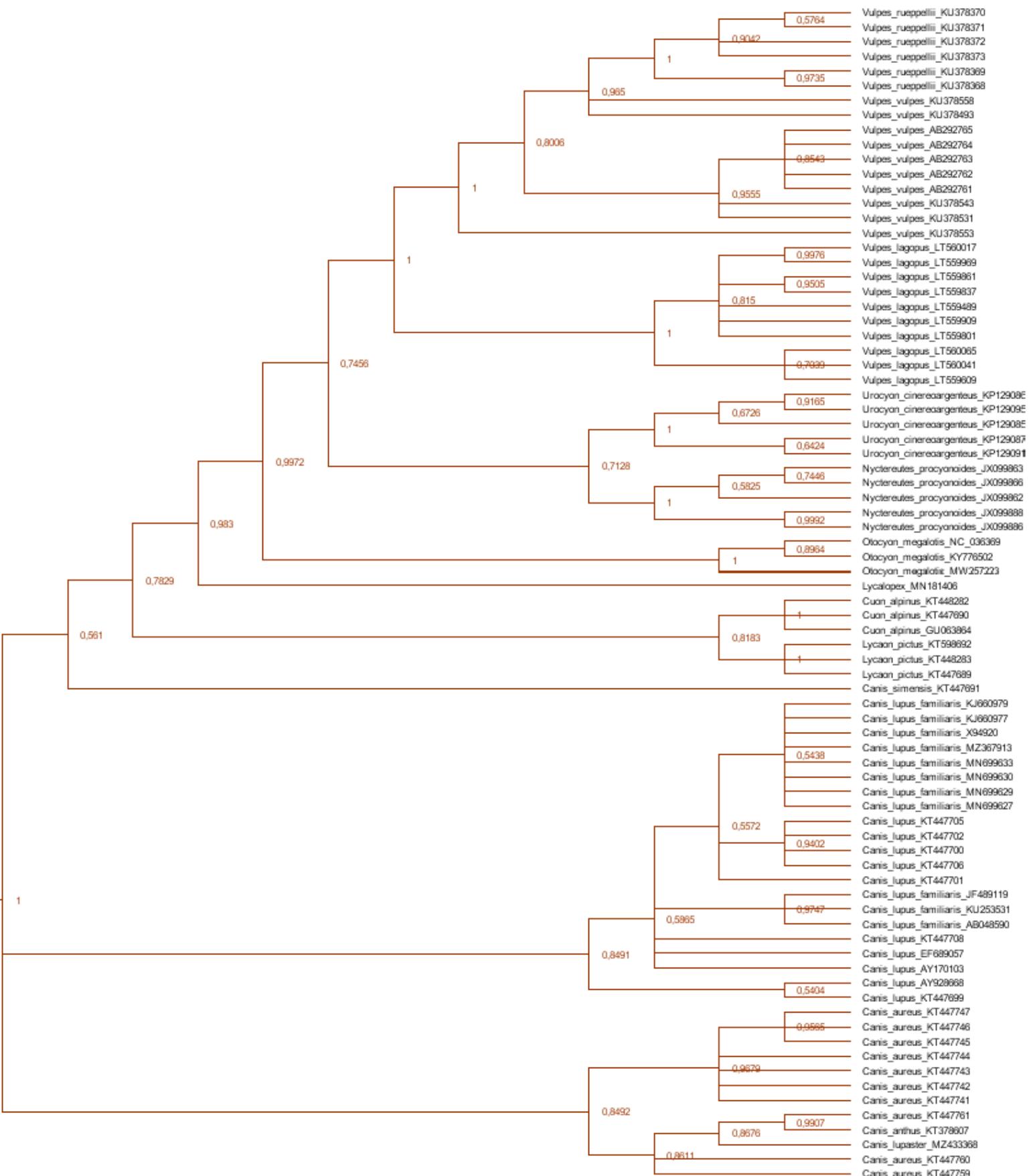
```
lset applyto=(1,3) nst=2 rates=gamma
lset applyto=(2) nst=2 rates=propinv
prset applyto=(3) pinvarpr=fixed(0)
mcmc ngen=100000
```

```
Using a relative burnin of 25.0 % for diagnostics
Chain results (100000 generations requested):
4495000 -- [-8231.18] (-8222.04) [-8182.85] (-8191.59) (-8247.70) (-8195.50) [-8190.74] (-8228.59) -- 0:00:48
Average standard deviation of split frequencies: 0.006128
4496000 -- (-8252.74) (-8213.49) [-8174.44] (-8203.88) * (-8236.75) [-8191.56] (-8227.54) (-8216.73) -- 0:00:38
4497000 -- (-8245.12) (-8217.27) (-8197.80) [-8195.76] * (-8235.18) (-8178.15) (-8228.12) [-8187.53] -- 0:00:28
4498000 -- (-8228.76) (-8240.66) (-8196.19) [-8199.06] * (-8220.54) (-8188.64) (-8226.57) [-8190.79] -- 0:00:19
4499000 -- (-8222.08) (-8226.95) (-8184.15) [-8178.65] * (-8220.39) [-8184.61] (-8222.67) (-8210.27) -- 0:00:09
4500000 -- (-8218.36) (-8222.63) [-8185.42] (-8211.06) * (-8226.96) [-8186.96] (-8253.11) (-8226.61) -- 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 -- (-8225.40) (-8262.63) (-8234.35) [-8219.14] * [-8207.29] (-8218.18) (-8243.42) (-8191.38) -- 0:16:10
12000 -- (-8219.23) (-8250.46) [-8213.93] (-8207.41) * [-8170.11] (-8211.67) (-8236.85) (-8203.96) -- 0:16:00
13000 -- [-8206.15] (-8228.83) (-8209.52) (-8208.08) * (-8192.85) (-8229.99) (-8235.80) [-8188.91] -- 0:15:50
14000 -- (-8215.96) (-8219.94) [-8207.30] (-8228.05) * (-8198.34) (-8208.28) (-8254.10) [-8184.85] -- 0:15:39
15000 -- (-8223.29) [-8208.20] (-8223.39) (-8239.67) * (-8190.43) (-8238.77) (-8252.06) [-8181.96] -- 0:15:40
Average standard deviation of split frequencies: 0.088651
```

Liczbę kroków w łańcuchach Markova, zwiększamy aż "Avarage standard deviation of split frequencies", nie spadnie poniżej 0.01. Gdy to się stanie, wykonujemy jeszcze trzykrotność dotychczasowych kroków. Dopiero po wykonaniu tych obliczeń, otrzymamy konsensusowe drzewo.

W naszym przypadku, wartość ta spadła poniżej 0.01, po około 1.1 miliona kroków. Sumarycznie, ilość wykonanych kroków wyniosła 4.5 miliona.

```
4495000 -- (-8231.18) (-8222.04) [-8182.85] (-8191.59) (-8247.70) (-8195.50) [-8190.74] (-8228.59) -- 0:00:48
Average standard deviation of split frequencies: 0.006128
4496000 -- (-8252.74) (-8213.49) [-8174.44] (-8203.88) * (-8236.75) [-8191.56] (-8227.54) (-8216.73) -- 0:00:38
4497000 -- (-8245.12) (-8217.27) (-8197.80) [-8195.76] * (-8235.18) (-8178.15) (-8228.12) [-8187.53] -- 0:00:28
4498000 -- (-8228.76) (-8240.66) (-8196.19) [-8199.06] * (-8220.54) (-8188.64) (-8226.57) [-8190.79] -- 0:00:19
4499000 -- (-8222.08) (-8226.95) (-8184.15) [-8178.65] * (-8220.39) [-8184.61] (-8222.67) (-8210.27) -- 0:00:09
4500000 -- (-8218.36) (-8222.63) [-8185.42] (-8211.06) * (-8226.96) [-8186.96] (-8253.11) (-8226.61) -- 0:00:00
Average standard deviation of split frequencies: 0.006143
Continue with analysis? (yes/no):
Additional number of generations:
```



Analiza drzewa filogenetycznego

Drzewo zostało ukorzenione do sekwencji organizmów niepsowatych. Zaczynając od korzenia, możemy dostrzec dwa rozgałęzienia:

- Dolne, w którym znajdują się prawie wszystkie gatunki wilków (*Canis*), jedynym wyjątkiem, jest *Canis Simensis*, który jest genetycznie oddalony od pozostałych, jednak powinien się znaleźć w tej gałęzi
- Górne, w którym znajdują się pozostałe

Kolejnym rozgałęzieniem, jest to w którym omylnie pojawił się *Canis Simensis*, stało się tak prawdopodobnie ze względu na jakość użytej sekwencji z bazy GenBank. Niska wartość prob, wskazuje na losowość tego dopasowania.

Kolejnym rozgałęzieniem jest oddzielenie grupy zewnętrznej lajkonów (*Lyacon Pictus*) oraz cyjonów (*Coua Alpinus*), wartości prob, w obu przypadkach są wysokie, jest to dobre dopasowanie.

W kolejnym rozgałęzieniu, z dużą wartością prob, oddziela się grupa zewnętrzna złożona z jednego gatunku - Nibylis (*Lycalopex*)

Następnym rozgałęzieniem jest oddzielenie się Otocjonów (*Otocyon Megalotis*). Również tutaj wartość prob jest bardzo duża.

Kolejne rozgałęzienie jest kluczowe:

- Niżej, znajdują się osobniki Urocjonów (*Urocyon cinereoargenteus*) oraz jenotów (*Nyctereutes procyonoides*)

- Wyżej znalazły się już tylko gatunki lisów (*Vulpes vulpes*, *Vulpes lagopus*, *Vulpes ruppelli*).

Wartości prob, ponownie były wysokie.

- Dalej różnice były już poza samymi tymi gatunkami, *Vulpes lagopus* odłączył się pierwszy, później *Vulpes vulpes*, na samym końcu *Vulpes ruppelli*. Różnice między dwoma ostatnimi są na tyle znikome, że dwie sekwencje *Vulpes vulpes* zostały przypisane do gałęzi na której znajduje się *Vulpes ruppeli*. Pozostałe różnice są już tylko wewnątrz danego gatunku.

Porównanie z literaturą

A molecular phylogeny of the Canidae based on six nuclear loci

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

Przechodząc do analizy drzewa filogenetycznego, porównanie naszego drzewa z literaturą wskazuje na pewne odstępstwa. Opisane zostaną tylko różnice.

- *Urocyon cinereoarctos* w naszym drzewie znajduje się w jednej gałęzi razem z *Nyctereutes procyonoides*, według literatury jednak ten gatunek odłączył się znacznie wcześniej.
- Gatunki *Cuon alpinus* oraz *Lycaon pictus*, według naszego drzewa nie należą do gałęzi na której znajdują się gatunki *Canis*, według literatury, należą do niej ale szybko się od niej odłączyły.
- *Vulpes lagopus* w naszym drzewie odłączył się przed *Vulpes vulpes*. Według literatury jednak *Vulpes lagopus* odłączył się ostatni.

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

- Ponownie porównując z literaturą, nasze gatunki *Cuon* i *Lycaon*, zostały umieszczone w złym miejscu. (zostały tam umieszczone z prob 0,561, co jest niską wartością).

Wszystkie nieprawidłowości które zostały wykazane w naszej analizie, można解释. Analiza którą przeprowadzaliśmy, odwoływała się wyłącznie do zmienności genu CytochromB, a nie całego genomu. Ponadto, drzewa literaturowe, do których porównywaliśmy nasze, nie możemy traktować jako pewne i idealne. Zostały one stworzone podobnymi technikami, tylko na podstawie bardziej rozbudowanych danych.

Źródła

<https://ocdn.eu/pulscms-transforms/1/xIwk9kpTURBXy8xYjVhYTE5NmY1MmNlODgzZDZhYmVjYWExZjNlOGJlYi5qcGeTIQMBzQE4zQ-gzOjKkwXNAxTNAbyTCaYwNTI4YzkGgaEwBQ/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>