

3. FYLOGENETICKÁ ANALÝZA ŠELEM NA ZÁKLADĚ MITOCHONDRIÁLNÍHO GENU *COX1*

Cíl úlohy

Rychlost a frekvence mutací mitochondriálního genomu jsou vhodné pro vyhodnocení genetických polymorfismů ve vztahu s fylogenetickým vývojem a taxonomií na úrovni řádů, čeledí a podčeledí. V následující úloze vycházejte z hypotézy, že všichni vybraní zástupci šelem měli společného předka. Tudíž tuto skupinu druhů budeme chápat jako monofyletickou. Cílem úlohy je vypracování fylogenetické analýzy na základě metody maximální věrohodnosti (Maximal Likelihood Method – ML). Prvním důležitým krokem je správně provést alignment výchozích sekvencí genu *Cox1*, který je frekventovaně používán nejen pro fylogenetické analýzy, ale je rovněž aplikován pro DNA barcoding systém zacílený na identifikaci druhů. Vzhledem k tomu, že *Cox1* je strukturální gen kódující protein, je vhodné zvolit takový typ alignmentu, který pracuje s kodóny. Analýzy mtDNA vychází z předpokladu homoplazmie a haplotypové analýzy. Haplotypy je nezbytné identifikovat vhodným programem. Cílem této úlohy je naučit pracovat studenty s programovým balíkem MEGA X, který umožňuje provedení několika typů alignmentu, výběru vhodného substitučního modelu pro metodu ML a následnou konstrukci a editaci stromu – fylogramu.

Vstupní data

- Sekvence mtDNA ve formátu FASTA, které byly získány z databáze NCBI. Sekvence představují vlákna s orientací 5'-3'.

Potřebné bioinformatické nástroje

- BioEdit 7.2 (<https://bioedit.software.informer.com/7.2/>)
- DNA Sequence Polymorphism v6.12.03 (<http://www.ub.edu/dnasp/downloadTv6.html>)
- MEGA X (https://www.megasoftware.net/dload_win_gui)
- NCBI (<https://www.ncbi.nlm.nih.gov/>)

Návod na řešení úlohy

1. Vstupními daty pro fylogenetickou analýzu jsou kompletní sekvence mitochondriálního genu *Cox1* u 35 druhů šelem. Sekvence představují vlákna DNA s orientací 5' - 3', která byla získána z mezinárodní bioinformatické databáze NCBI (<https://www.ncbi.nlm.nih.gov/>). Jako tak zvaný outgroup genotyp použijte sekvenci *Cox1* genu ježury (*Tachyglossus aculeatus*). Vstupní data jsou uložena ve formátu FASTA v příloze této úlohy (soubor COX1 šelmy před alignmentem.fas)

BioEdit Sequence Alignment Editor - [E:\Základy bioinformatiky\Fylogenetická analýza\COX1 šelmy před alignmentem.fas]

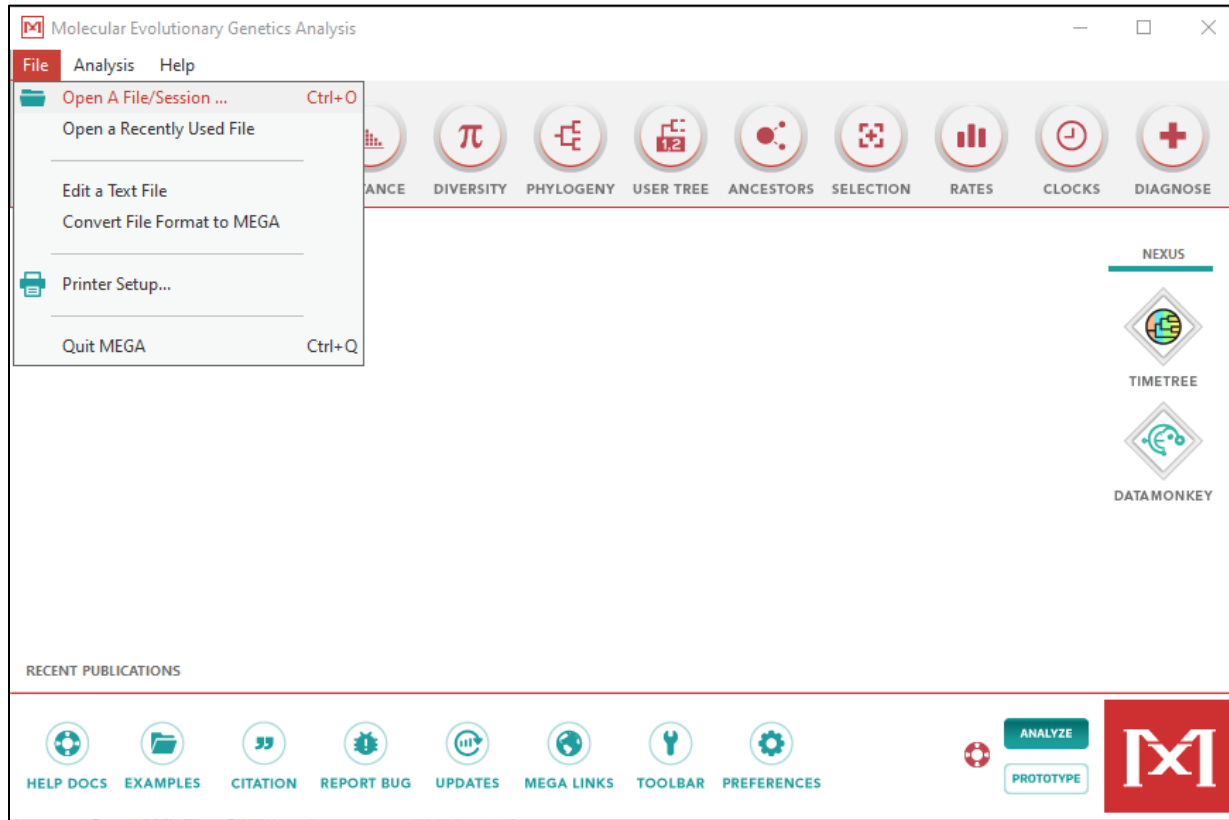
File Edit Sequence Alignment View Accessory Application RNA World Wide Web Options Window Help

36 total sequences

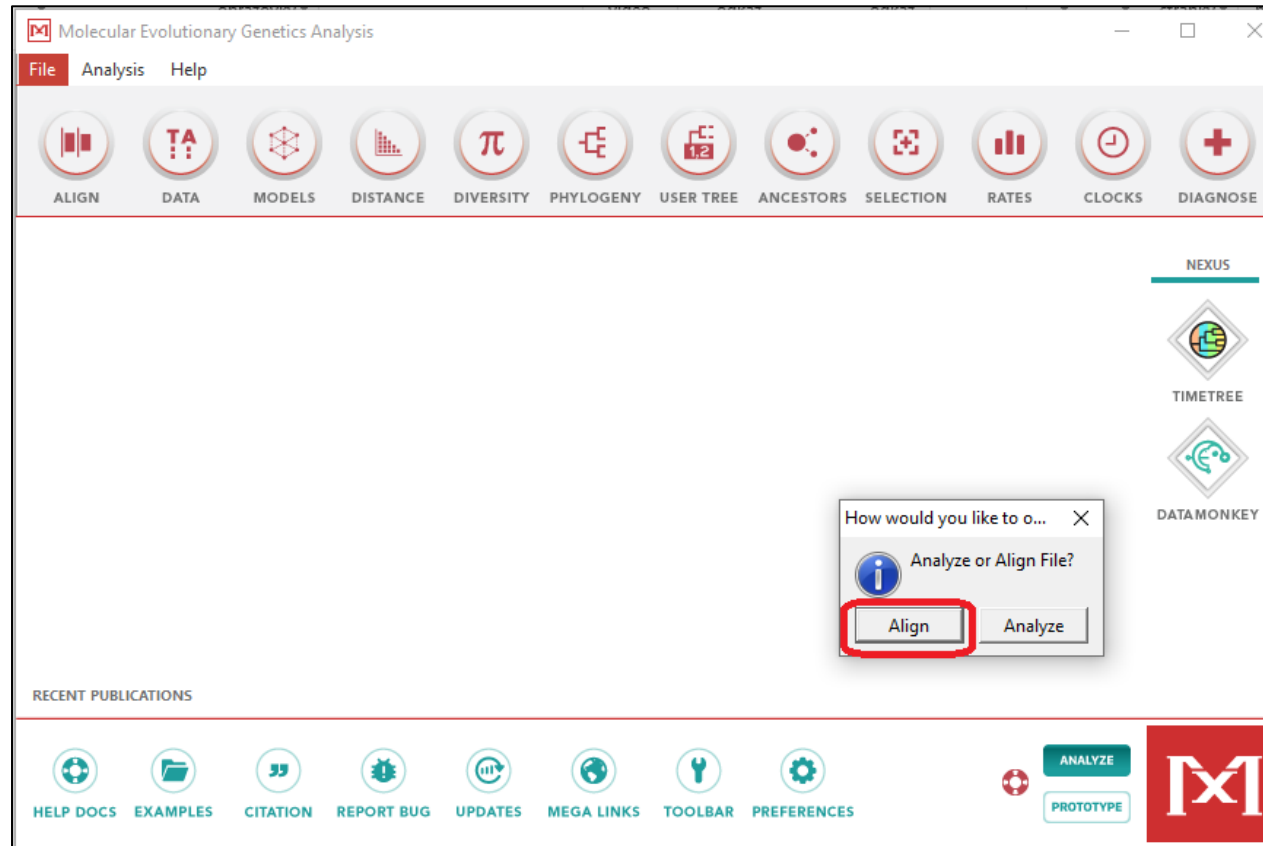
Mode: Select / Slide Selection: 0 Position: Sequence Mask: None Numbering Mask: None Start ruler at: 1

Canis latrans
Canis lupus familiaris
Canis lupus lupus
Crocota crocota
Felis catus
Felis silvestris
Hyaena hyaena
Lutra lutra
Lutra sumatrana
Lycaon pictus
Lynx canadensis
Lynx lynx
Lynx pardinus
Lynx rufus
Martes flavigula
Martes martes
Martes zibellina
Meles anakuma
Meles leucurus
Meles meles
Mustela erminea
Mustela nivalis
Mustela putorius
Mustela sibirica
Panthera leo
Panthera pardus
Parahyaena brunnea
Puma concolor
Tachyglossus aculeatus
Ursus americanus
Ursus arctos
Ursus maritimus
Ursus thibetanus thibetanus
Vulpes lagopus
Vulpes vulpes
Vulpes zerda

2. Pro fylogenetickou analýzu použijte program MEGA X. Prvním krokem je zadání souboru se vstupními daty. Jedná se o soubor ve formátu FASTA, který byl představen na předchozím obrázku. Otevřete tento soubor z umístění na vašem počítači.



3. Před vlastní analýzou je nutné provést mnohonásobné porovnávání analyzovaných sekvencí, pro které se mnohem častěji používá anglický termín Multiple Alignment. Po zadání vstupních dat je nutné potvrdit, že s daty budeme provádět právě tuto proceduru. Zadejte proto červeně onačenou možnost Align.



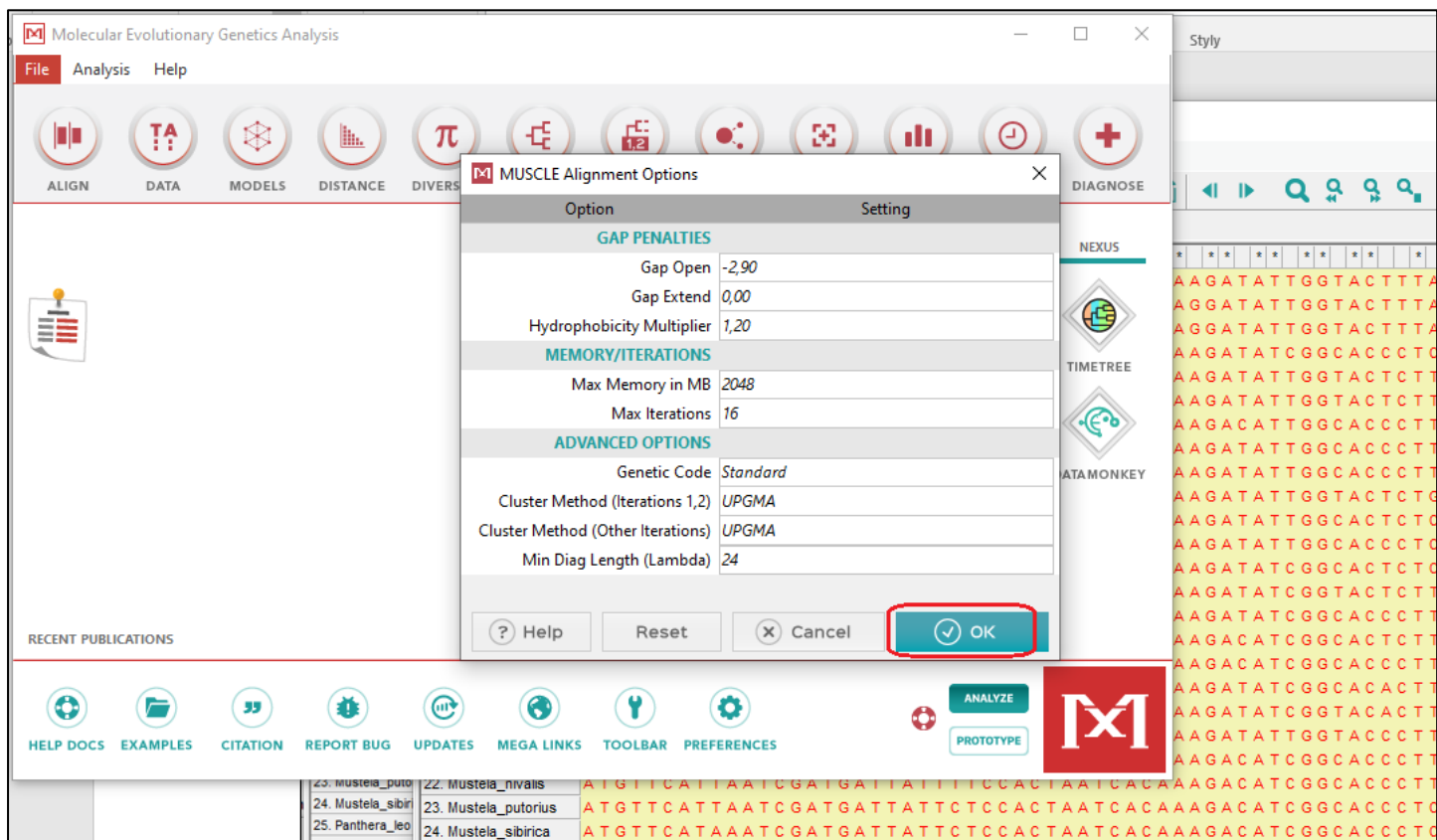
- Pro alignment sekvencí použijte algoritmus MUSCLE a to variantu, která porovnává sekvence na základě kodónů. Tento algoritmus je možné použít, protože se jedná o kódující sekvence, kdy první triplet kóduje první aminokyselinu. Jedná o sekvence mitochondriálního genu, který je tvořen pouze exonovými (kódujícími) oblastmi. Parametry analýzy, které musíte zvolit, jsou vyznačeny červenými rámečky.

The screenshot shows the MX: Alignment Explorer software interface. The title bar reads "MX: Alignment Explorer (COX1 želmy před alignmentem.fas)". The menu bar includes "Data", "Edit", "Search", "Alignment", "Web", "Sequencer", "Display", and "Help". The "Alignment" menu is open, showing options: "Align by ClustalW", "Align by ClustalW (Codons)", "Align by MUSCLE", and "Align by MUSCLE (Codons)". The "Align by MUSCLE (Codons)" option is highlighted with a red box. Other menu items include "Mark/Unmark Site (Ctrl+M)", "Align Marked Sites (Ctrl+L)", "Unmark All Sites", "Delete Gap-Only Sites", and "Auto-Fill Gaps".

The main window displays a multiple sequence alignment of COX1 sequences from 30 different species. The sequences are listed on the left, and the alignment is shown in the center. The alignment is color-coded by amino acid, with gaps represented by dashes. The species listed are: 1. *Canis latrans*, 2. *Canis lupus familiaris*, 3. *Canis lupus lupus*, 4. *Crocuta crocuta*, 5. *Felis catus*, 6. *Felis silvestris*, 7. *Hyaena hyaena*, 8. *Lutra lutra*, 9. *Lutra sumatrana*, 10. *Lycaon pictus*, 11. *Lynx canadensis*, 12. *Lynx lynx*, 13. *Lynx pardinus*, 14. *Lynx rufus*, 15. *Martes flavigula*, 16. *Martes martes*, 17. *Martes zibellina*, 18. *Meles anakuma*, 19. *Meles leucurus*, 20. *Meles meles*, 21. *Mustela erminea*, 22. *Mustela nivalis*, 23. *Mustela putorius*, 24. *Mustela sibirica*, 25. *Panthera leo*, 26. *Panthera pardus*, 27. *Parahyaena brunne*, 28. *Puma concolor*, 29. *Tachyglossus acule*, and 30. *Ursus americanus*.

At the bottom of the window, there is a "Site #" field with the value "1" and a "with" button.

6. Pro analýzu ponechejte všechny implicitně nastavené parametry. Alignment spusťte zvolením nabídky OK.



7. Před zahájením analýzy je nezbytné potvrdit, že při analýze budou kompletně vymazány tak zvané Gapy. Gapy jsou mezery, které algoritmus vytvoří porovnáváním sekvencí. Jedná se v podstatě o nukleotidy nebo skupiny nukleotidů, které algoritmus detekuje pouze u některých porovnávaných druhů. Jedná se o inserčně – deleční typ polymorfismů. Sekvence způsobující vznik gapů se nacházejí pouze u některých srovnávaných druhů. Proto je nelze využít k hodnocení substitučních změn napříč celým spektrem hodnocených druhů. Proto jsou ve fylogenetických analýzách obvykle oblasti gapů kompletně odstraněny.

Molecular Evolutionary Genetics Analysis

File Analysis Help

ALIGN DATA MODELS DISTANCE DIVERSITY PHYLOGENY USER TREE ANCESTORS SELECTION RATES CLOCKS DIAGNOSE

NEXUS

TIMETREE

DATAMONKEY

Remove Gaps

Would you like to remove gaps before alignment?

Yes No

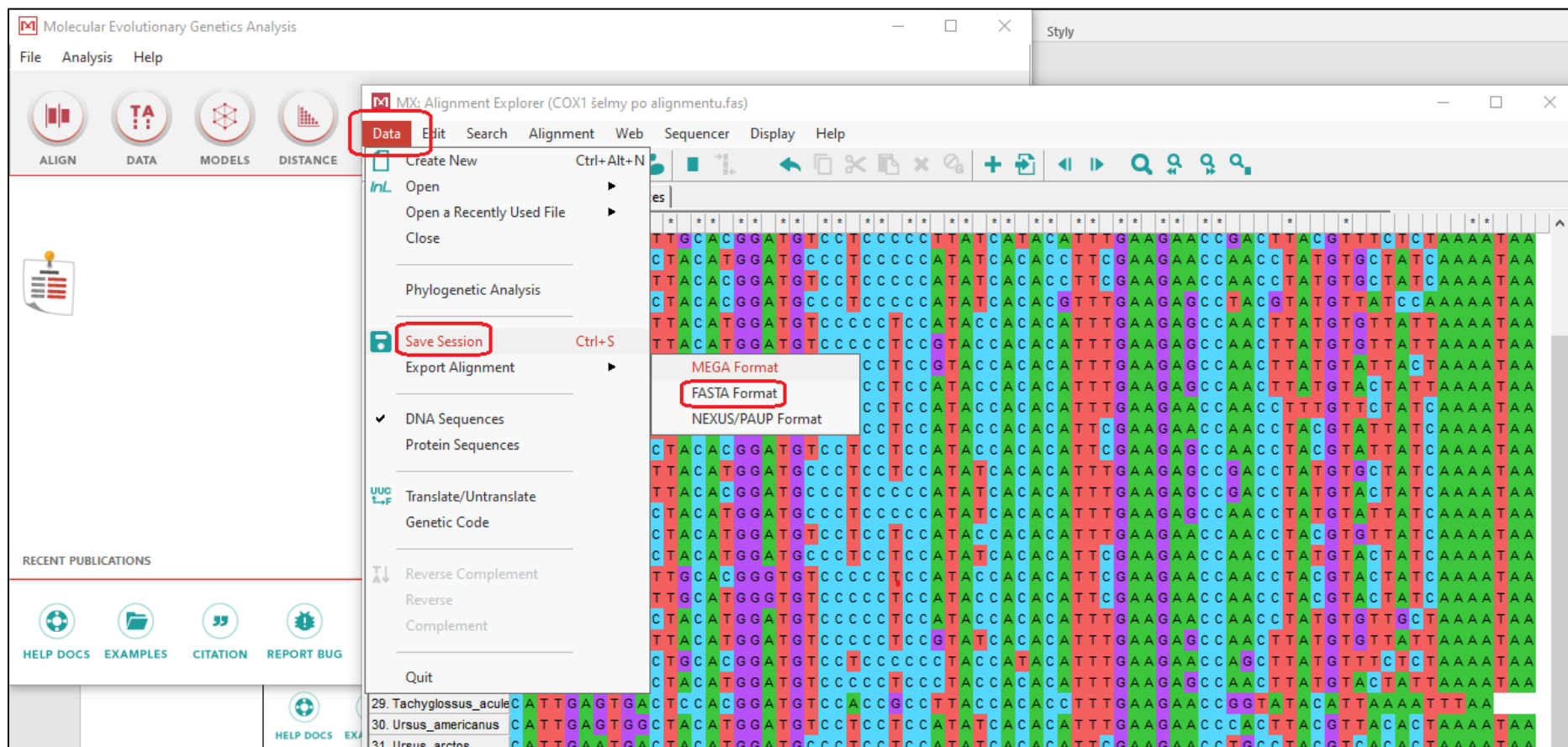
RECENT PUBLICATIONS

HELP DOCS EXAMPLES CITATION REPORT BUG UPDATES MEGA LINKS TOOLBAR PREFERENCES

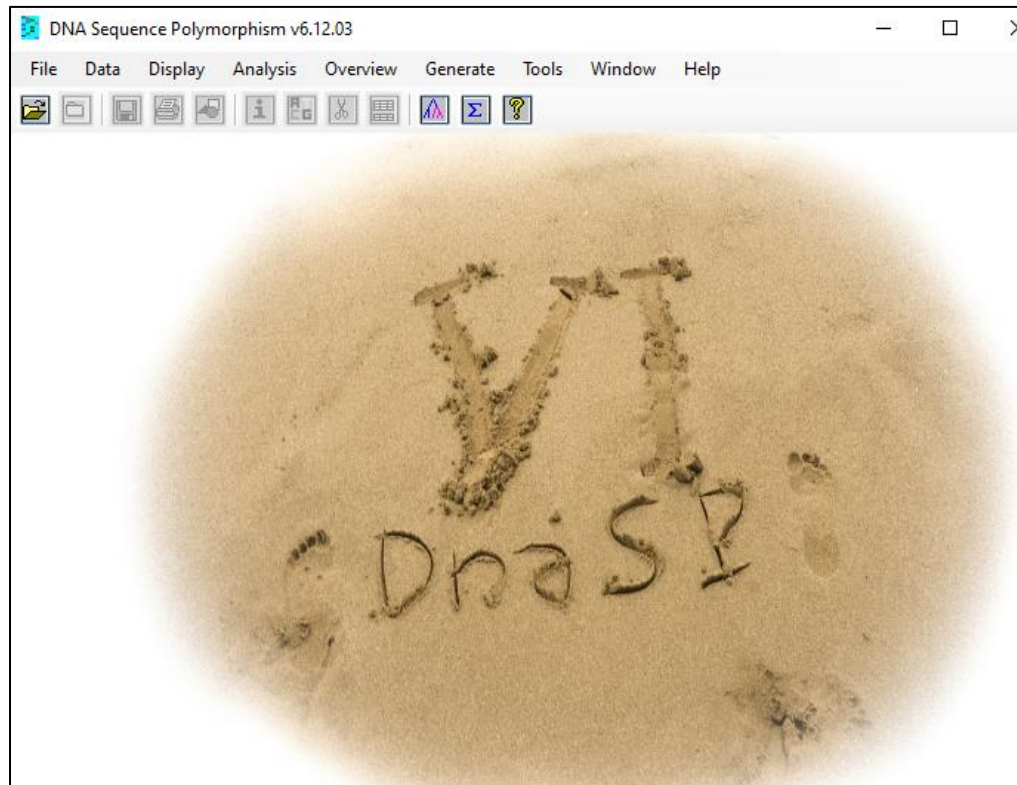
ANALYZE PROTOTYPE

22. *Mustela_nivalis* ATGTTTCATTAATCGATGATTATTTTCACATAATCACAAGACATCGGGCACCCT
23. *Mustela_putorius* ATGTTTCATTAATCGATGATTATTTTCACATAATCACAAGACATCGGGCACCCT
24. *Mustela_sibirica* ATGTTTCATAAATCGATGATTATTTTCACATAATCACAAGACATCGGGCACCCT
25. *Panthera_leo* ATGTTTCATAAACCGCTGACTATTTTCAACCAATCACAAGACATTTGGAACCT
26. *Panthera_pardus* ATGTTTCATAAACCGCTGACTATTTTCAACCAATCACAAGATATTGGAACCT

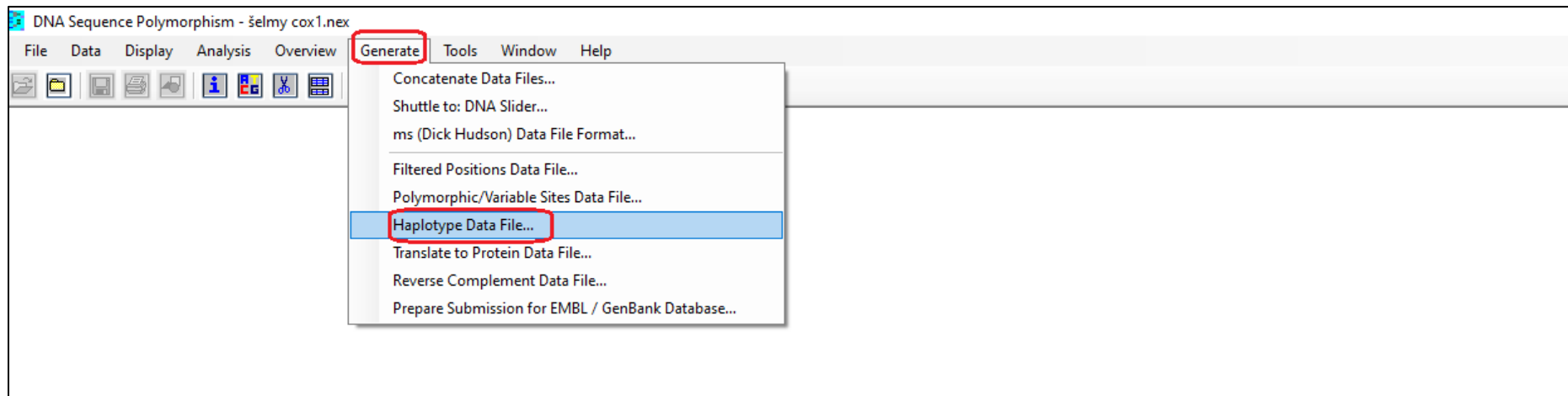
8. Pokud alignment proběhl, je nezbytné uložit jeho výsledek do vašeho počítače, a to ve formátu FASTA. Zvolte takové jméno souboru, aby bylo patrné, že představuje data po alignmentu.



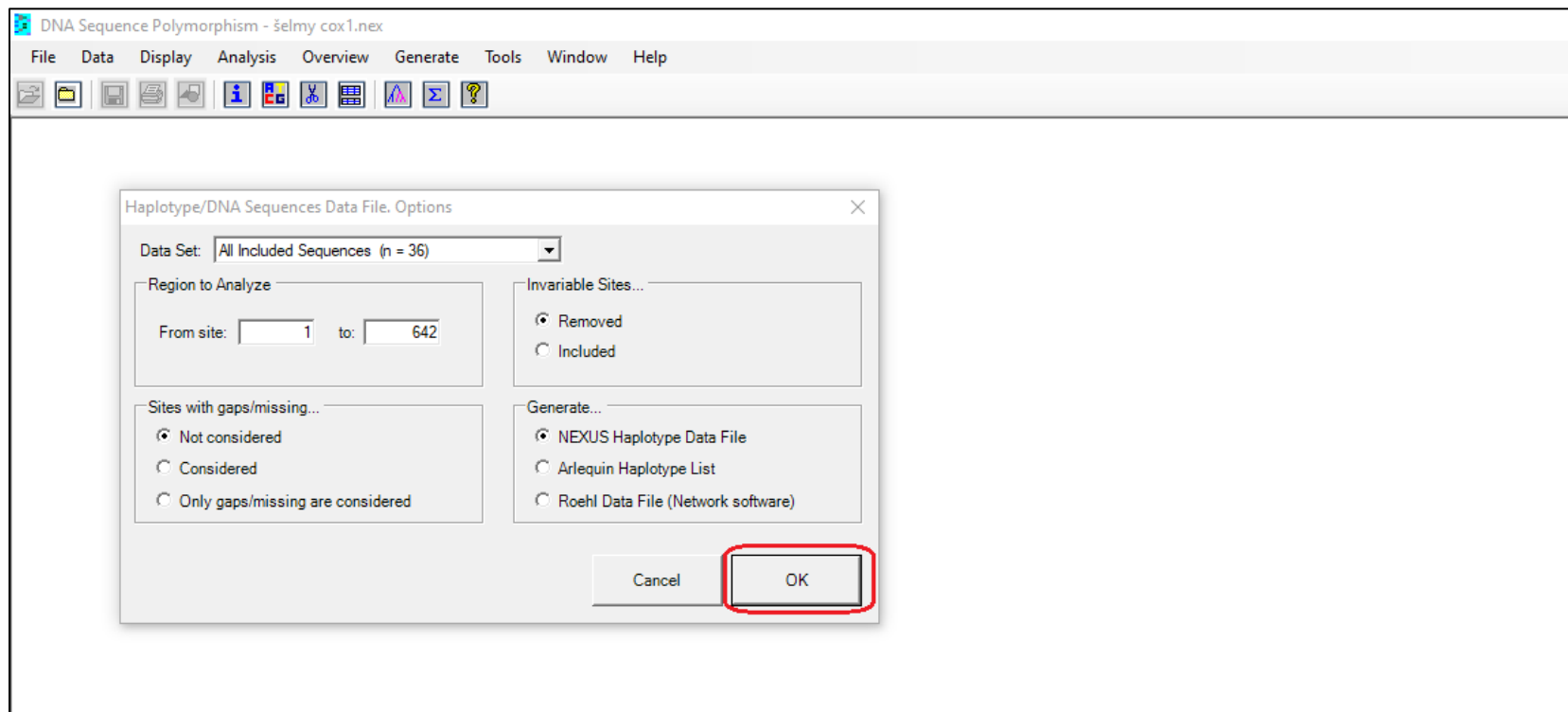
9. Analyzované sekvence mitochondriálního genu *Cox1* můžeme označit jako haplotypy. Mitochondriální genom je charakteristický matroklinní dědičností a homoplazmií. Pro fylogenetickou analýzu je potřeba pracovat s takovým datovým souborem, který bude obsahovat každý haplotyp pouze jedenkrát. Vzhledem k tomu, že ve vstupních datech je každý zoologický druh zastoupen pouze jedním jedincem, můžeme předpokládat, že každý zoologický druh představuje jeden haplotyp a sekvence haplotypů se tudíž vzájemně odlišují. Nicméně v souboru hodnocených šelem se nacházejí dvě domestikované formy (pes domácí a kočka domácí) včetně jejich nedomestikovaných předchůdců. V těchto případech je mnohem větší pravděpodobnost, že shodný haplotyp by se mohl vyskytovat jak u volně žijící, tak i u domestikované formy. Z těchto důvodů je nutné použít haplotypovou analýzu, a to pomocí programu DNA Sequence Polymorphism v6.12.03.



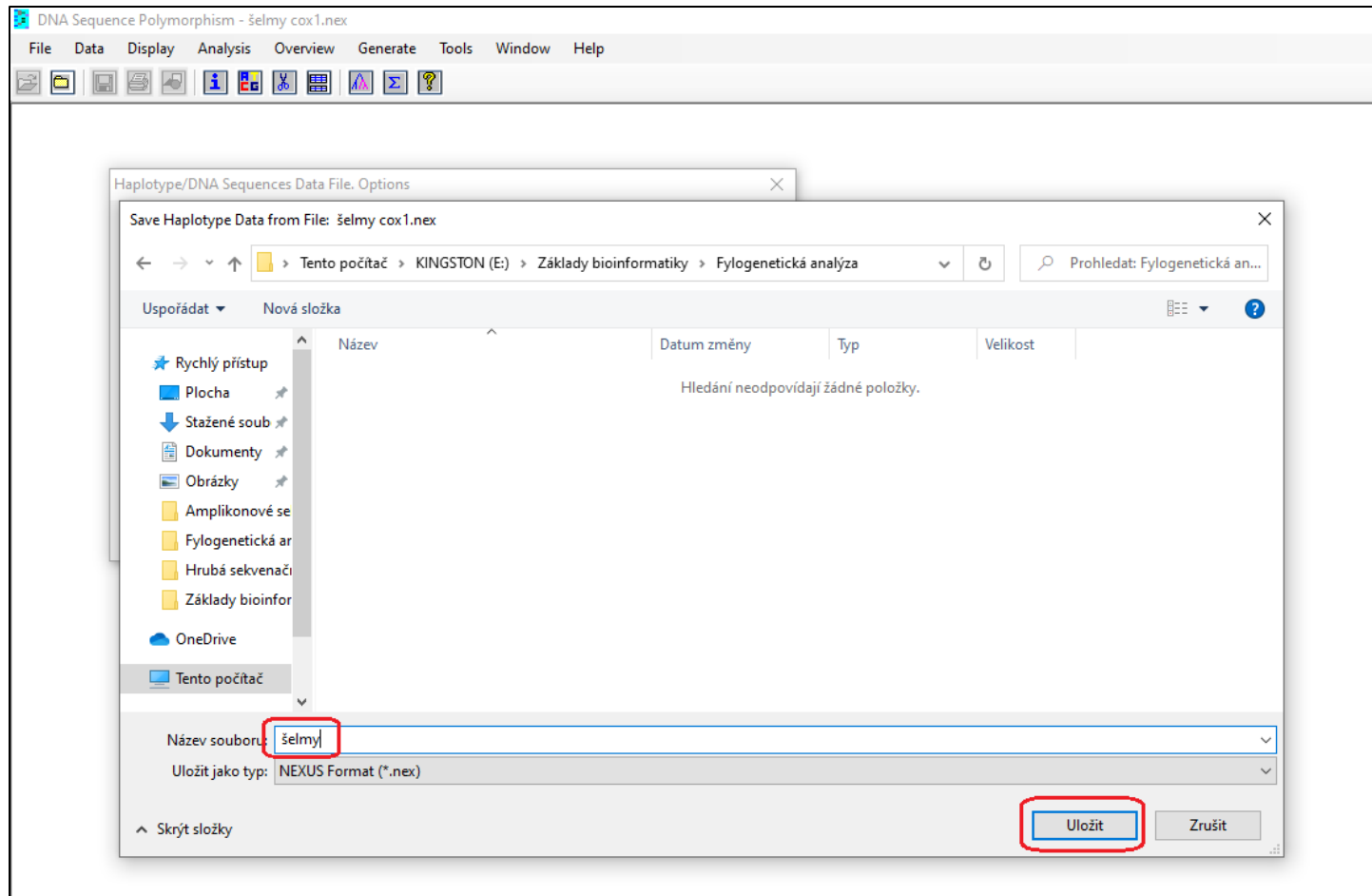
10. Vstupními daty pro haplotypovou analýzu je soubor se sekvencemi, u kterých byl proveden alignment pomocí programu MEGA X a algoritmu Muscle s využitím kodónů. Tento soubor jste vytvořili a uložili ve formátu FASTA v přechodících krocích analýzy. Otevřete tento program v prostředí programu DNA Sequence Polymorphism v6.12.03. Po načtení dat vám program potvrdí, že bylo načteno všech 36 sekvencí. Pro identifikaci haplotypů zvolte nabídku Generate – Haplotype Data File



11. V dalším kroku haplotypové analýzy program podá informaci o parametrech vlastního hodnocení. Ponechejte všechny implicitně nastavené hodnoty. Program podává informaci o počtu hodnocených sekvencí (36), o hodnocené oblasti (1. – 642. nukleotid). Nepolymorfni oblasti nejsou hodnoceny. Oblasti s výskytem gapů nejsou rovněž hodnoceny. Analýza bude pokračovat po zvolení nabídky OK.



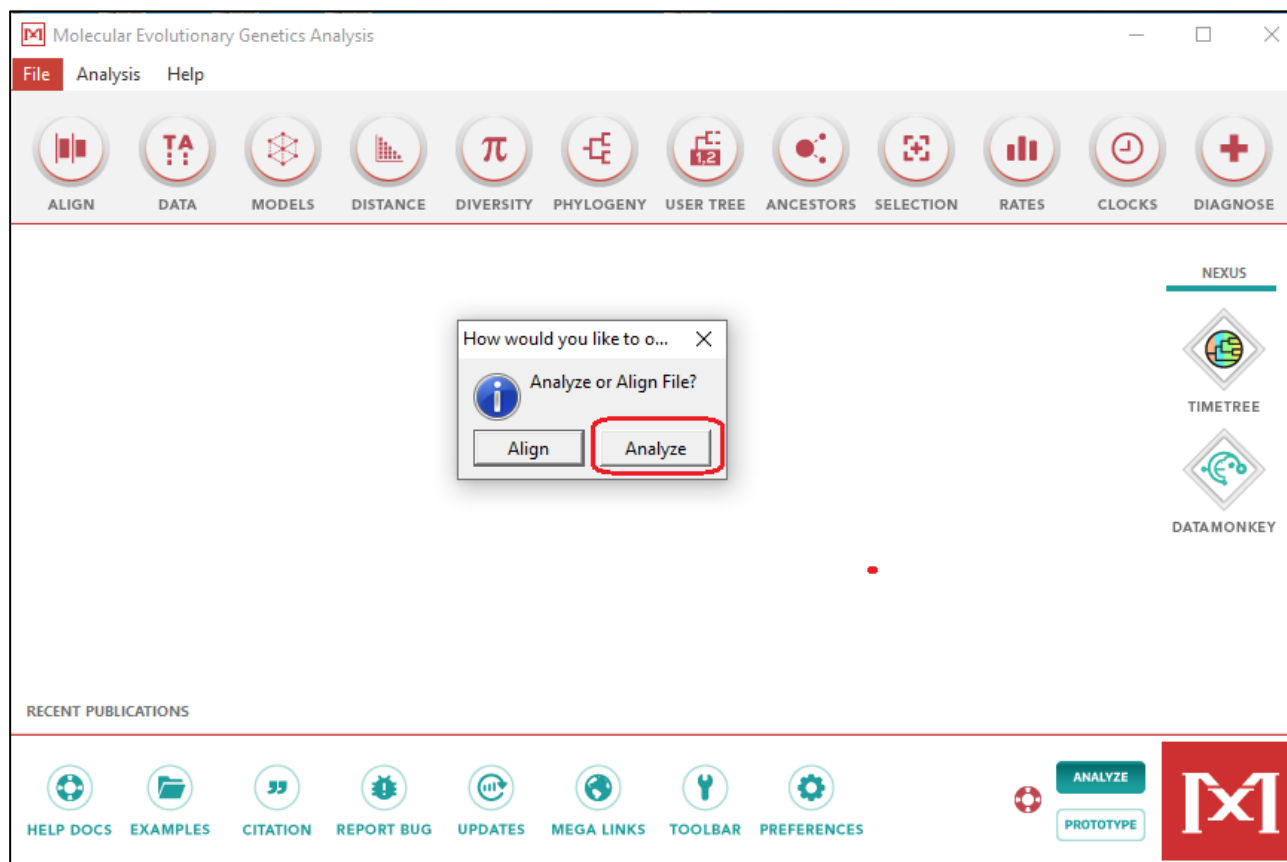
12. Výsledky haplotypové analýzy uložte do vašeho počítače ve formátu NEXUS.



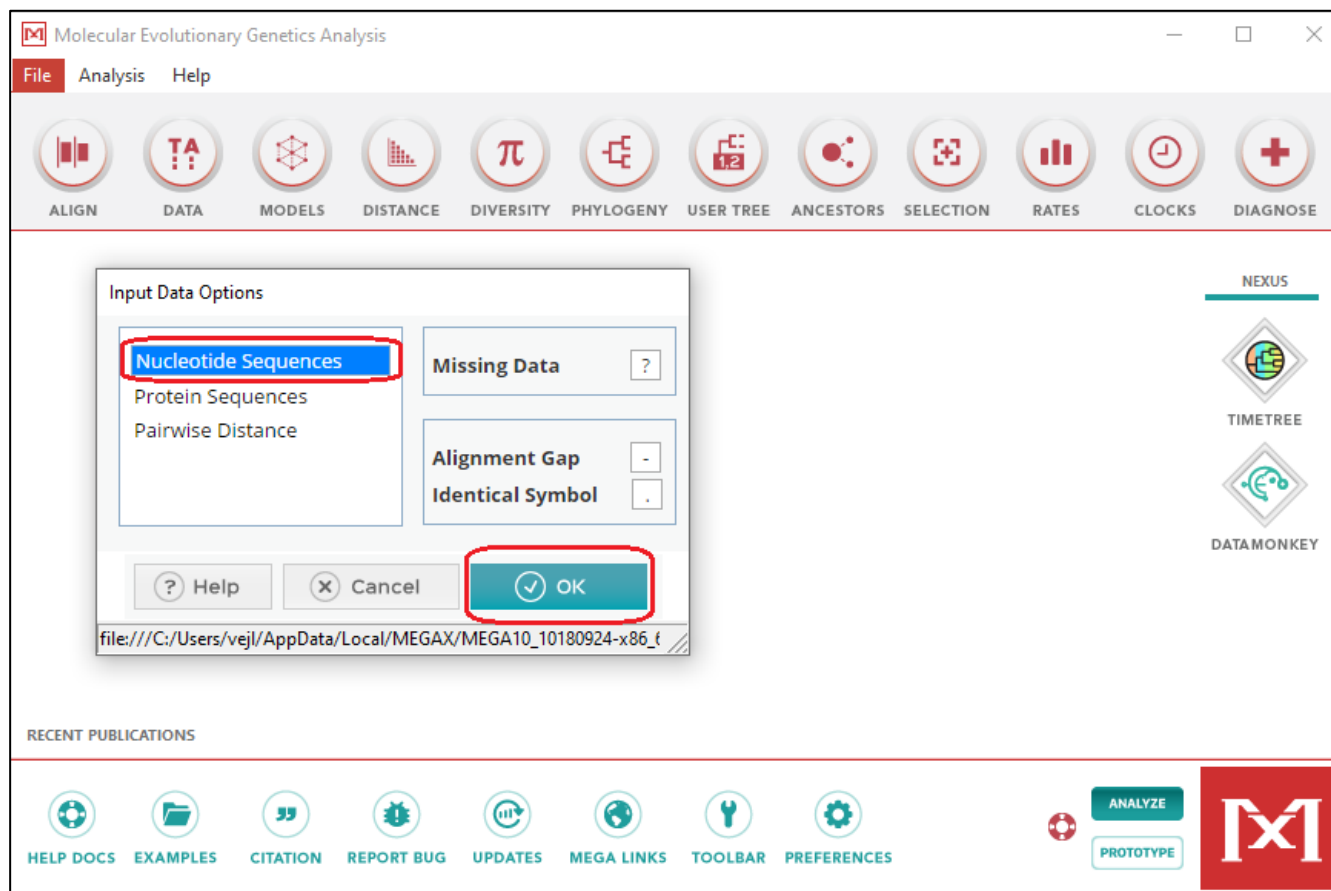
13. V tuto chvíli je haplotypová analýza kompletní. Byl detekovaný shodný počet haplotypů (36) jako je počet hodnocených sekvencí. Můžeme tudíž tvrdit, že všechny haplotypy, které budou použity pro fylogenetickou studii, se vzájemně sekvenčně odlišují. Haplotypová diverzita H_D je tudíž rovna 1,00.

```
DNA Sequence Polymorphism - šelmy cox1.nex
File Data Display Analysis Overview Generate Tools Window Help
Output of: šelmy cox1.nex
Haplotype/DNA Sequences Data File
Input Data File: C:\...\šelmy cox1.nex
Number of sequences: 36 Number of sequences used: 36
Selected region: 1-642 Number of sites: 642
Total number of sites (excluding sites with gaps / missing data): 642
Sites with alignment gaps: not considered
Number of variable sites: 642
===== Haplotype Distribution =====
Number of haplotypes, h: 36
Haplotype diversity, Hd: 1,0000
Hap_1: 1 [1]
Hap_2: 1 [2]
Hap_3: 1 [3]
Hap_4: 1 [4]
Hap_5: 1 [5]
Hap_6: 1 [6]
Hap_7: 1 [7]
Hap_8: 1 [8]
Hap_9: 1 [9]
Hap_10: 1 [10]
Hap_11: 1 [11]
Hap_12: 1 [12]
Hap_13: 1 [13]
Hap_14: 1 [14]
Hap_15: 1 [15]
Hap_16: 1 [16]
Hap_17: 1 [17]
Hap_18: 1 [18]
Hap_19: 1 [19]
Hap_20: 1 [20]
Hap_21: 1 [21]
Hap_22: 1 [22]
Hap_23: 1 [23]
Hap_24: 1 [24]
Hap_25: 1 [25]
Hap_26: 1 [26]
Hap_27: 1 [27]
Hap_28: 1 [28]
Hap_29: 1 [29]
Hap_30: 1 [30]
Hap_31: 1 [31]
Hap_32: 1 [32]
Hap_33: 1 [33]
Hap_34: 1 [34]
Hap_35: 1 [35]
Hap_36: 1 [36]
Hap_1: 1 [Hap_1]
Hap_2: 1 [Hap_2]
Hap_3: 1 [Hap_3]
Hap_4: 1 [Hap_4]
Hap_5: 1 [Hap_5]
Hap_6: 1 [Hap_6]
Hap_7: 1 [Hap_7]
Hap_8: 1 [Hap_8]
Hap_9: 1 [Hap_9]
Hap_10: 1 [Hap_10]
Hap_11: 1 [Hap_11]
Hap_12: 1 [Hap_12]
Hap_13: 1 [Hap_13]
Hap_14: 1 [Hap_14]
Hap_15: 1 [Hap_15]
Hap_16: 1 [Hap_16]
```

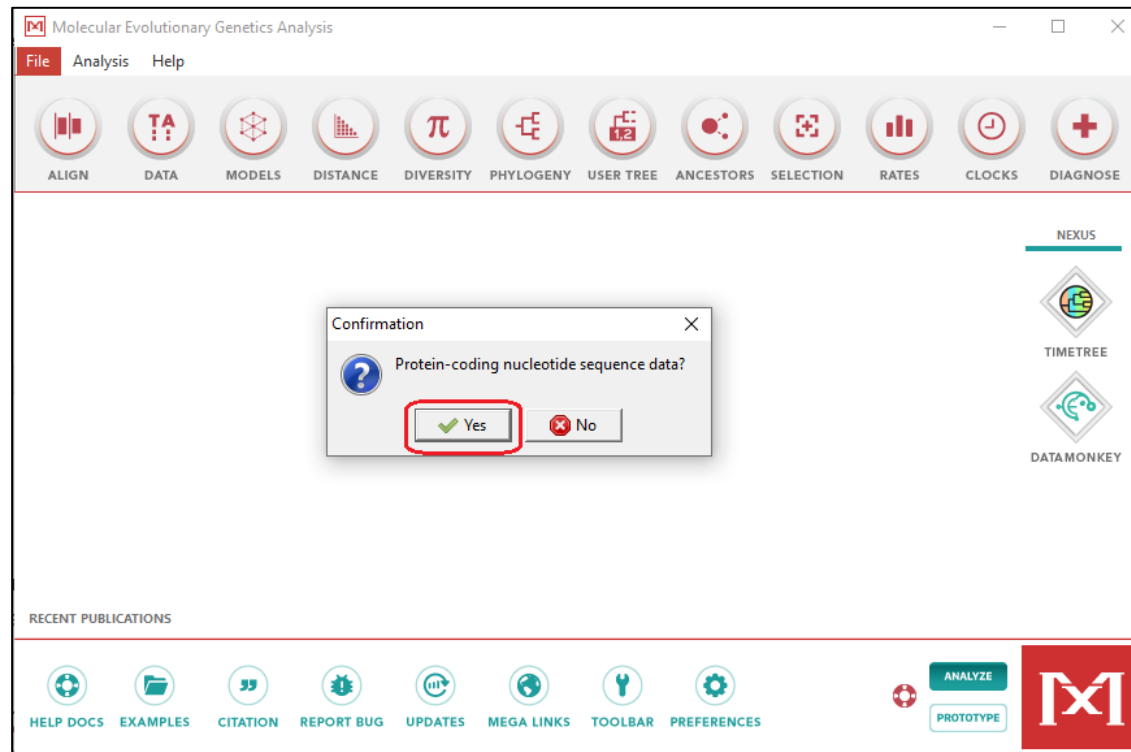
14. Nyní opusťte program DNA Sequence Polymorphism v6.12.03. a vraťte se zpět do programu MEGA X, ve kterém budete pokračovat ve fylogenetické analýze. Existuje velké množství algoritmů a strategií pro fylogenetické analýzy a konstrukce fylogenetických stromů. Jednotlivé metody se liší nejen svou sofistikovaností, spolehlivostí a věrohodností výsledků, ale rovněž i náročností na hardwarové vybavení nebo čas, po který analýzy probíhají. Pro kurz Základů bioinformatiky byla zvolena metoda maximální věrohodnosti (Maximum Likelihood Method). Spusťte znovu program MEGA X a otevřete vámi uložený datový soubor po alignmentu ve formátu FASTA. Po výběru souboru zvolte, že data budete analyzovat. Alignment jste již provedli v předešlých krocích analýzy.



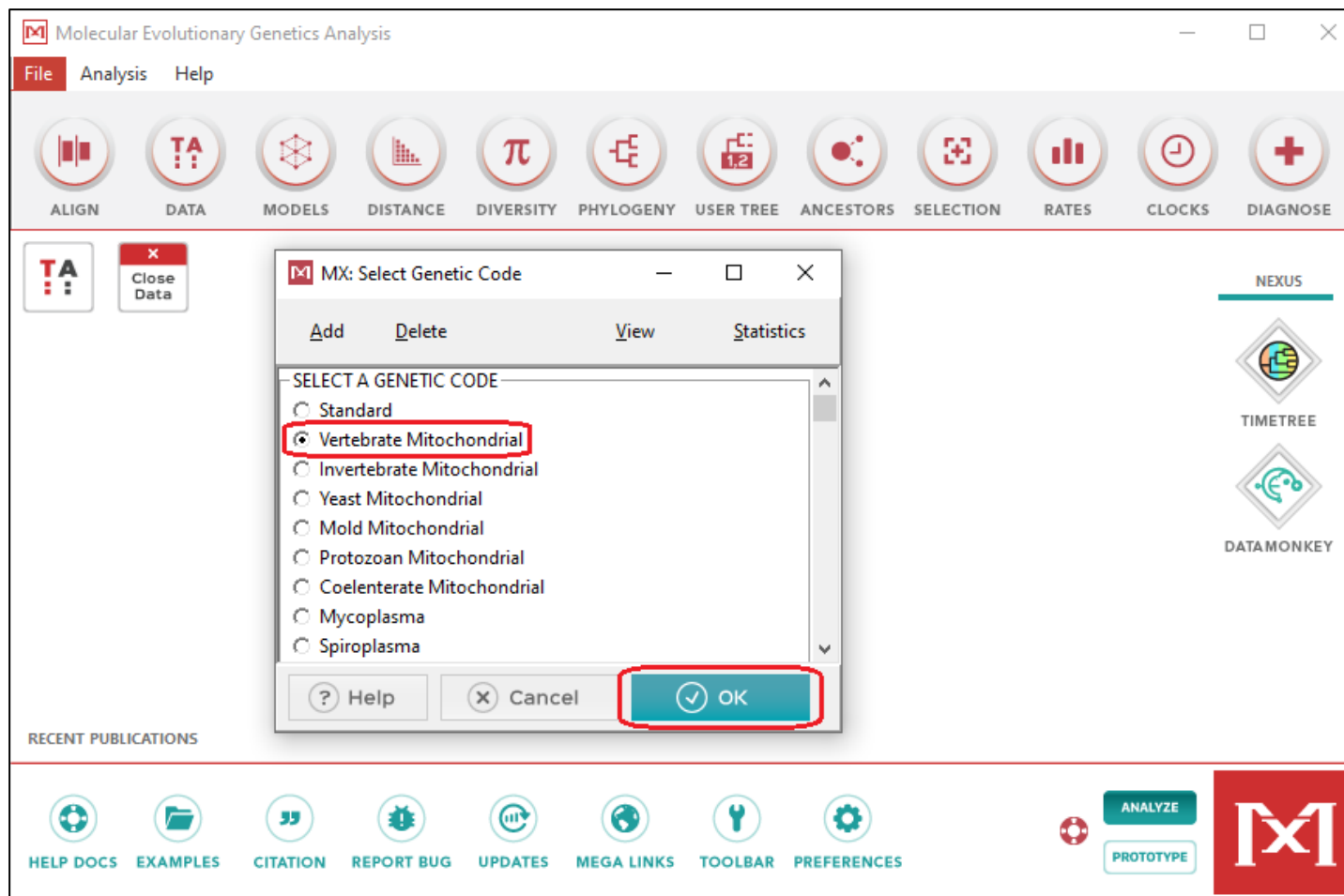
15. Program MEGA X nyní ověří, jakým způsobem je použitý pro označování jednotlivých pozic u hodnocených sekvencí. Myší označte, že se jedná o nukleotidové sekvence. U ostatních parametrů ponechejte implicitně nastavené hodnoty a zadejte OK.



16. V dalším kroku analýzy budete vyzváni k potvrzení, zdali se jedná o sekvenci DNA kódující proteiny. *Cox1* gen je strukturální gen kódující protein, a proto tuto nabídku potvrďte.



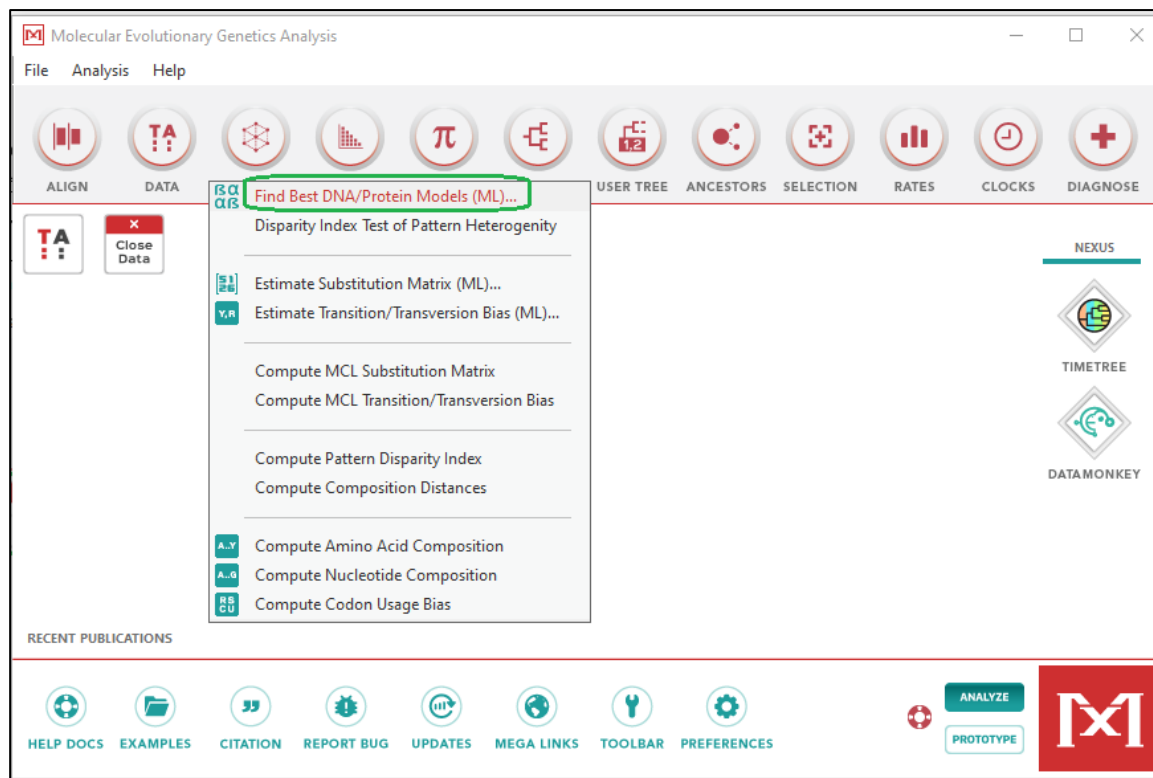
17. Následně vyberte typ genetického kódu. Analyzujete mitochondriální gen *Cox1* u šelem – obratlovců (Vertebrate Mitochondrial).



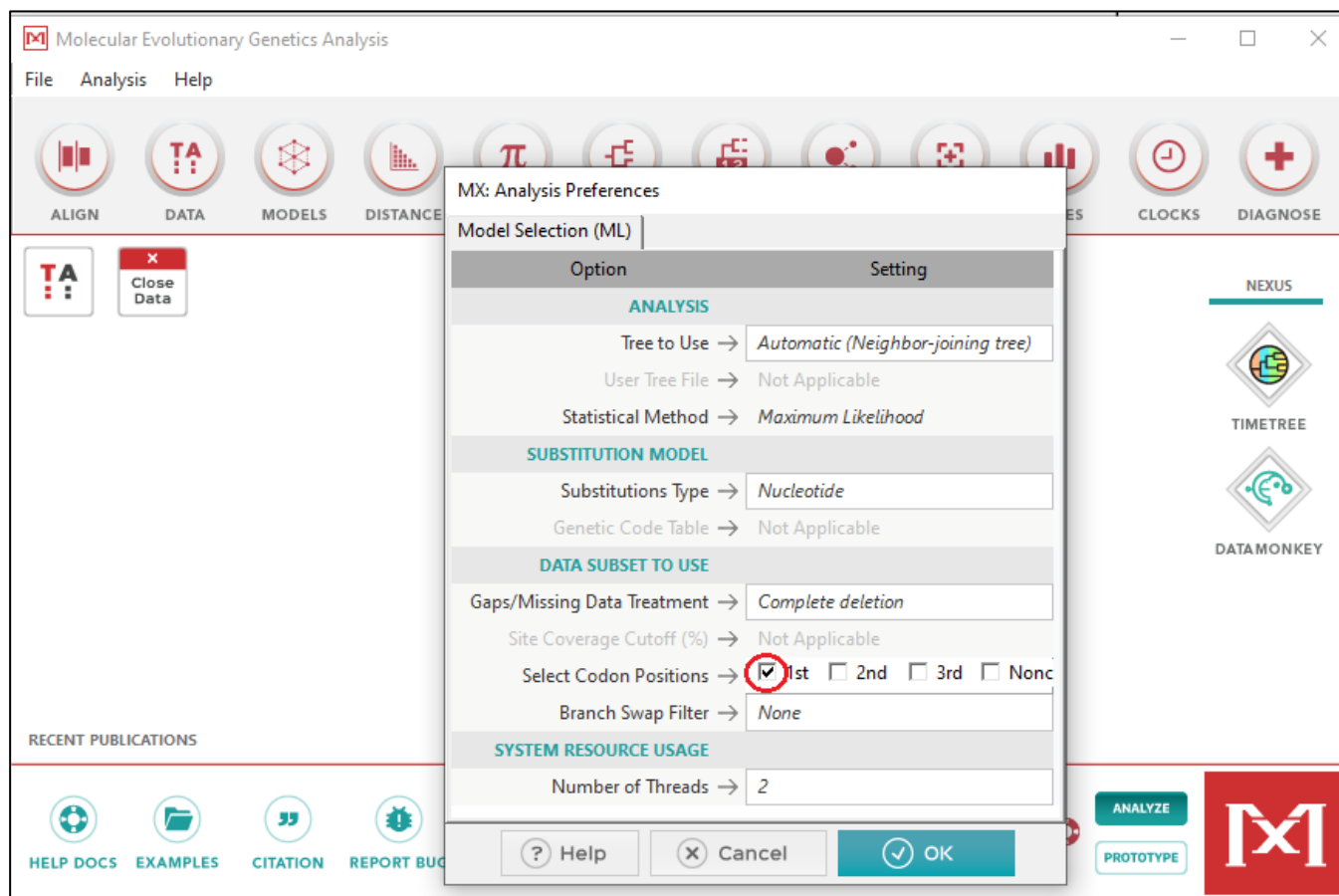
18. Nyní jsou data načtena a program MEGA X má zadané informace, jak bude k sekvencím přistupovat při fylogenetické analýze. Přítomnost dat v paměti počítači označuje zeleně označená ikona. Během analýzy nikdy neklikajte na červeně označenou ikonu, která provede vymazání všech zadaných dat a tím i ukončení všech probíhajících analýz.



19. V předchozí části postupu bylo zmíněno, že pro analýzu použijeme metodu maximální věrohodnosti (Maximum Likelihood Method). Prvním krokem je volba vhodného substitučního modelu, který popisuje, jakým způsobem probíhaly substituční mutace, které ve finále mohly diverzifikovat historického společného předka na recentní druhy hodnocených šelem. V hlavní liště zvolte nabídku MODELS a konkrétní nabídku Find Best DNA/Protein Models (ML). Pro kontrolu budete dotázáni, zdali chcete skutečně analyzovat aktuálně otevřený datový soubor.

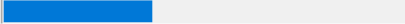


20. V dalším kroku analýzy ponechte všechny implicitně nastavené hodnoty s výjimkou Select Codon Position. Zde nechte označenou pouze hodnotu 1st, a to proto, že víte, že v hodnocených sekvencích představuje první kodon skutečně první 3 nukleotidy. Parametr Number of Threads se nastavuje automaticky podle počtu jader procesoru ve vašem počítači. Po zadání nabídky OK bude analýza pokračovat.



21. Nyní bude program testovat a hodnotit jednotlivé substituční modely (24)


MX: Progress

Progress 

Status/Options

Run Status

Start time	22.01.2021 14:14:50
Status	Making initial tree
Thread-2	HKY + I
Thread-1	TN93 + G + I

<  >

Analysis Options

```
Tree to Use           : Automatic (Neighbor-joining tree)
  Statistical Method   : Maximum Likelihood
Substitution Model
  Substitutions Type   : Nucleotide
Data Subset to Use
  Gaps/Missing Data Treatment : Complete deletion
  Select Codon Positions  : 1st,2nd,3rd,Non-Coding
  Branch Swap Filter     : None
System Resource Usage
  Number of Threads     : 2
```


22. Výsledky hodnocení modelů zobrazí program tabulkovou formou, kdy nejvěrohodnější substituční model je vybrán na základě BIC skóre. Jako nejlepší byl vybrán substituční model TN93+G. Bližší odkazy na tento model jsou uvedeny pod tabulkou.

MEGA Caption Expert: Find Best-Fit Substitution Model (ML)

File Edit View Help

Results

Table. Maximum Likelihood fits of 24 different nucleotide substitution models

Model	Parameters	BIC	AICc	lnL	(+I)	(+G)	R	r(A)	r(T)	r(C)	r(G)	r(AT)	r(AC)	r(AG)	r(TA)	r(TC)	r(TG)	r(CA)	r(CT)	r(CG)	r(GA)	r(GT)	r(GC)
TN93+G	75	5205,813	4619,501	-2234,441	n/a	0,11	14,99	0,266	0,239	0,203	0,292	0,007	0,006	0,069	0,008	0,372	0,008	0,008	0,437	0,008	0,063	0,007	0,006
TN93+G+I	76	5215,143	4621,022	-2234,193	0,09	0,12	14,81	0,266	0,239	0,203	0,292	0,007	0,006	0,072	0,008	0,369	0,009	0,008	0,434	0,009	0,066	0,007	0,006
GTR+G	78	5223,486	4613,747	-2228,539	n/a	0,11	13,52	0,266	0,239	0,203	0,292	0,012	0,006	0,069	0,014	0,370	0,013	0,008	0,435	0,000	0,063	0,011	0,000
K2+G+I	72	5229,676	4666,792	-2261,111	0,46	0,20	16,61	0,250	0,250	0,250	0,250	0,007	0,007	0,236	0,007	0,236	0,007	0,007	0,236	0,007	0,236	0,007	0,007
GTR+G+I	79	5232,649	4615,101	-2228,207	0,10	0,12	13,27	0,266	0,239	0,203	0,292	0,013	0,006	0,073	0,014	0,366	0,013	0,008	0,430	0,000	0,066	0,011	0,000
T92+G+I	73	5240,116	4669,423	-2261,418	0,46	0,20	16,47	0,252	0,252	0,248	0,248	0,007	0,007	0,233	0,007	0,233	0,007	0,007	0,238	0,007	0,238	0,007	0,007
K2+G	71	5242,694	4687,621	-2272,533	n/a	0,11	16,57	0,250	0,250	0,250	0,250	0,007	0,007	0,236	0,007	0,236	0,007	0,007	0,236	0,007	0,236	0,007	0,007
TN93+I	75	5252,700	4666,388	-2257,885	0,74	n/a	13,41	0,266	0,239	0,203	0,292	0,008	0,007	0,078	0,009	0,361	0,009	0,009	0,424	0,009	0,072	0,008	0,007
T92+G	72	5252,916	4690,033	-2272,731	n/a	0,11	16,46	0,252	0,252	0,248	0,248	0,007	0,007	0,233	0,007	0,233	0,007	0,007	0,238	0,007	0,238	0,007	0,007
GTR+I	78	5269,655	4659,916	-2251,623	0,74	n/a	11,45	0,266	0,239	0,203	0,292	0,016	0,005	0,078	0,018	0,357	0,015	0,007	0,420	0,000	0,071	0,012	0,000
HKY+G+I	75	5279,423	4693,111	-2271,246	0,47	0,20	17,86	0,266	0,239	0,203	0,292	0,006	0,006	0,276	0,007	0,192	0,008	0,007	0,226	0,008	0,252	0,006	0,006
HKY+G	74	5294,997	4716,494	-2283,946	n/a	0,11	17,54	0,266	0,239	0,203	0,292	0,007	0,006	0,276	0,007	0,192	0,008	0,007	0,226	0,008	0,251	0,007	0,006
K2+I	71	5311,368	4756,295	-2306,870	0,78	n/a	12,87	0,250	0,250	0,250	0,250	0,009	0,009	0,232	0,009	0,232	0,009	0,009	0,232	0,009	0,232	0,009	0,009
T92+I	72	5320,801	4757,918	-2306,674	0,78	n/a	12,81	0,252	0,252	0,248	0,248	0,009	0,009	0,230	0,009	0,230	0,009	0,009	0,234	0,009	0,234	0,009	0,009
HKY+I	74	5374,023	4795,520	-2323,459	0,78	n/a	12,97	0,266	0,239	0,203	0,292	0,009	0,007	0,270	0,010	0,188	0,011	0,010	0,221	0,011	0,246	0,009	0,007
TN93	74	5750,357	5171,854	-2511,626	n/a	n/a	10,50	0,266	0,239	0,203	0,292	0,010	0,008	0,080	0,011	0,352	0,012	0,011	0,413	0,012	0,073	0,010	0,008
GTR	77	5782,689	5180,759	-2513,053	n/a	n/a	7,43	0,266	0,239	0,203	0,292	0,012	0,014	0,081	0,013	0,337	0,016	0,018	0,396	0,016	0,074	0,013	0,011
JC+G+I	71	5783,341	5228,268	-2542,857	0,43	0,23	0,50	0,250	0,250	0,250	0,250	0,083	0,083	0,083	0,083	0,083	0,083	0,083	0,083	0,083	0,083	0,083	0,083
JC+G	70	5790,805	5243,542	-2551,501	n/a	0,12	0,50	0,250	0,250	0,250	0,250	0,083	0,083	0,083	0,083	0,083	0,083	0,083	0,083	0,083	0,083	0,083	0,083
JC+I	70	5838,838	5291,576	-2575,518	0,77	n/a	0,50	0,250	0,250	0,250	0,250	0,083	0,083	0,083	0,083	0,083	0,083	0,083	0,083	0,083	0,083	0,083	0,083
K2	70	5900,870	5353,608	-2606,534	n/a	n/a	10,33	0,250	0,250	0,250	0,250	0,011	0,011	0,228	0,011	0,228	0,011	0,011	0,228	0,011	0,228	0,011	0,011
T92	71	5908,788	5353,715	-2605,580	n/a	n/a	10,32	0,252	0,252	0,248	0,248	0,011	0,011	0,226	0,011	0,226	0,011	0,011	0,230	0,011	0,230	0,011	0,011
HKY	73	5959,334	5388,641	-2621,028	n/a	n/a	10,28	0,266	0,239	0,203	0,292	0,011	0,009	0,265	0,012	0,185	0,013	0,012	0,217	0,013	0,242	0,011	0,009
JC	69	6393,253	5853,801	-2857,639	n/a	n/a	0,50	0,250	0,250	0,250	0,250	0,083	0,083	0,083	0,083	0,083	0,083	0,083	0,083	0,083	0,083	0,083	0,083

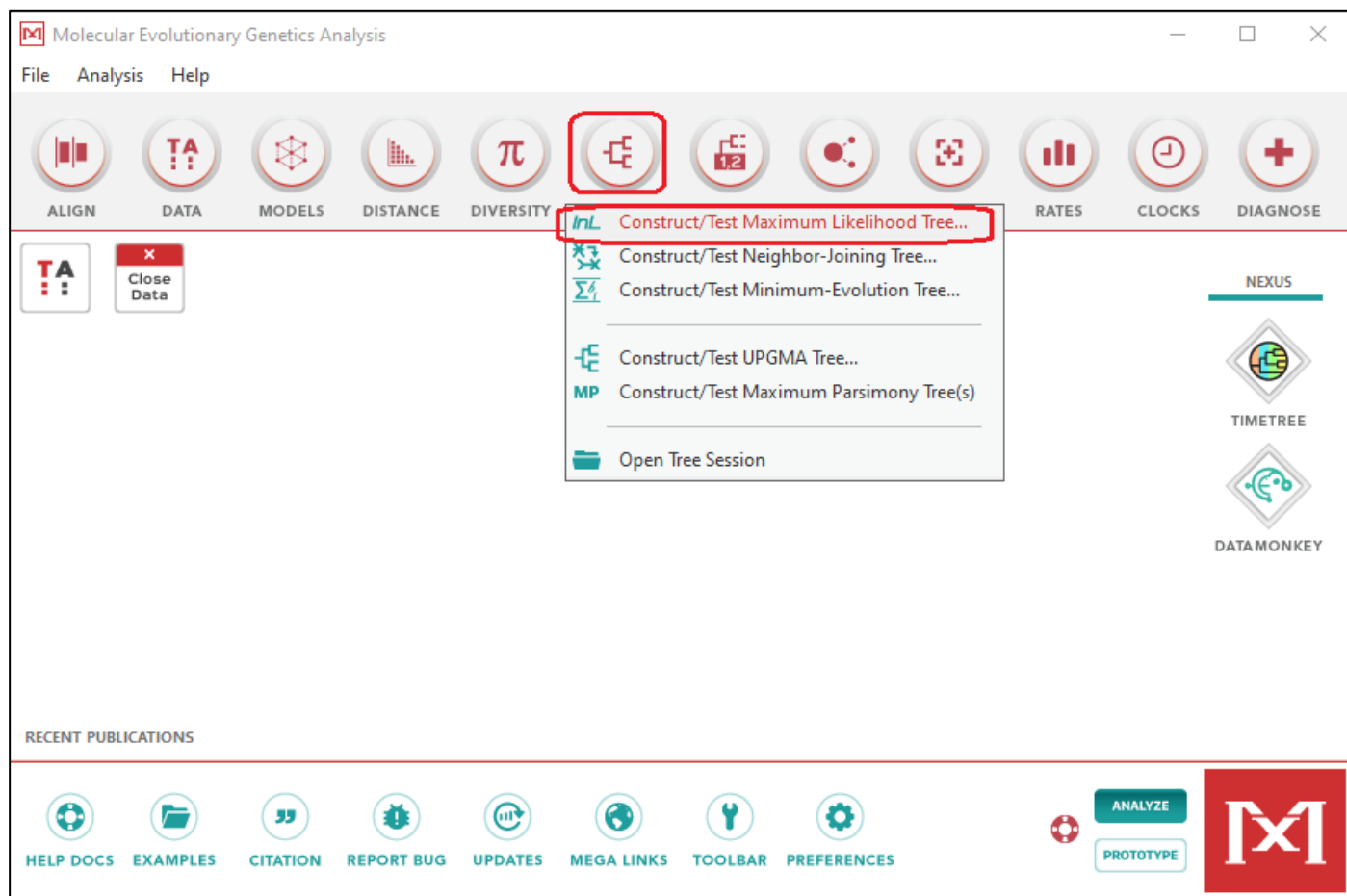
NOTE.-- Models with the lowest BIC scores (Bayesian Information Criterion) are considered to describe the substitution pattern the best. For each model, AICc value (Akaike Information Criterion, corrected), Maximum Likelihood value (lnL), and the number of parameters (including branch lengths) are also presented [1]. Non-uniformity of evolutionary rates among sites may be modeled by using a discrete Gamma distribution (+G) with 5 rate categories and by assuming that a certain fraction of sites are evolutionarily invariable (+I). Whenever applicable, estimates of gamma shape parameter and/or the estimated fraction of invariant sites are shown. Assumed or estimated values of transition/transversion bias (R) are shown for each model, as well. They are followed by nucleotide frequencies (r) for each nucleotide pair. Relative values of instantaneous r should be considered when evaluating them. For simplicity, sum of r values is made equal to 1 for each model. For estimating ML values, a tree topology was automatically computed. This analysis involved 36 nucleotide sequences. Codon positions included were 1st. All positions containing gaps and missing data were eliminated (complete deletion option). There were a total of 514 positions in the final dataset. Evolutionary analyses were conducted in MEGA X [2].

Abbreviations: TR: General Time Reversible; HKY: Hasegawa-Kishino-Yano; TN93: Tamura-Nei; T92: Tamura 3-parameter; K2: Kimura 2-parameter; JC: Jukes-Cantor./div>

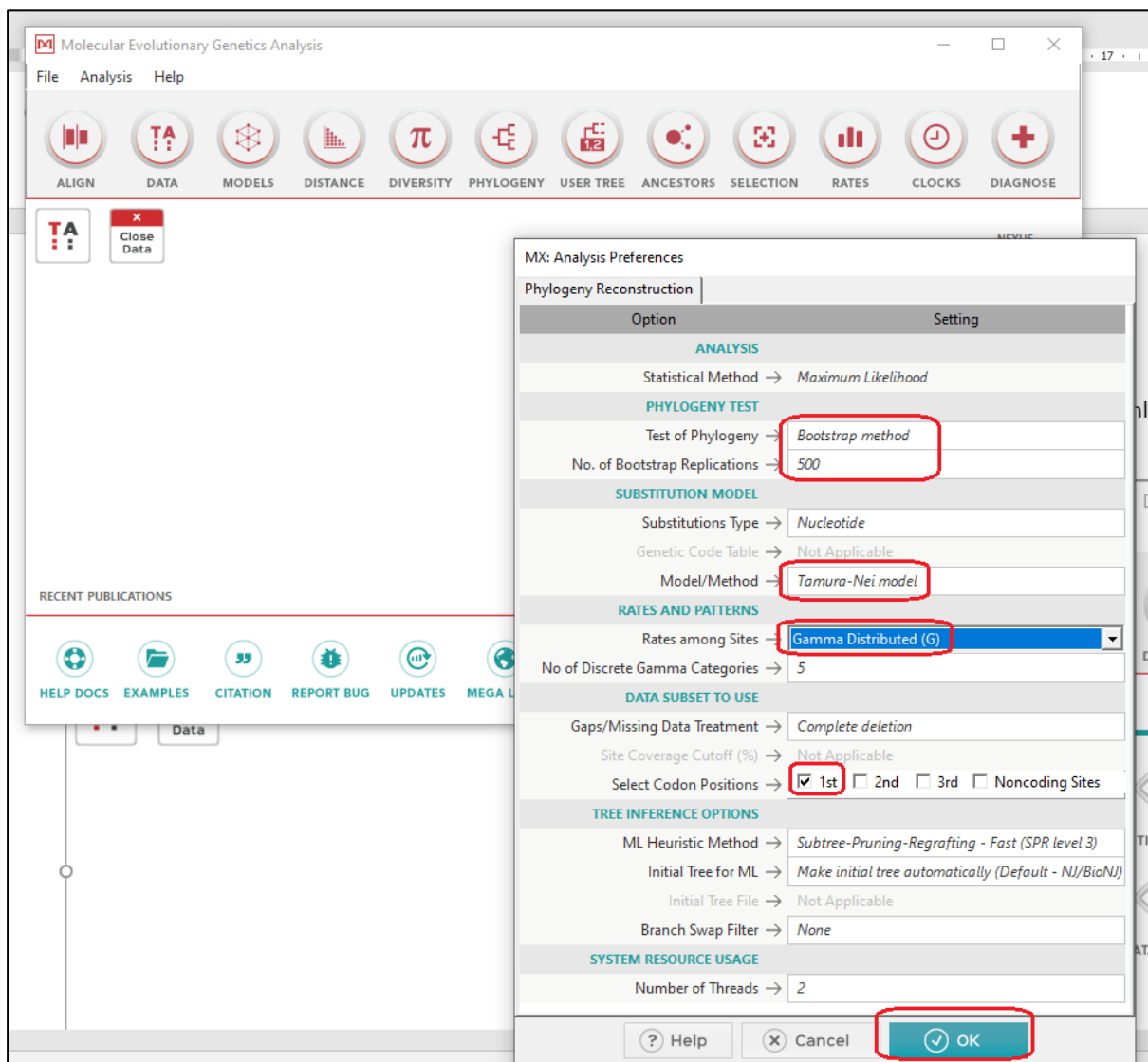
1. Nei M. and Kumar S. (2000). *Molecular Evolution and Phylogenetics*. Oxford University Press, New York.

2. Kumar S., Stecher G., Li M., Niyaz C., and Tamura K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Molecular Biology and Evolution* 35:1547-1549.

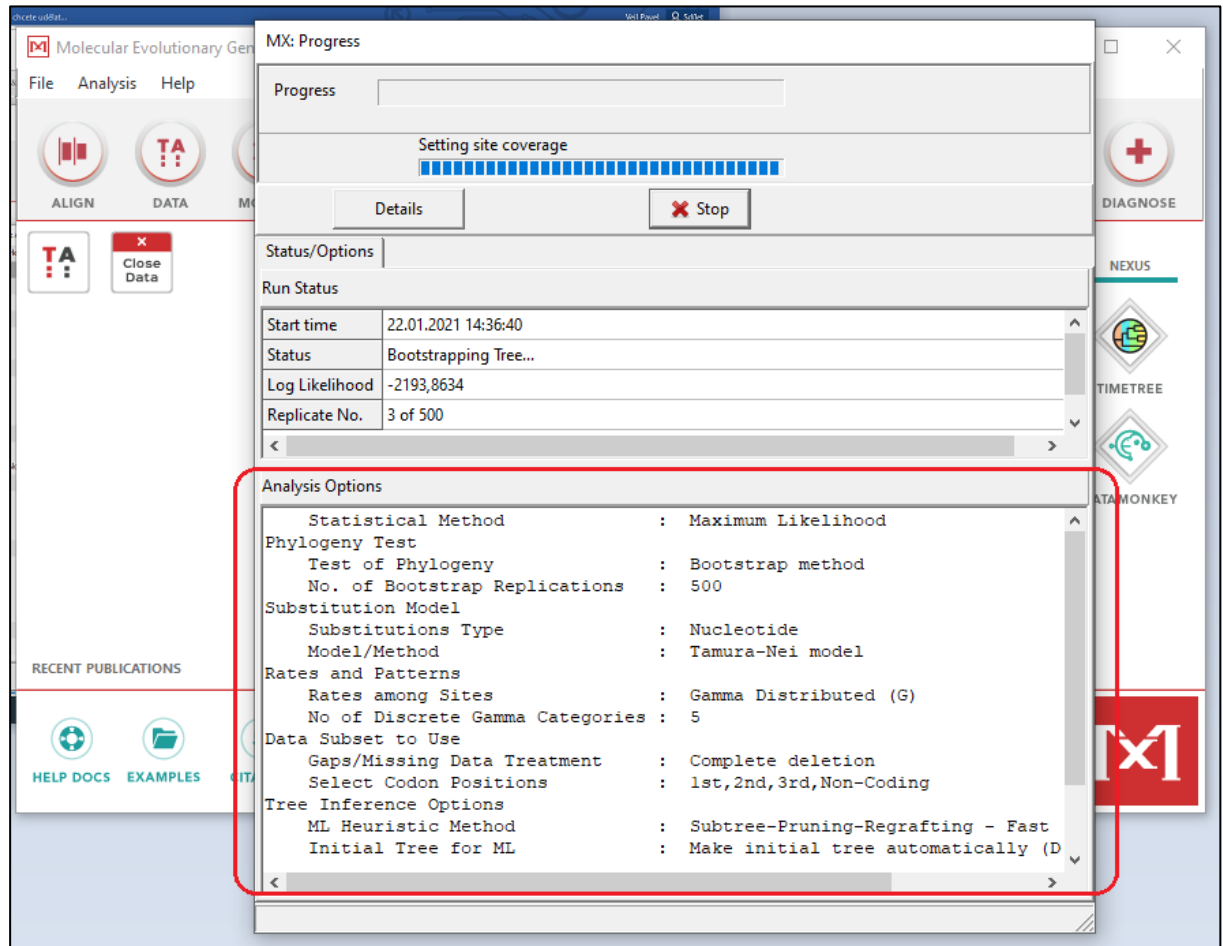
23. Po nalezení optimálního substitučního modelu přistupte ke konstrukci fylogenetického stromu. Na hlavní liště zvolte funkci PHYLOGENY a v další nabídce zvolte vybranou metodu Construct/Test Maximum Likelihood Tree.



24. V tuto chvíli jste dospěli k poslední fázi nastavení parametrů vlastní fylogenetické analýzy. Program MEGA X implicitně nastavuje test stromu, respektive test zařazení jednotlivých druhů do větví pomocí tak zvané Bootstrap metody, která se implicitně opakuje 500x. Zde doporučuji ponechat implicitní nastavení 500x. Je potřeba si uvědomit, že právě nastavení hodnoty opakování výrazně zpomaluje celou analýzu a v případě nedostatečného hardwarového vybavení bude celý chod počítače výrazně zpomalen a analýza může trvat až několik hodin. Pokud nastanou tyto problémy, doporučuji při opakovaném spuštění analýzy hodnotu opakování snížit například na 200.
- V červeně označených polích je provedeno správné nastavení substitučního modelu TN93+G. Tyto parametry je nutné manuálně nastavit. U parametru Select Codon Position zkontrolujte, zda je správně nastavený 1. kodón (1st). Vlastní konstrukci stromu zahájíte stisknutím nabídky OK.



25. V tuto chvíli bude program zpracovávat data a konstruovat větve fylogenetického stromu tak, aby zařazení druhů do jednotlivých větví bylo provedeno na základě maximální věrohodnosti. Analýza bude probíhat v závislosti na hardwarovém vybavení několik desítek minut až hodin. Během analýzy budou na monitoru zobrazeny parametry, které byly pro tvorbu stromu nastaveny.



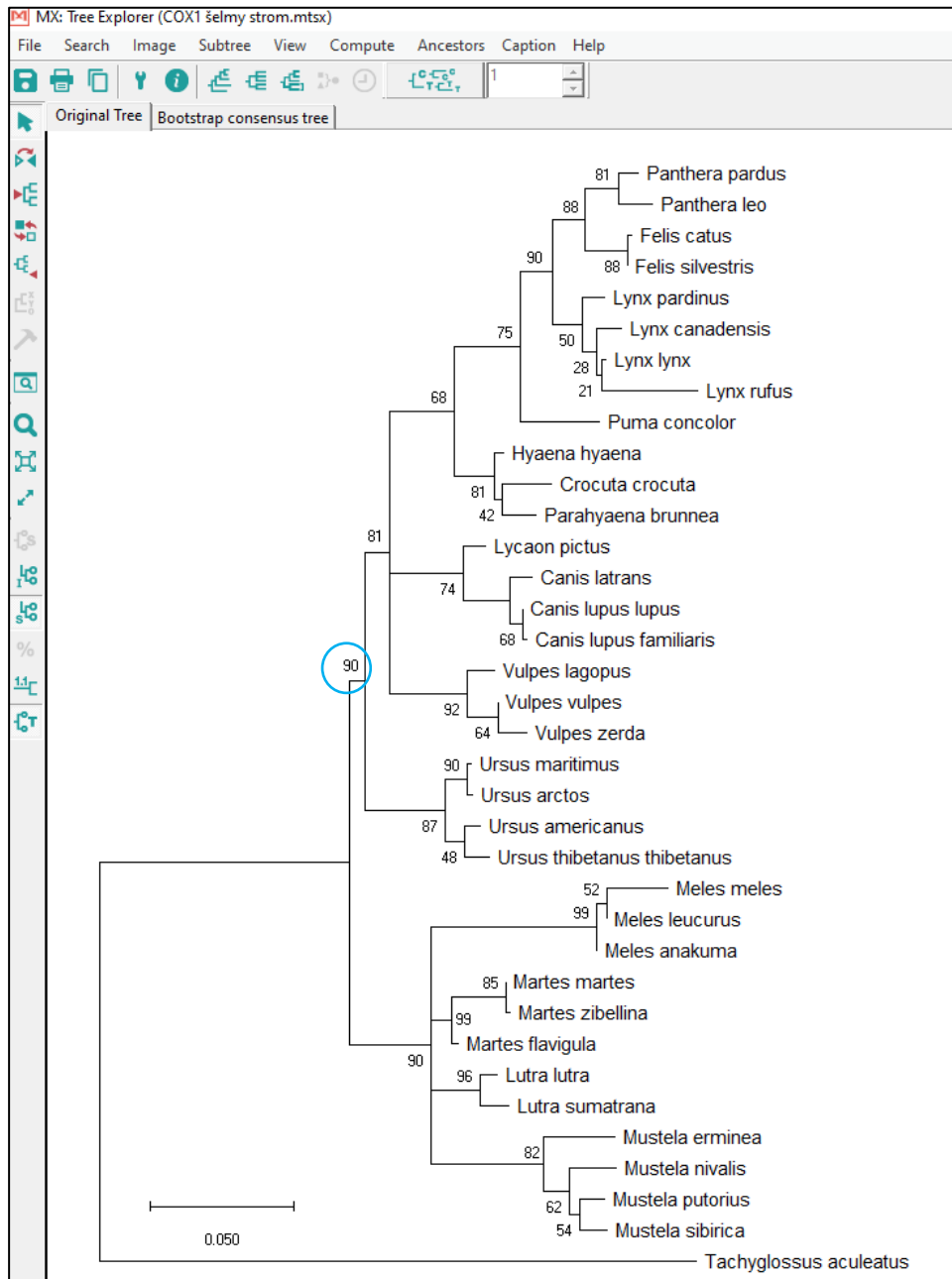
Molecular Evolutionary Genomics (MEGA) software interface showing the 'MX: Progress' dialog box. The dialog displays a progress bar for 'Setting site coverage' and a 'Details' section with 'Status/Options' and 'Run Status' tabs. The 'Run Status' tab shows:

Start time	22.01.2021 14:36:40
Status	Bootstrapping Tree...
Log Likelihood	-2193,8634
Replicate No.	3 of 500


The 'Analysis Options' tab is highlighted with a red box and lists:

Statistical Method	: Maximum Likelihood
Phylogeny Test	
Test of Phylogeny	: Bootstrap method
No. of Bootstrap Replications	: 500
Substitution Model	
Substitutions Type	: Nucleotide
Model/Method	: Tamura-Nei model
Rates and Patterns	
Rates among Sites	: Gamma Distributed (G)
No of Discrete Gamma Categories	: 5
Data Subset to Use	
Gaps/Missing Data Treatment	: Complete deletion
Select Codon Positions	: 1st, 2nd, 3rd, Non-Coding
Tree Inference Options	
ML Heuristic Method	: Subtree-Pruning-Regrafting - Fast
Initial Tree for ML	: Make initial tree automatically (D)

26. Po skončení analýzy je výsledný strom zobrazen pomocí podprogramu Tree Explorer, který umožní mimo jiné grafické editace stromu. Hodnoty bootstrap replikátů jsou uvedeny u kořenů větví (kládů) a na následujícím obrázku jsou označeny modrým kruhem. V prostředí programu Tree Explorer je možné provést například takzvané zakořenění stromu na základě přítomnosti separované větve s outgroup druhu (*Tachyglossus aculeatus*).



27. Výstupem programu MEGA X je rovněž kompletní popis fylogenetické analýzy včetně odkazů na primární literární zdroje, kde jsou popsány algoritmy, na základě kterých byl vybírán optimální substituční model a konstruován fylogenetický strom.



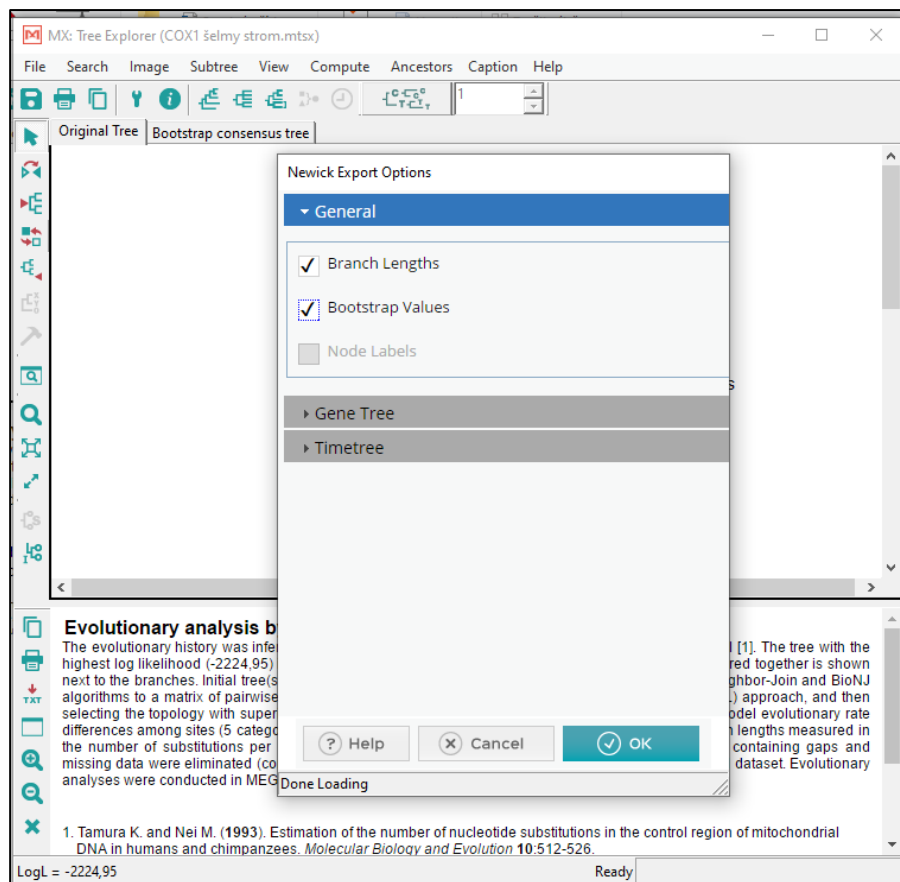
Evolutionary analysis by Maximum Likelihood method

The evolutionary history was inferred by using the Maximum Likelihood method and Tamura-Nei model [1]. The tree with the highest log likelihood (-2224,95) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 0,1072)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. This analysis involved 36 nucleotide sequences. All positions containing gaps and missing data were eliminated (complete deletion option). There were a total of 514 positions in the final dataset. Evolutionary analyses were conducted in MEGA X [2].

1. Tamura K. and Nei M. (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution* 10:512-526.
2. Kumar S., Stecher G., Li M., Niyaz C., and Tamura K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Molecular Biology and Evolution* 35:1547-1549.

Disclaimer: Although utmost care has been taken to ensure the correctness of the caption, the caption text is provided "as is" without any warranty of any kind. Authors advise the user to carefully check the caption prior to its use for any purpose and report any errors or problems to the authors immediately (www.megasoftware.net). In no event shall the authors and their employers be liable for any damages, including but not limited to special, consequential, or other damages. Authors specifically disclaim all other warranties expressed or implied, including but not limited to the determination of suitability of this caption text for a specific purpose, use, or application.

28. Vytvořený strom je možné rovněž exportovat v NEWICK formátu. Jedná se standardní formát, ve kterém jsou uložena kompletní data potřebná k otevření stromu a jeho další editace či analýzy. S formátem NEWICK pracuje většina programů zaměřených na fylogenetické analýzy. Pomocí tohoto formátu jsou exportovány konkrétní hodnoty analyzovaného stromu, jako jsou například délky větví (Branch Legths) nebo hodnoty bootstrapů (Bootstrap Values).



29. Po ukončení analýz je možné veškerá vstupní i výstupní data uložit formou projektu (session) ve formátu MEGA. Uložený projekt je možné opětovně otevřít a upravovat analýzy či provádět analýzy nové.

