

PODSTAWY BIOINFORMATYKI

WYKŁAD 6

ANALIZA FILOGENETYCZNA

- 1. Wstęp - filogenetyka**
- 2. Struktura drzewa filogenetycznego**
- 3. Metody konstrukcji drzewa**
- 4. Etapy konstrukcji drzewa filogenetycznego**
- 5. Oprogramowanie - przykłady**

OKREŚLENIE POWIĄZAŃ POMIĘDZY

1. gatunkami
2. populacjami
3. osobnikami
4. genami
5. sekwencjami powtarzalnymi
6. strukturami 2 rzędowymi białek
7. ścieżkami metabolicznymi

INFORMACJE

- Sekwencje aminokwasów, nukleotydów
- Całe genomy
- Genomy organelli (np. mitochondrium)
- Pojedyncze geny

OKREŚLENIE POWIĄZAŃ OPIERA SIĘ NA METODACH

- Klastrowania
- Kladystycznych

WIZUALIZACJA

- Drzewa filogenetyczne

ZASTOSOWANIE ANALIZY FILOGENETYCZNEJ

1. Ewolucjonizm

- określenie powiązań ewolucyjnych

2. Medycyna

- określenie zróżnicowania genetycznego patogenów

3. Predykcja funkcji genów

4. Bioróżnorodność

FILOGENEZA GENÓW A FILOGENEZA GATUNKÓW

- Ewolucja konkretnej sekwencji nie zawsze jest zgodna z drogą ewolucji gatunku
- Ewolucja gatunku jest łącznym efektem ewolucji wielu genów tworzących genom

PODOBIENSTWO SEKWENCJI

HOMOLOGIA

pochodzące od wspólnego przodka

ORTOLOGIA

- homologia pomiędzy gatunkami
- wynikła na drodze specjacji
- białka mają podobne domeny i strukturę 3-D
- podobna funkcja
- predykcja funkcji nowych genów

PARALOGIA

- homologia wewnątrz gatunku
- wynikła na drodze duplikacji genu
- białka mają zwykle odmienne funkcje

HOMOPLAZJA

niepochodzące bezpośrednio od wspólnego przodka

PARALELIZM

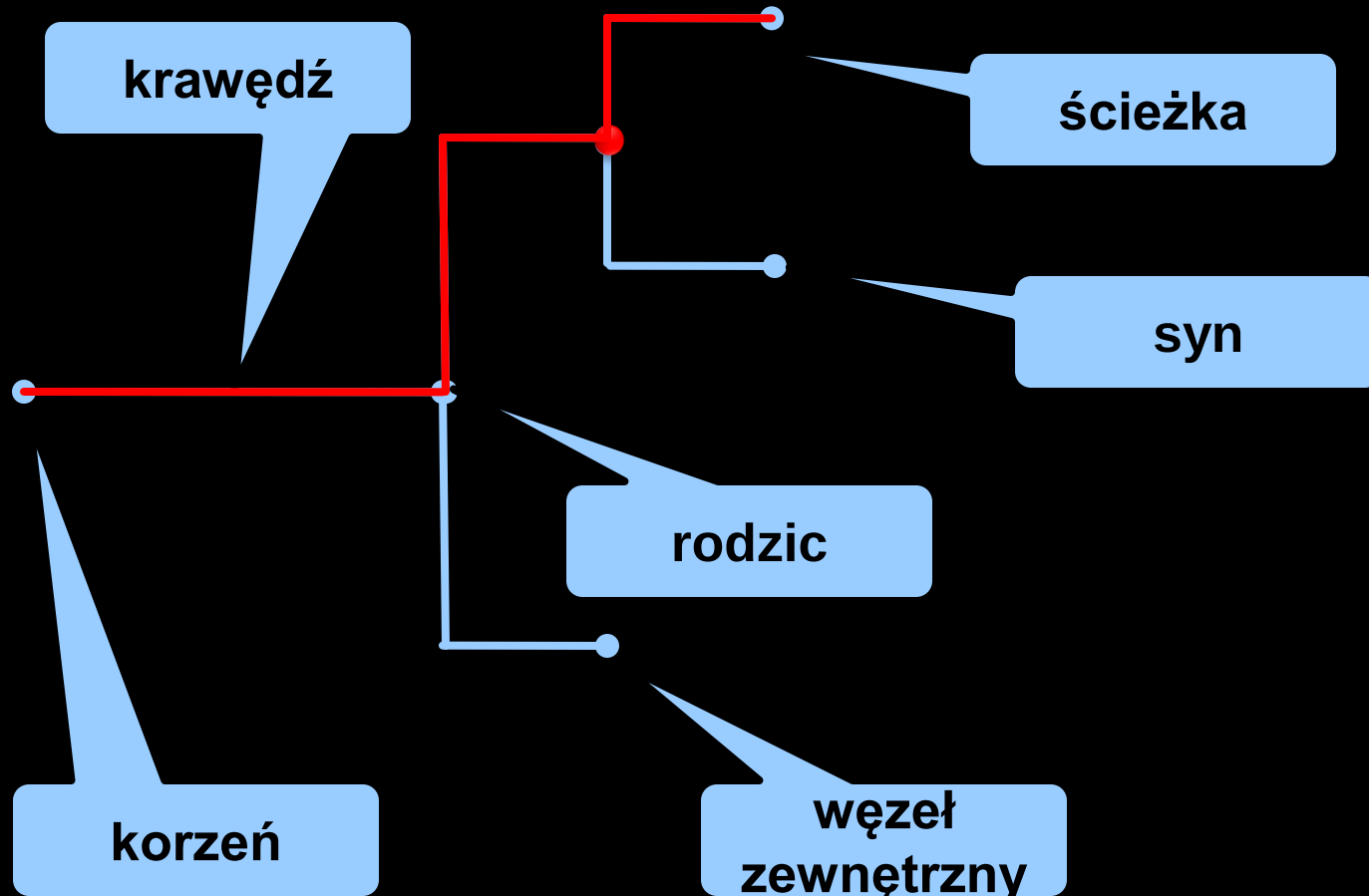
- niezależne zmiany podobnych sekwencji w tym samym kierunku

KONWERGENCJA

- niezależne zmiany różnych sekwencji w tym samym kierunku

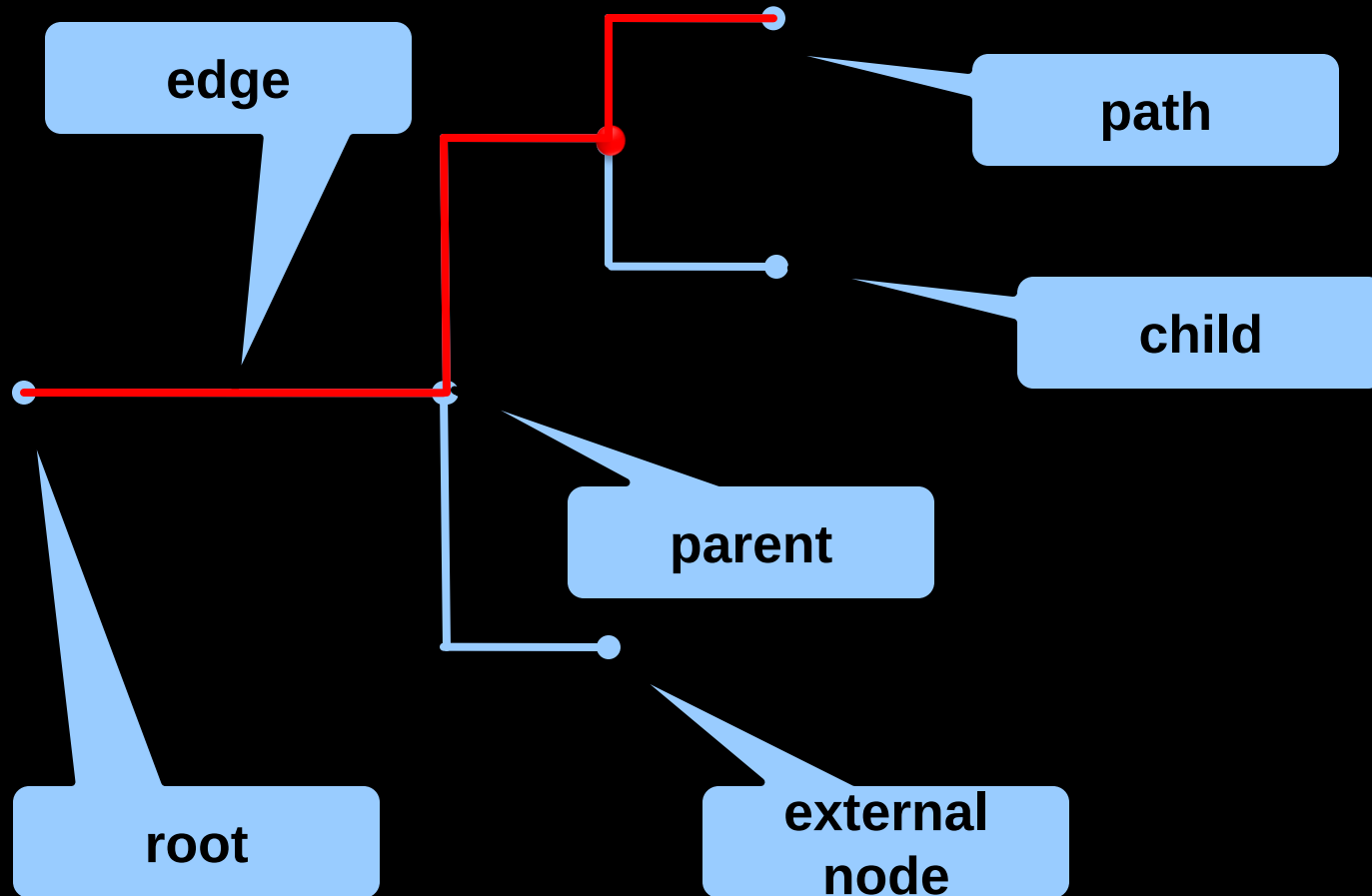
STRUKTURA DRZEWA FILOGENETYCZNEGO

TERMINOLOGIA TEORII GRAFÓW



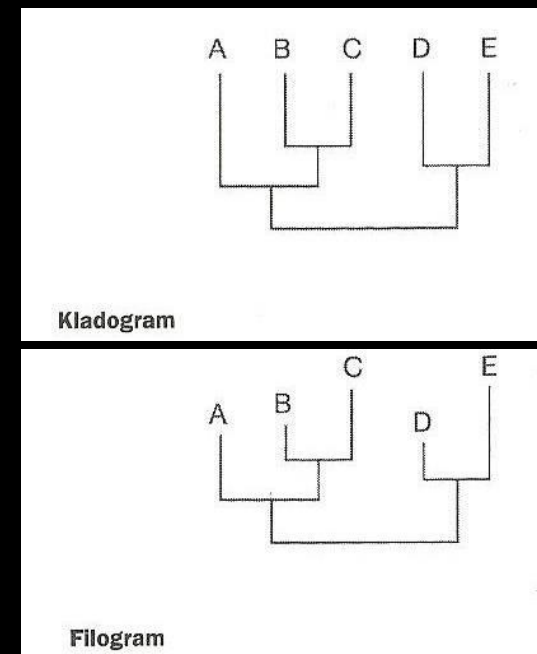
STRUKTURA DRZEWA FILOGENETYCZNEGO

GRAPH



DRZEWA FILOGENETYCZNE

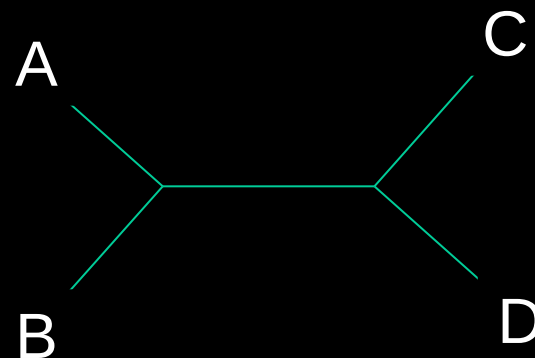
- ukorzenione i nieukorzenione
- binarność konstrukcji
- topologia
- długość gałęzi = czas ewolucji ?
 - kladogram
 - filogram



Drzewo
ukorzenione



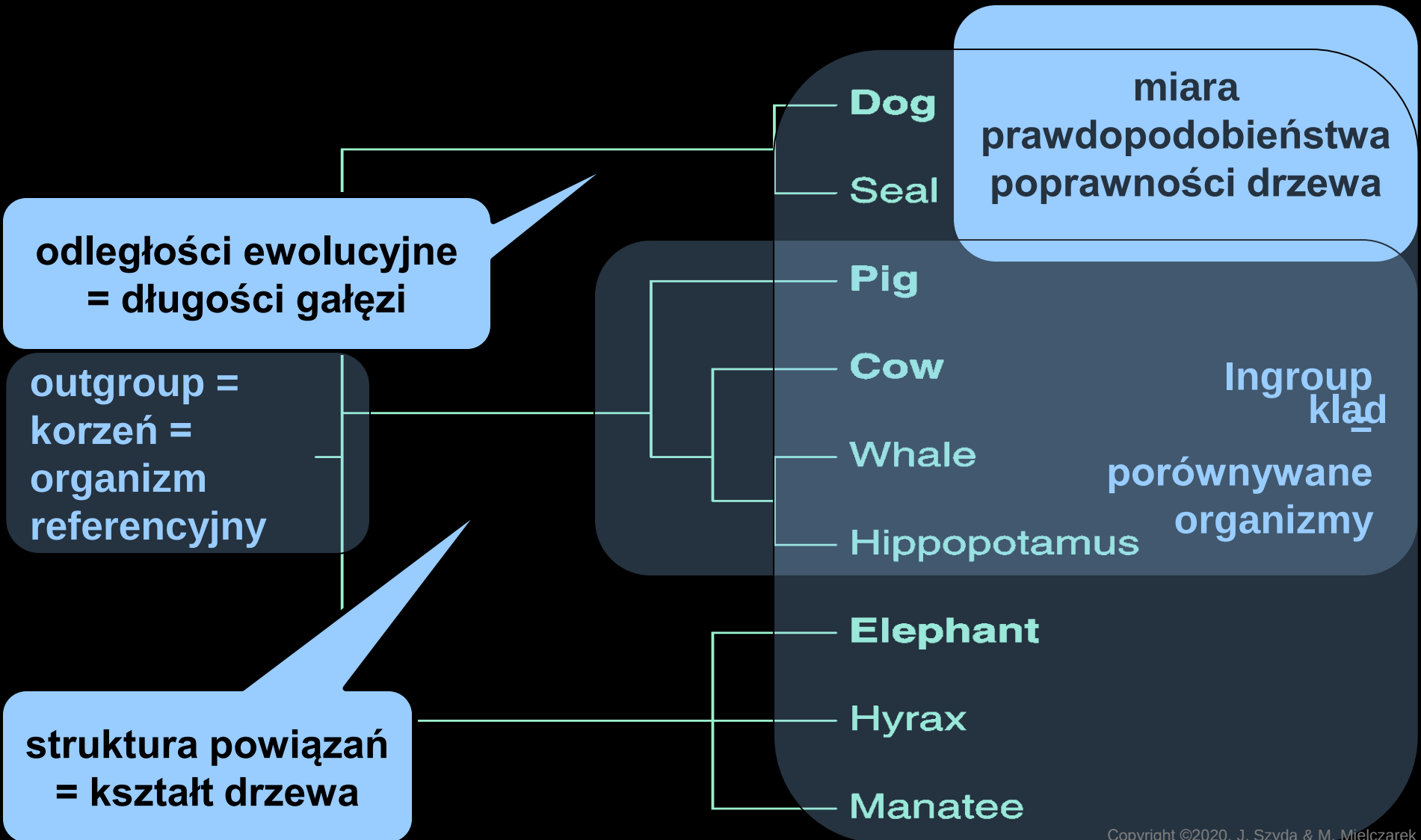
Drzewo
nieukorzenione



STRUKTURA DRZEWA FILOGENETYCZNEGO

DRZEWO UKORZENIONE

ssaki łożyskowe



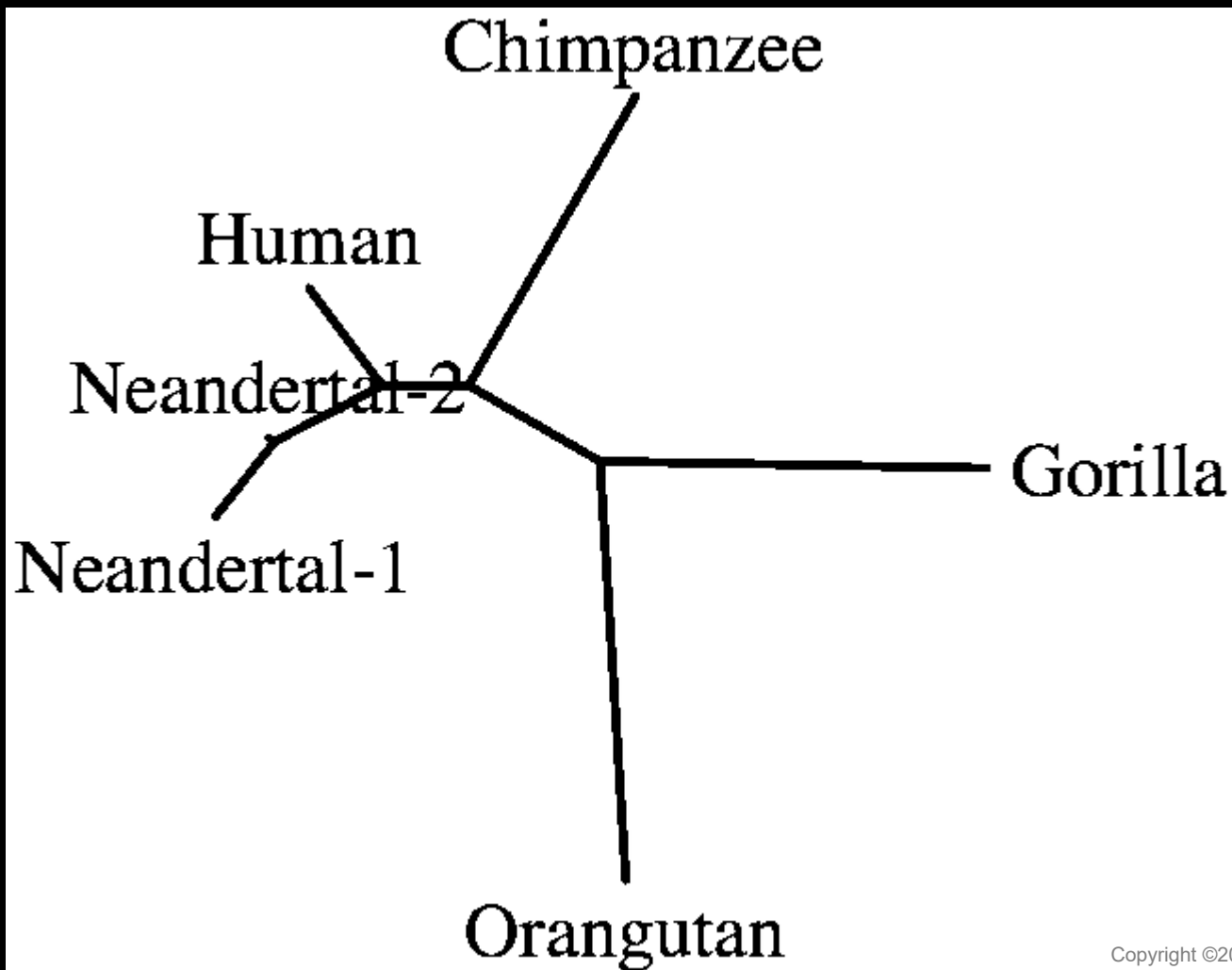
UKORZENIENIE DRZEWA

- Grupa zewnętrzna – gatunek, który jako pierwszy oddzielił się od pozostałych
- Zwykle odległy ewolucyjnie od porównywanych organizmów

STRUKTURA DRZEWA FILOGENETYCZNEGO

DRZEWO NIEUKORZENIONE

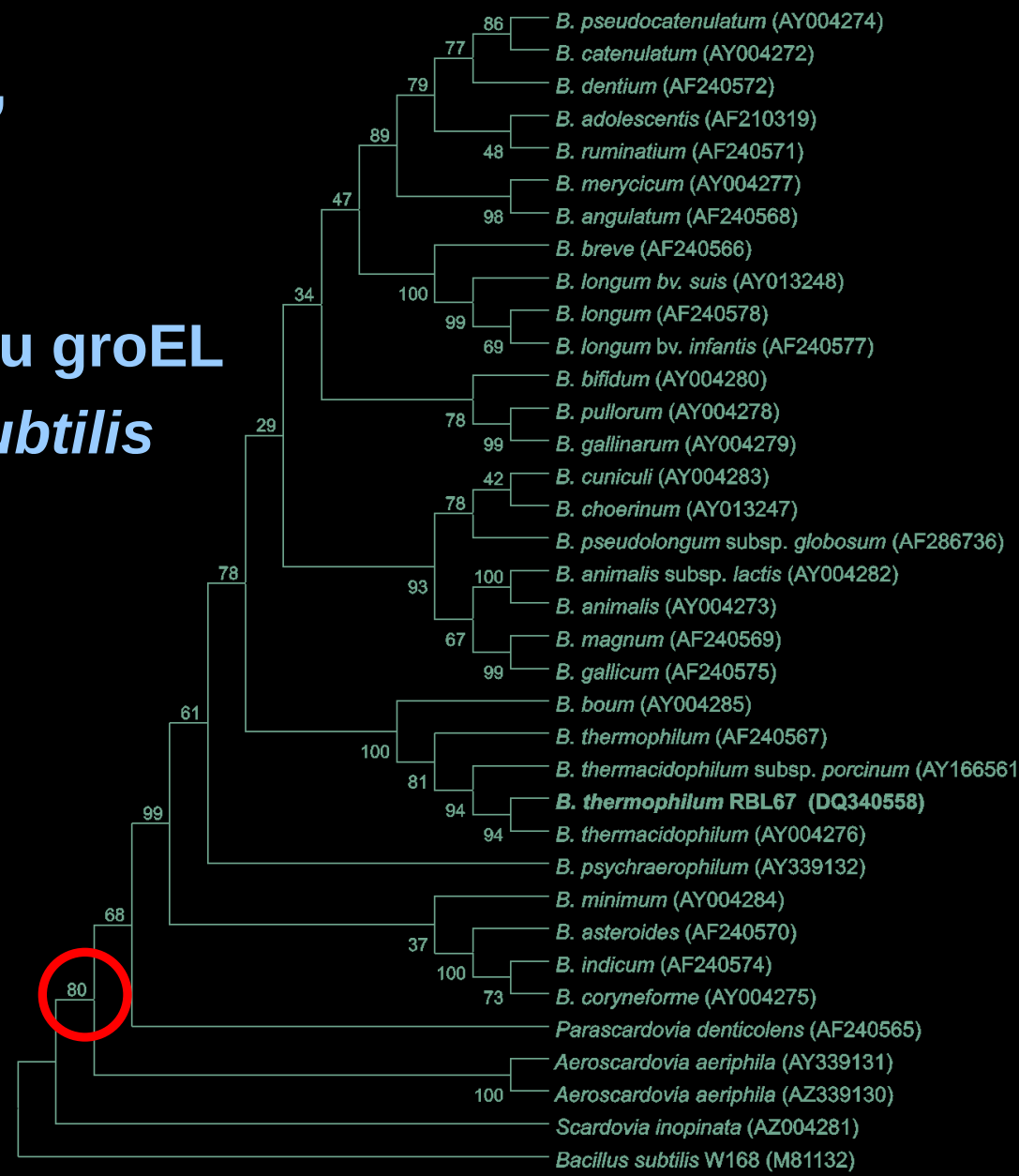
naczelne, mtDNA



STRUKTURA DRZEWA FILOGENETYCZNEGO

DRZEWO UKORZENIONE,
BIFURKACYJNE

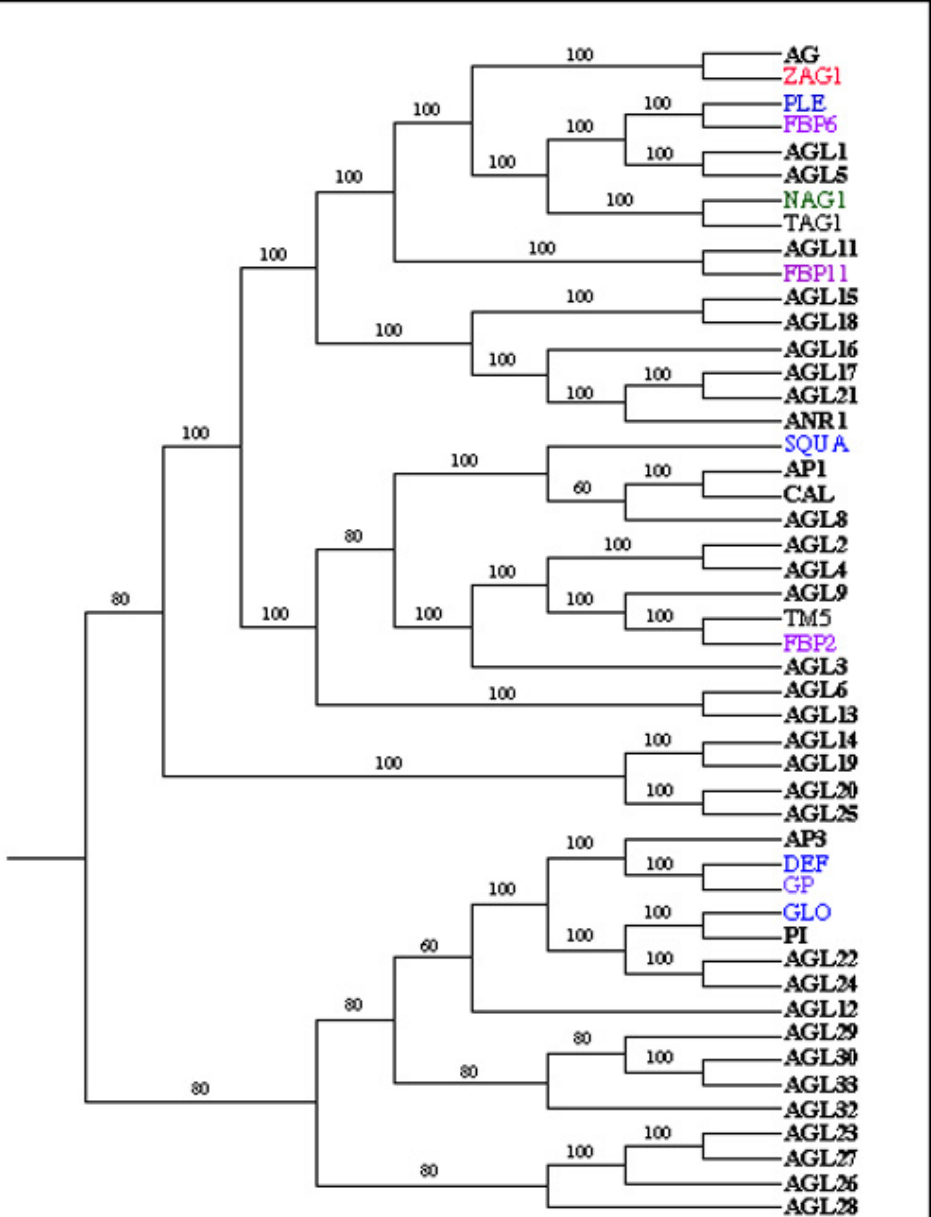
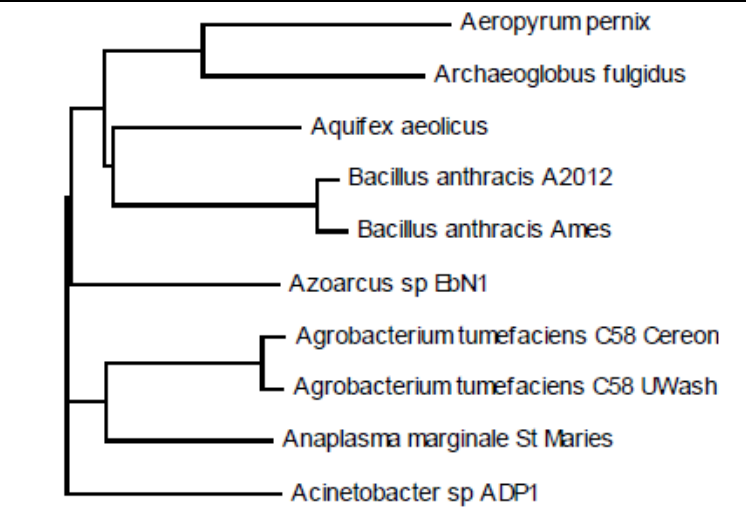
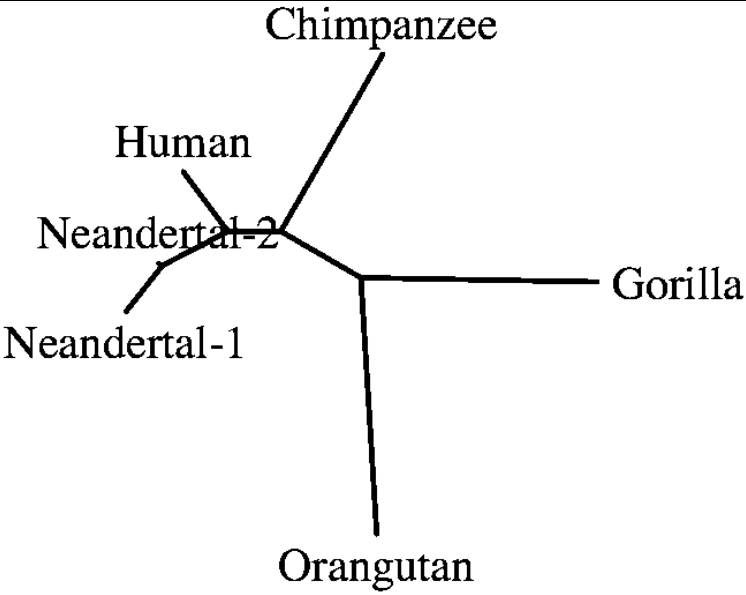
sekwencja fragmentu genu groEL
ukorzeniona w *Bacillus subtilis*



0.5

STRUKTURA DRZEWA FILOGENETYCZNEGO

PRZYKŁADY





Research article

Open Access

Analysis of complete mitochondrial genome sequences increases phylogenetic resolution of bears (Ursidae), a mammalian family that experienced rapid speciation

Li Yu*^{†1}, Yi-Wei Li^{†1}, Oliver A Ryder³ and Ya-Ping Zhang*^{2,1}

Abstract

Background: Despite the small number of ursid species, bear phylogeny has long been a focus of study due to their conservation value, as all bear genera have been classified as endangered at either the species or subspecies level. The Ursidae family represents a typical example of rapid evolutionary radiation. Previous analyses with a single mitochondrial (mt) gene or a small number of mt genes either provide weak support or a large unresolved polytomy for ursids. We revisit the contentious relationships within Ursidae by analyzing complete mt genome sequences and evaluating the performance of both entire mt genomes and constituent mtDNA genes in recovering a phylogeny of extremely recent speciation events.

Results: This mitochondrial genome-based phylogeny provides strong evidence that the spectacled bear diverged first, while within the genus *Ursus*, the sloth bear is the sister taxon of all the other five ursines. The latter group is divided into the brown bear/polar bear and the two black bears/sun bear assemblages. These findings resolve the previous conflicts between trees using partial mt genes. The ability of different categories of mt protein coding genes to recover the correct phylogeny is concordant with previous analyses for taxa with deep divergence times. This study provides a robust Ursidae phylogenetic framework for future validation by additional independent evidence, and also has significant implications for assisting in the resolution of other similarly difficult phylogenetic investigations.

Conclusion: Identification of base composition bias and utilization of the combined data of whole mitochondrial genome sequences has allowed recovery of a strongly supported phylogeny that is upheld when using multiple alternative outgroups for the Ursidae, a mammalian family that underwent a rapid radiation since the mid- to late Pliocene. It remains to be seen if the reliability of mt genome analysis will hold up in studies of other difficult phylogenetic issues. Although the whole mitochondrial DNA sequence based phylogeny is robust, it remains in conflict with phylogenetic relationships suggested by analysis of limited nuclear-encoded data, a situation that will require gathering more nuclear DNA sequence information.

PRZYKŁAD Z LITERATURY

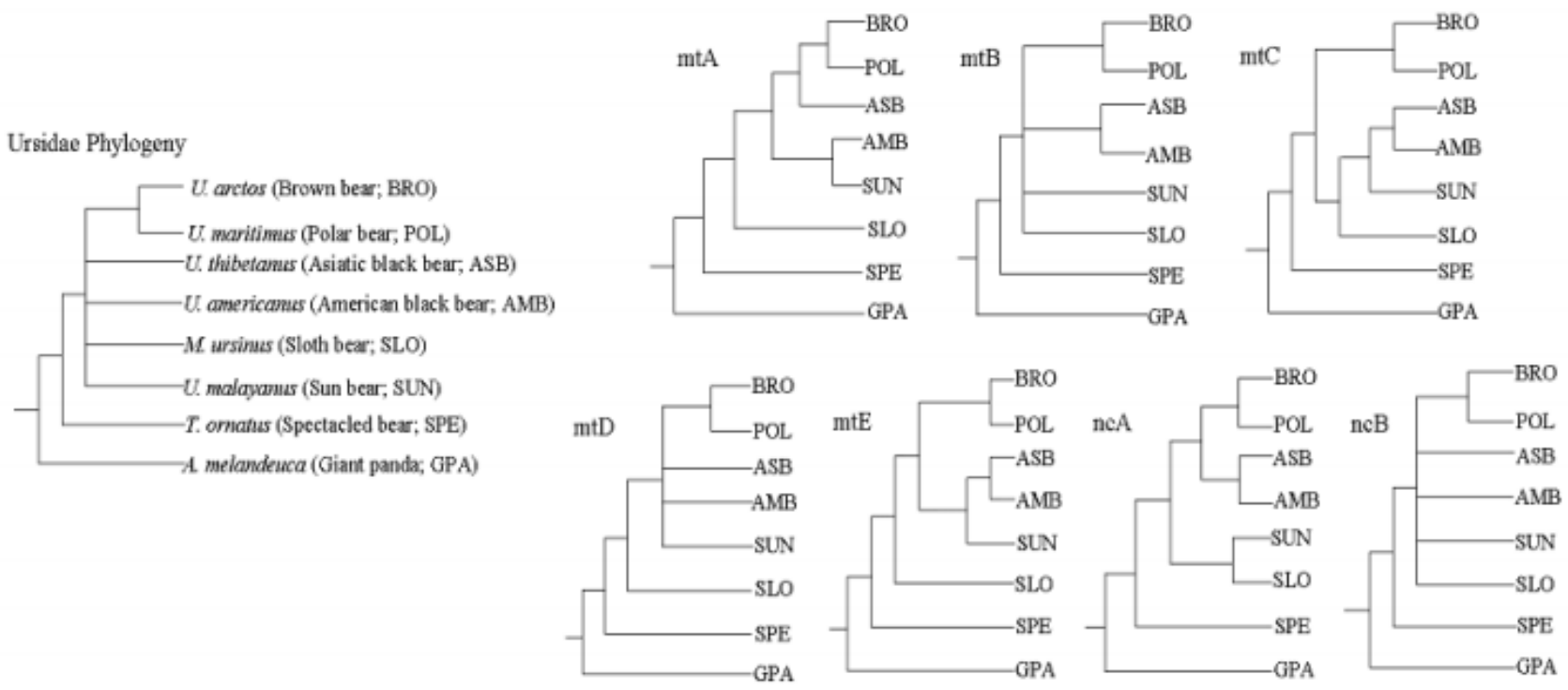


Figure 1
Long-standing unresolved Ursidae phylogeny (tree in the left) and competing hypotheses proposed based on previous sequence data (mtA-E and nuA-B). Trees were constructed from (mtA) combined analysis of partial control region, 12SrRNA, CYTB, tRNA^{Pro}, and tRNA^{Thr} mt genes [11], (mtB) MP analysis of complete CYTB, tRNA^{Pro}, and tRNA^{Thr} mt genes [12], (mtC) NJ analysis of complete CYTB, tRNA^{Pro}, and tRNA^{Thr} mt genes [12], (mtD) combined analysis of partial control region, CYTB, ND4, ND5, COII, and 16SrRNA mt genes [8], (mtE) combined analysis of partial control region, 12SrRNA, complete CYTB, tRNA^{Pro}, and tRNA^{Thr} mt genes [13], (nuA) combined analysis of interphotoreceptor retinoid binding protein (IRBP) exon I and transthyretin (TTR) intron I nuclear genes [13], and (nuB) combined analysis of four type I sequence-tagged sites (STS) and IRBP exon I nuclear genes [60].

METODY KONSTRUKCJI DRZEW FILOGENETYCZNYCH

ODLEGŁOŚCIOWE

- konwertowanie znaków na odległości ewolucyjne

- Metoda średnich połączeń (Unweighted pair group method with arithmetic mean; UPGMA)

- Przyłączania sąsiadów (Neighbor-joining; NJ)

OPARTE NA ZNAKACH

- oparte bezpośrednio na znakach sekwencji (zliczanie mutacji)

- Metoda parsymonii (Maximum Parsimony; MP)

- Metoda największej wiarygodności (Maximum likelihood; ML)



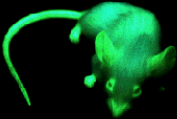
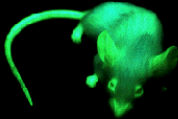
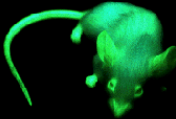
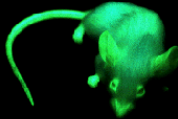
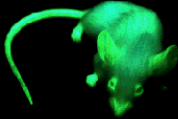
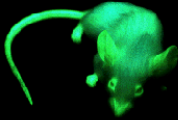
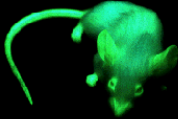
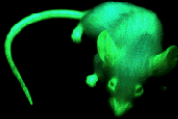
METODA ODLEGŁOŚCIOWA (KLASTROWANIA)

ETAPY → UPGMA

1. Obliczyć macierz zróżnicowania pomiędzy osobnikami
2. Wybór najbardziej podobnych osobników = węzeł
3. Obliczenie nowej macierzy zróżnicowania
4. ... powrót do punktu 2
5. ↺

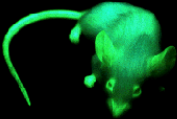
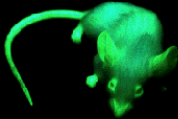
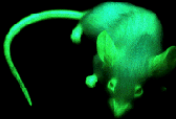
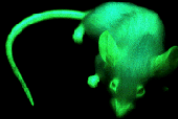
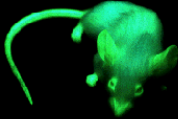
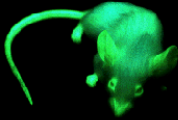
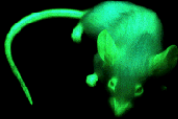
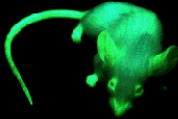
METODY KONSTRUKCJI DRZEWA - UPGMA

1. Obliczyć macierz różnicowania pomiędzy osobnikami

	 ATCC	 ATGC	 TTCG	 TCGG
 ATCC	0	1	2	4
 ATGC		0	3	3
 TTCG			0	2
 TCGG				0

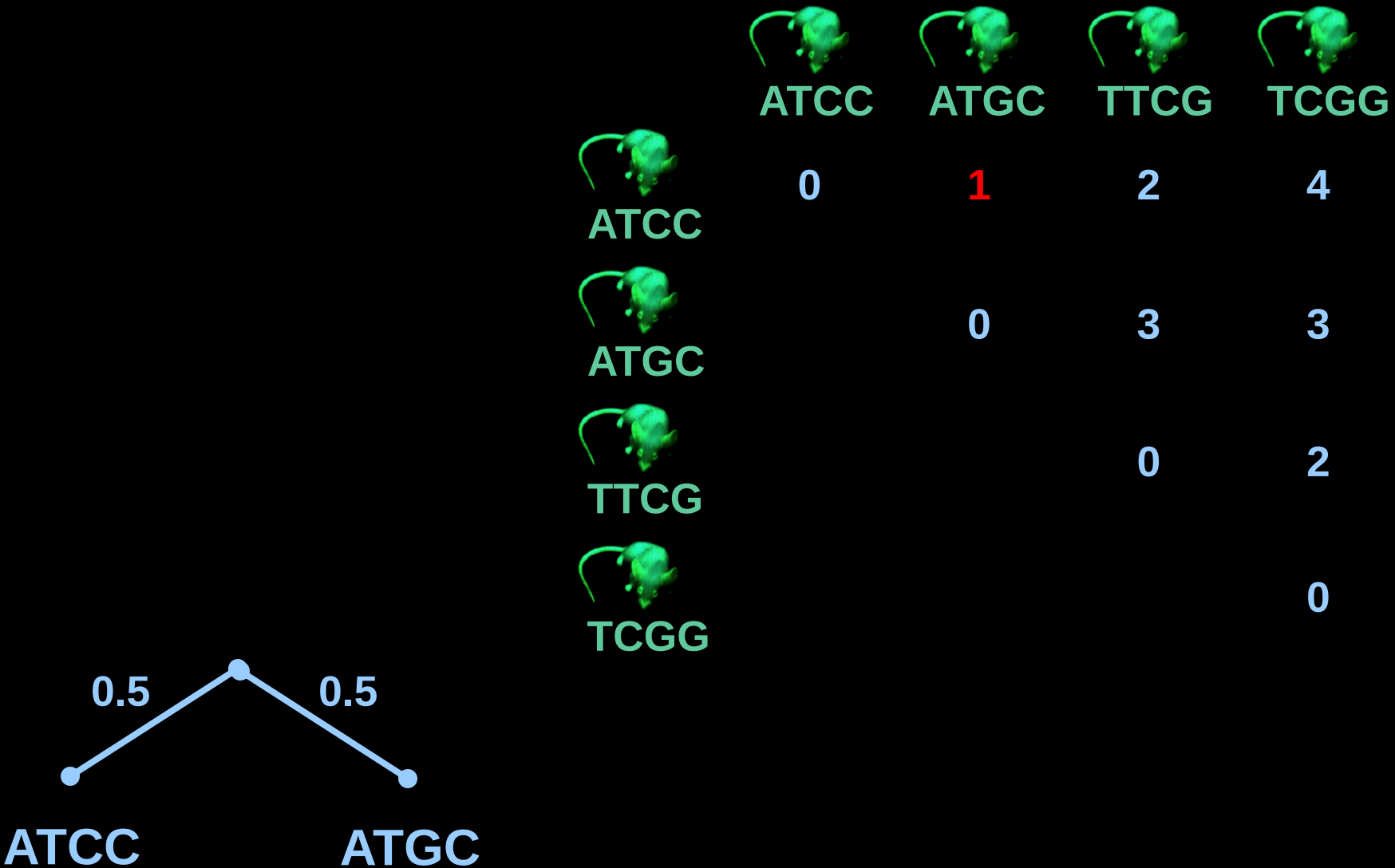
METODY KONSTRUKCJI DRZEWA - UPGMA

2. Wybór najbardziej podobnych osobników = węzeł

	 ATCC	 ATGC	 TTCG	 TCGG
 ATCC	0	1	2	4
 ATGC		0	3	3
 TTCG			0	2
 TCGG				0

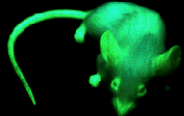
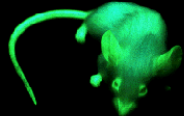
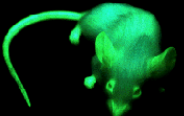
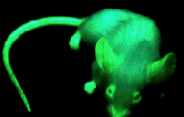
METODY KONSTRUKCJI DRZEWA - UPGMA

2. Wybór najbardziej podobnych osobników = węzeł



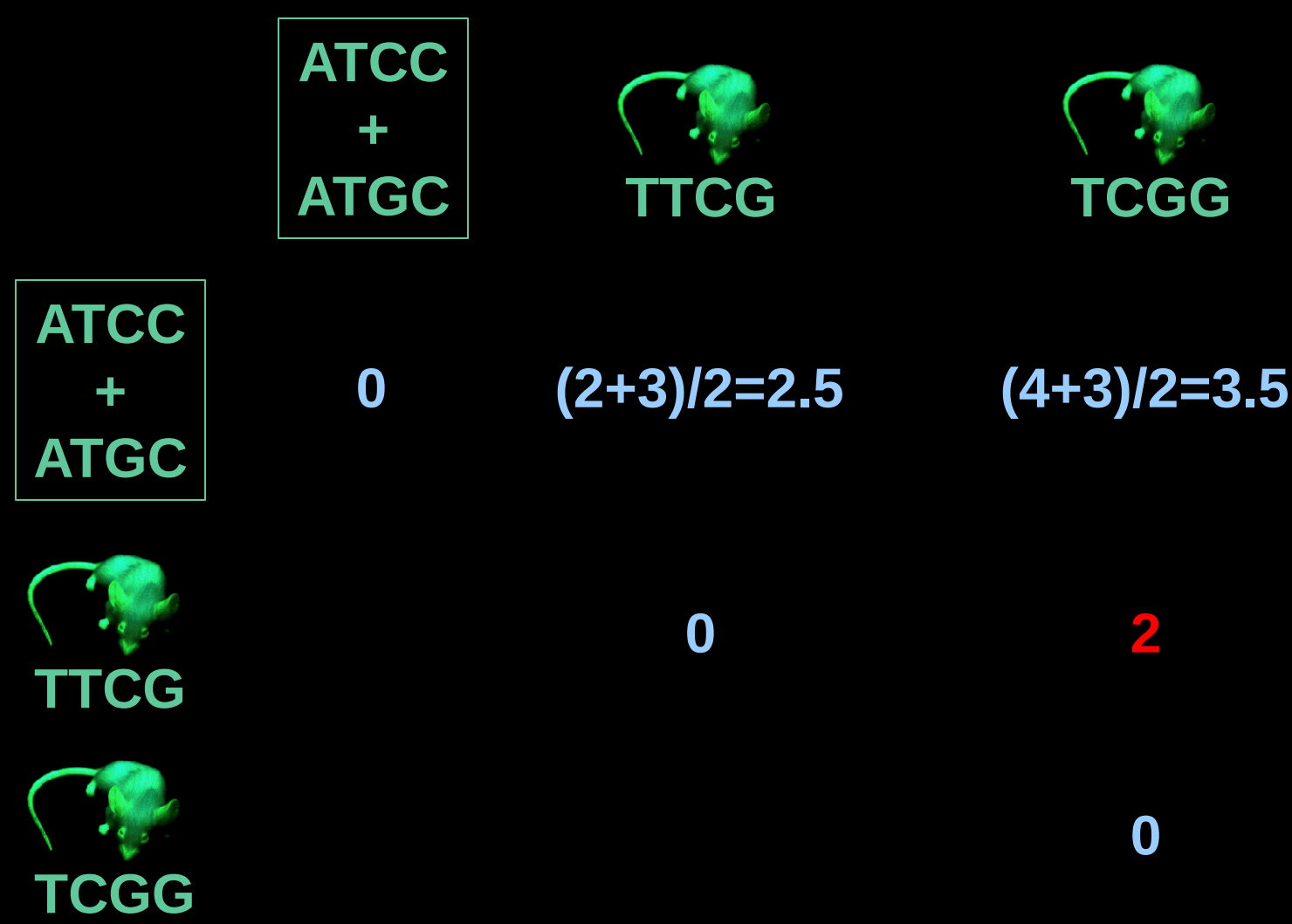
METODY KONSTRUKCJI DRZEWA - UPGMA

3. Obliczenie nowej macierzy zróżnicowania

	ATCC + ATGC	 TTCG	 TCGG
ATCC + ATGC	0	$(2+3)/2=2.5$	$(4+3)/2=3.5$
 TTCG		0	2
 TCGG			0

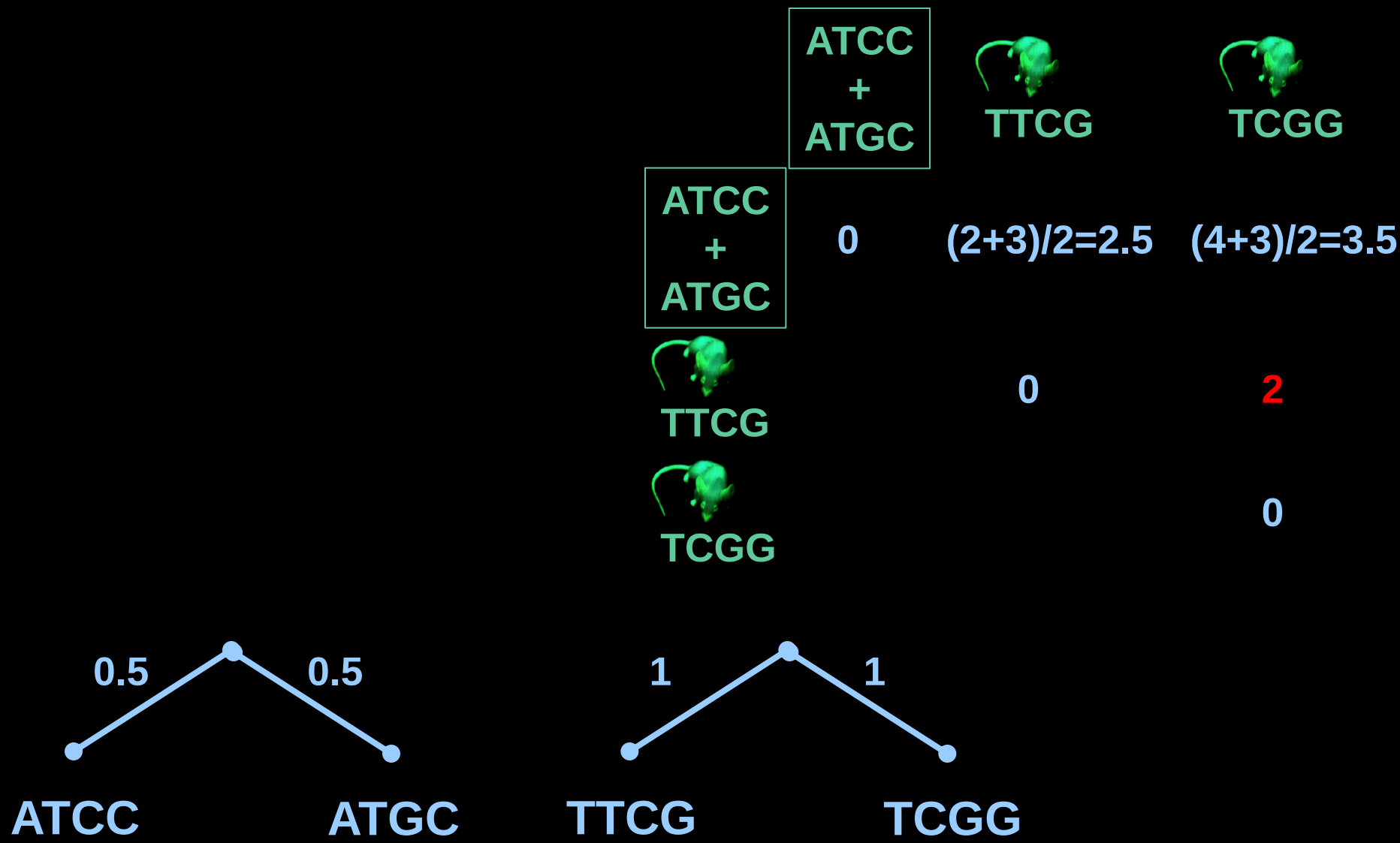
METODY KONSTRUKCJI DRZEWA - UPGMA

4. Wybór najbardziej podobnych osobników = węzeł

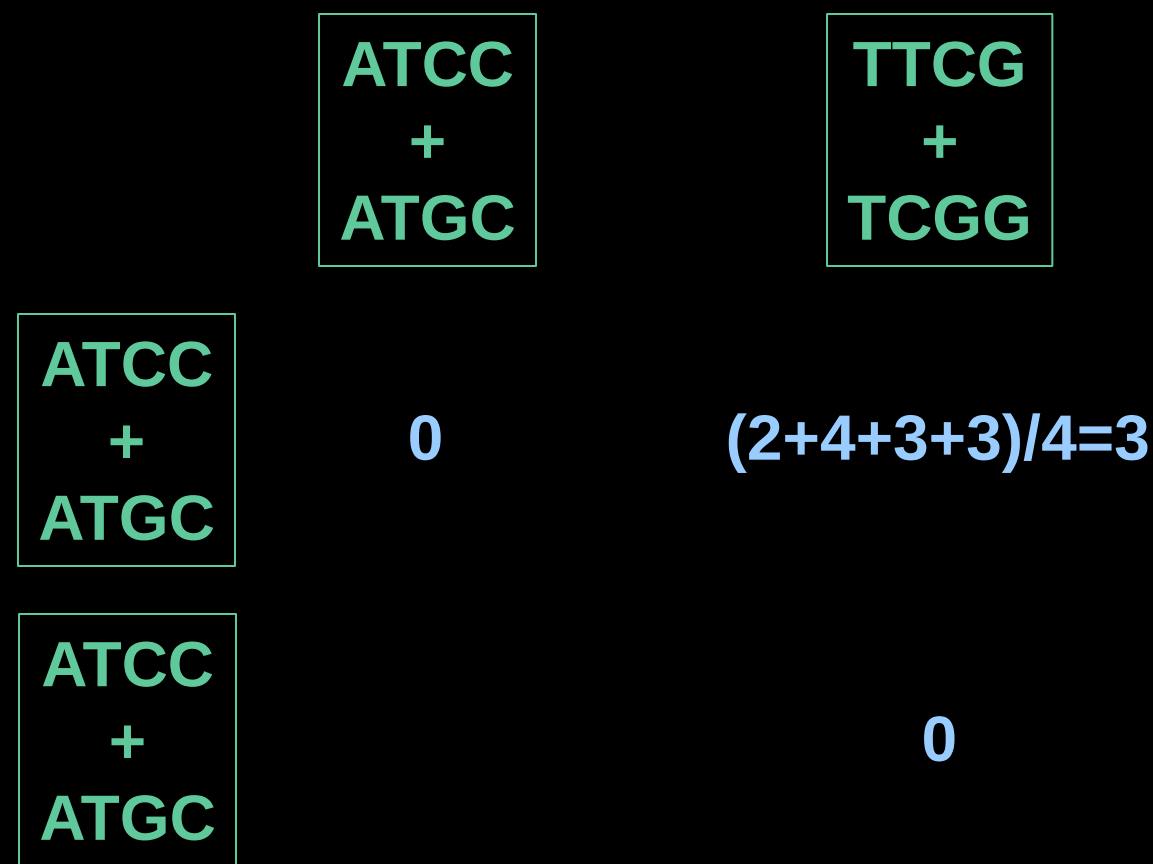


METODY KONSTRUKCJI DRZEWA - UPGMA

4. Wybór najbardziej podobnych osobników = węzeł

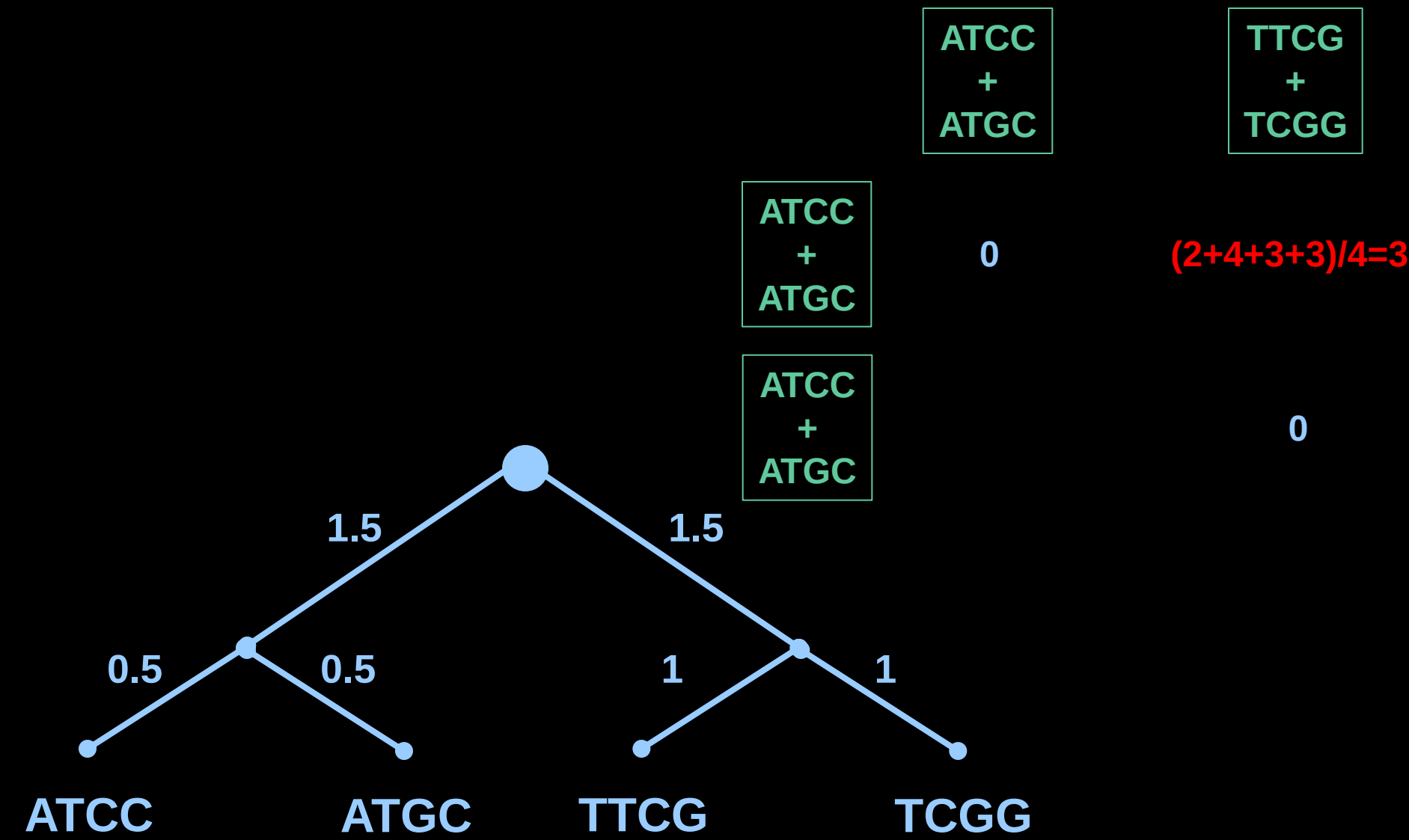


5. Obliczenie nowej macierzy zróżnicowania



METODY KONSTRUKCJI DRZEWA - UPGMA

6. Wybór najbardziej podobnych osobników = węzeł



METODY KONSTRUKCJI DRZEWA - UPGMA

1. Najprostsza metoda tworzenia drzew
2. Bardzo szybka
3. Tworzy drzewa ukorzenione
4. Przyjmuje założenie „zegara molekularnego” → identyczna szybkość mutacji na poszczególnych ścieżkach



1. Pary sekwencji o najmniejszej odległości ewolucyjnej = **sąsiedzi**
2. Stosunkowo szybka
3. Tworzy drzewa nieukorzenione
4. Uwzględnia zróżnicowane tempo ewolucji
5. ...wyniki zależne od założonego modelu ewolucyjnego

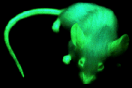
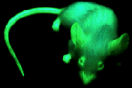
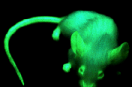
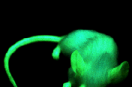


METODA OPARTA NA ZNAKACH (KLADYSTYCZNA)

ETAPY → METODA MAKSYMALNEGO PODOBIEŃSTWA

1. Dopasowanie sekwencji kilku organizmów
2. Konstrukcja (wszystkich) możliwych drzew
3. Wybór drzewa wymagającego najmniejszej liczby mutacji

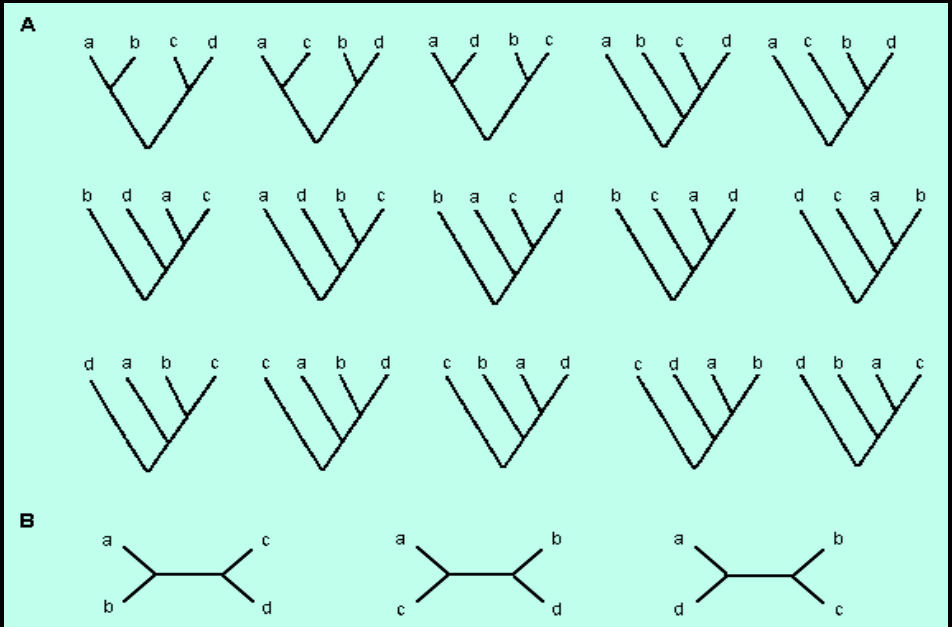
1. Dopasowanie sekwencji kilku organizmów

1		A	A	C	C	G	A	T
2		A	A	C	C	G	C	A
3		A	G	T	C	G	T	T
4		A	G	T	C	G	G	A

- Jednakowe wartości → sekwencja nieinformatywna
- Różne wartości → sekwencja nieinformatywna
- **Powtarzalne wartości → sekwencja informatywna**

2. Konstrukcja (wszystkich) możliwych drzew

n	N _R	N _U
3	3	1
4	15	3
5	105	15
10	34 459 425	2 027 025

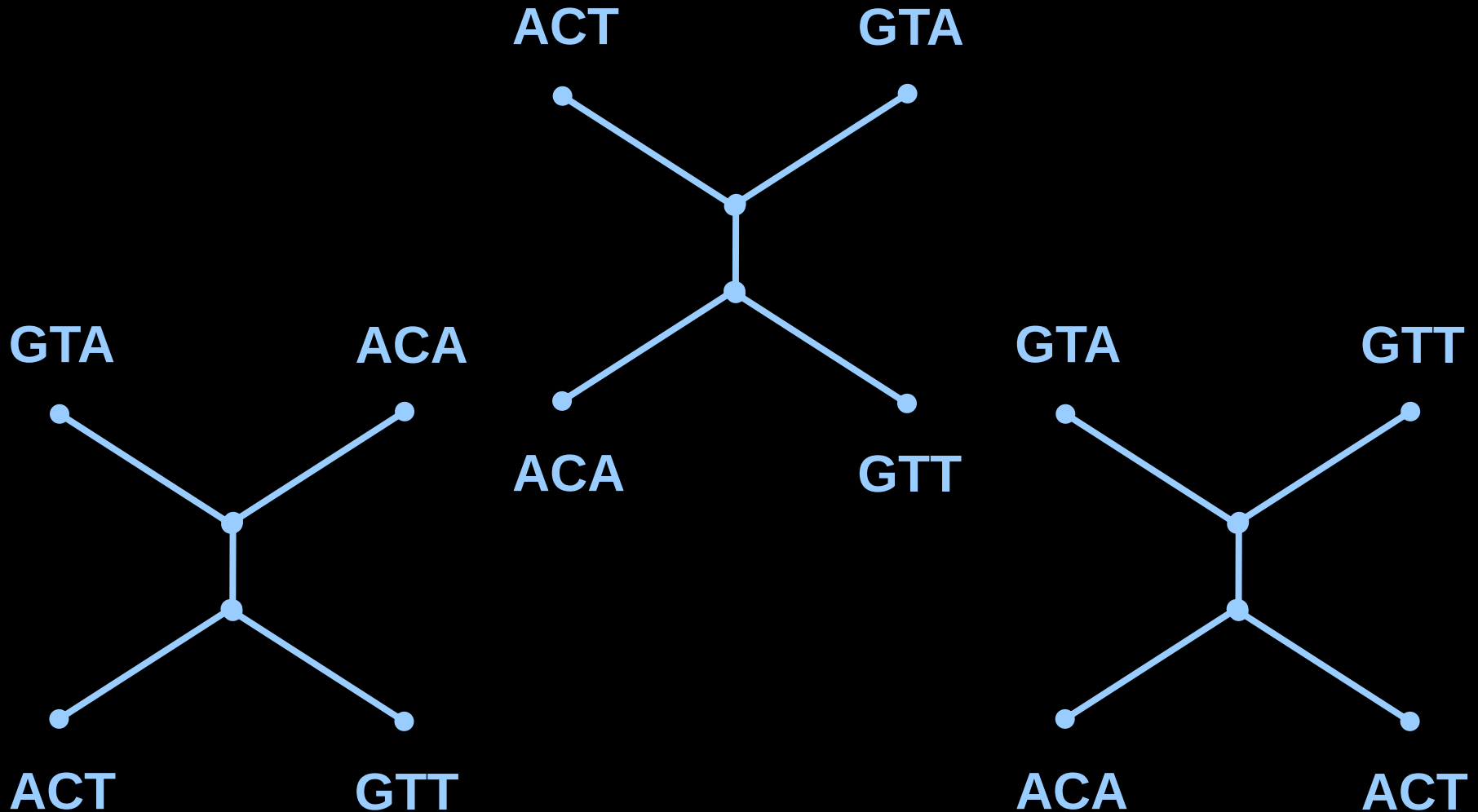


Liczba możliwych drzew:

- ukorzenionych $N_R = (2n-3)!/2^{n-2}(n-2)!$
- nieukorzenionych $N_U = (2n-5)!/2^{n-3}(n-3)!$

n= liczba taksonów/sekwencji

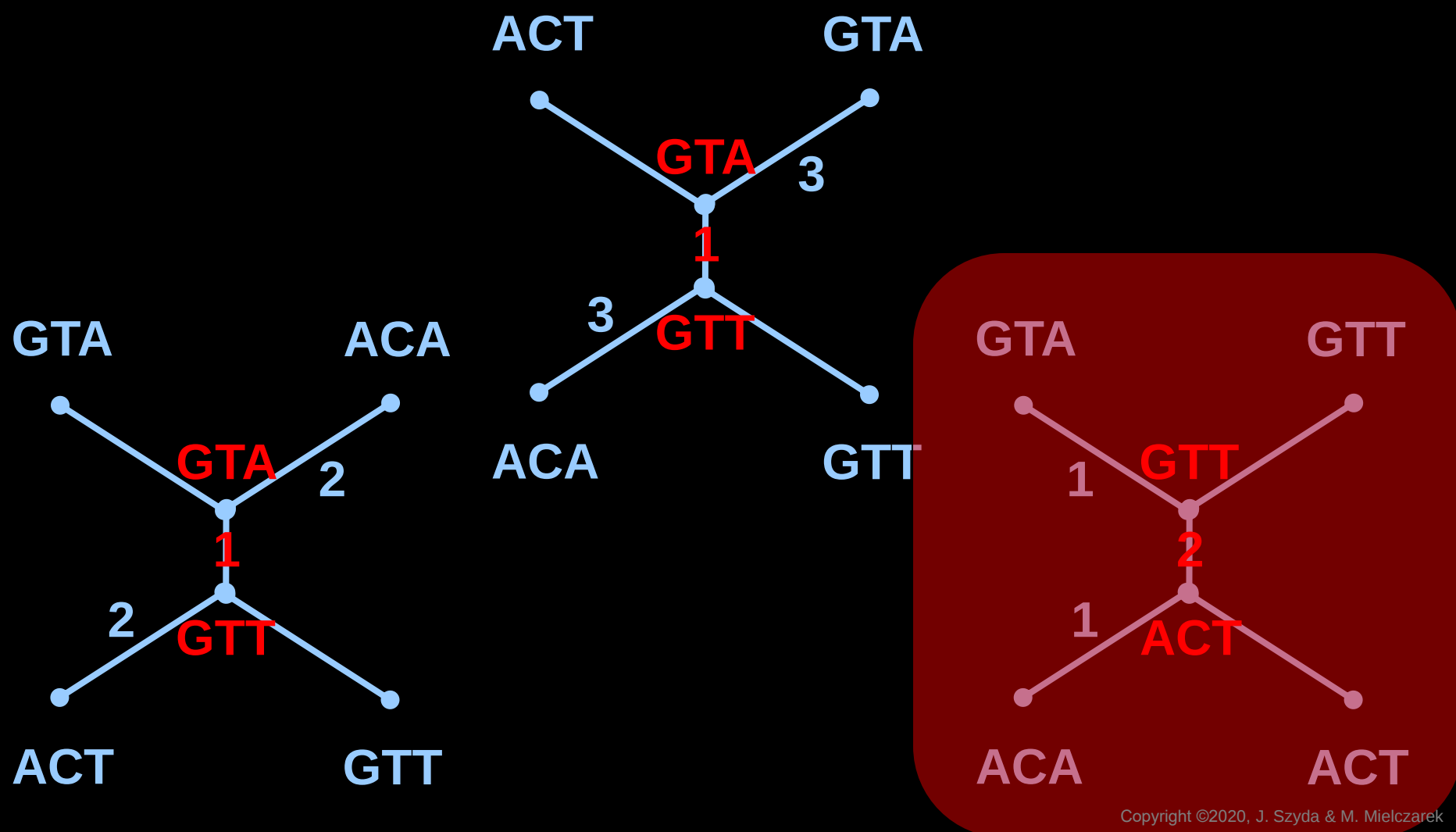
2. Konstrukcja (wszystkich) możliwych drzew



3. Wybór drzewa wymagającego najmniejszej liczby mutacji

- Porównujemy najbliższe węzły między sobą
- Wybieramy po 1 przedstawicielu z tych węzłów i sprawdzamy, co da najmniejsza liczbę zmian

3. Wybór drzewa wymagającego najmniejszej liczby mutacji



Inaczej metoda największej oszczędności

- wybiera drzewa z najmniejszą liczbą zmian ewolucyjnych**
- opiera się na zasadzie „brzytwy Ockhama”**
- Wykorzystuje pozycje informatywne (skrócenie czasu obliczeń)**

METODY KONSTRUKCJI DRZEWA - NAJWYŻSZE PRAWDOP.

- 1. Uwzględnia zróżnicowane prawdopodobieństwo poszczególnych mutacji → modele substytucyjne**
- 2. Uwzględnia każdą pozycję (nie tylko informatywne)**
- 3. Bardzo powolna**
- 4. Daje dokładne wyniki (małe prawdopodobieństwo uzyskania błędnego drzewa)**
- 5. Brak efektu przyciągania długich gałęzi**
- 6. Określa prawdopodobieństwa poprawności danego drzewa**

METODY KONSTRUKCJI DRZEWA

UPGMA

- łatwa, szybka
- analiza dużych zbiorów danych możliwa

METODA PARSYMONII

- powolna, duża liczba możliwych drzew
- analiza dużych zbiorów danych problematyczna

ETAPY KONSTRUKCJI DRZEWA FILOGENETYCZNEGO

1. Wybór i zgromadzenie sekwencji do budowy drzewa
2. Przyrównanie sekwencji (alignment)
3. Konstrukcja drzewa za pomocą jednej z kilku metod
4. Ocena drzewa i przedstawienie drzewa w formie odpowiedniej dla odbiorców

ETAPY KONSTRUKCJI DRZEWA FILOGENETYCZNEGO

1. Wybór i zgromadzenie sekwencji do budowy drzewa

- Zależny od problemu!
- Filogeneza konkretnych organizmów (wybór dowolnego genu)
- Ewolucja konkretnego genu (sekwencje pochodzące od wielu gatunków)
- Wybór markerów → ewolucja ani zbyt szybka, ani zbyt powolna, regiony zmienne, regiony konserwatywne

ETAPY KONSTRUKCJI DRZEWA FILOGENETYCZNEGO

1. Wybór i zgromadzenie sekwencji do budowy drzewa

NCBI

Resources

How To

rzekotka

My NCBI

Sign Out

Nucleotide

Nucleotide

(ursus arctos[Organism]) AND cytochrome b

Search

Create alert

Advanced

Help

NCBI

Resources

How To

rzekotka

My NCBI

Sign Out

Nucleotide

Nucleotide

Search

Advanced

Help

GenBank

Send

Change region shown

Customize view

Analyze this sequence

Run BLAST

Pick Primers

Highlight Sequence Features

Find in this Sequence

Related information

Protein

PubMed

Taxonomy

U.arctos mitochondrial cytochrome b gene

GenBank: X82308.1

FASTA

Graphics

Go to

LOCUS X82308 1140 bp DNA linear MAM 26-JUL-2016

DEFINITION U.arctos mitochondrial cytochrome b gene.

ACCESSION X82308

VERSION X82308.1

KEYWORDS cytochrome b.

SOURCE mitochondrion Ursus arctos (brown bear)

ORGANISM Ursus arctos

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Carnivora; Caniformia; Ursidae; Ursus.

REFERENCE 1

AUTHORS Arnason,U., Bodin,K., Gullberg,A., Ledje,C. and Mouchaty,S.

TITLE A molecular view of pinniped relationships with particular emphasis

Copyright ©2020, J. Szyda & M. Mielczarek

ETAPY KONSTRUKCJI DRZEWA FILOGENETYCZNEGO

1. Wybór i zgromadzenie sekwencji do budowy drzewa

blastn blastp blastx tblastn tblastx

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) [Clear](#) Query subrange

X82308.1

From To

Or, upload file [Wybierz plik](#) Nie wybrano pliku

Job Title

Enter a descriptive title for your BLAST search

☐ Align two or more sequences

Choose Search Set

Database ☒ Standard databases (nr etc.): ☐ rRNA/ITS databases ☐ Genomic + transcript databases ☐ Betacoronavirus

Nucleotide collection (nr/nt)

Organism Optional

Enter organism name or id—completions will be suggested ☐ exclude

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown

Exclude Optional

☐ Models (XM/XP) ☐ Uncultured/environmental sample sequences

Limit to Optional

☐ Sequences from type material

Entrez Query Optional

Enter an Entrez query to limit search

Program Selection

Op ☐ Highly similar sequences (megablast) ☒ More dissimilar sequences (discontiguous megablast) ☐ Somewhat similar sequences (blastn)

Choose a BLAST algorithm

BLAST

Search database Nucleotide collection (nr/nt) using Discontiguous megablast (Optimize for more dissimilar)

☐ Show window

[Algorithm parameters](#)

Descriptions Graphic Sum

Sequences producing significant alignments

☒ select all 100 sequences selected

Manage Columns Show 100

GenBank Graphic Distance tree of results

	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
☒ Ursus arctos isolate Tver203 mito	2057	2057	100%	0.0	100.00%	KY419702.1
☒ Ursus arctos isolate Tver202 mito	2057	2057	100%	0.0	100.00%	KY419701.1
☒ Ursus arctos isolate Tver201 mito	2057	2057	100%	0.0	100.00%	KY419700.1
☒ Ursus arctos isolate Sver219 mito	2057	2057	100%	0.0	100.00%	KY419694.1
☒ Ursus arctos isolate Sver217 mito	2057	2057	100%	0.0	100.00%	KY419693.1
☒ Ursus arctos isolate Sver216 mito	2057	2057	100%	0.0	100.00%	KY419692.1
☒ Ursus arctos isolate Sver215 mito	2057	2057	100%	0.0	100.00%	KY419691.1
☒ Ursus arctos isolate Sver213 mito	2057	2057	100%	0.0	100.00%	KY419690.1
☒ Ursus arctos isolate Sver209 mito	2057	2057	100%	0.0	100.00%	KY419689.1
☒ Ursus arctos isolate Rom11 mito	2057	2057	100%	0.0	100.00%	KY419687.1
☒ Ursus arctos isolate Rom10 mito	2057	2057	100%	0.0	100.00%	KY419686.1
☒ Ursus arctos isolate Rom11 mito	2057	2057	100%	0.0	100.00%	KY419684.1
☒ Ursus arctos isolate Rom10 mito	2057	2057	100%	0.0	100.00%	KY419683.1
☒ Ursus arctos isolate Perm90 mito	2057	2057	100%	0.0	100.00%	KY419681.1
☒ Ursus arctos isolate Perm89 mito	2057	2057	100%	0.0	100.00%	KY419679.1
☒ Ursus arctos isolate Perm79 mito	2057	2057	100%	0.0	100.00%	KY419678.1
☒ Ursus arctos isolate Perm78 mito	2057	2057	100%	0.0	100.00%	KY419677.1
☒ Ursus arctos isolate Perm77 mito	2057	2057	100%	0.0	100.00%	KY419676.1
☒ Ursus arctos isolate Perm76 mito	2057	2057	100%	0.0	100.00%	KY419675.1
☒ Ursus arctos isolate Perm75 mito	2057	2057	100%	0.0	100.00%	KY419674.1

→ „The lower the E-value, or the closer it is to zero, the more "significant" the match is"

→ BLAST FAQ

ETAPY KONSTRUKCJI DRZEWA FILOGENETYCZNEGO

1. Wybór i zgromadzenie sekwencji do budowy drzewa



Sequences producing significant alignments									
Download Manage Columns Show 20000 ?									
select all 20000 sequences selected									
GenBank Graphics Distance tree of results									
	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession		
☒	Ursus arctos isolate Tver203 mitochondrion, partial genome	2057	2057	100%	0.0	100.00%	KY419702.1		
☒	Ursus arctos isolate Tver202 mitochondrion, partial genome	2057	2057	100%	0.0	100.00%	KY419701.1		
☒	Ursus arctos isolate Tver201 mitochondrion, partial genome	2057	2057	100%	0.0	100.00%	KY419700.1		
☒	Ursus arctos isolate Sver219 mitochondrion, partial genome	2057	2057	100%	0.0	100.00%	KY419694.1		
☒	Ursus arctos isolate Sver217 mitochondrion, partial genome	2057	2057	100%	0.0	100.00%	KY419693.1		
☒	Ursus arctos isolate Sver216 mitochondrion, partial genome	2057	2057	100%	0.0	100.00%	KY419692.1		
☒	Ursus arctos isolate Sver215 mitochondrion, partial genome	2057	2057	100%	0.0	100.00%	KY419691.1		
☒	Ursus arctos isolate Sver213 mitochondrion, partial genome	2057	2057	100%	0.0	100.00%	KY419690.1		
☒	Ursus arctos isolate Sver209 mitochondrion, partial genome	2057	2057	100%	0.0	100.00%	KY419689.1		
☒	Ursus arctos isolate Sver206 mitochondrion, partial genome	2057	2057	100%	0.0	100.00%	KY419688.1		
☒	Ursus spelaeus complete mitochondrial genome, isolate SP1325	1714	1714	100%	0.0	93.33%	FN390849.1		
☒	Ursus spelaeus complete mitochondrial genome, isolate SP370	1714	1714	100%	0.0	93.33%	FN390847.1		
☒	Ursus arctos isolate UA141 cytochrome b (CYTB) gene, partial cds; tRNA-Thr and tRNA-Pro genes, complete sequence; and D-loop, partial cds	1712	1712	90%	0.0	96.89%	MF927787.1		
☒	Ursus arctos isolate UA151 cytochrome b (CYTB) gene, partial cds; tRNA-Thr and tRNA-Pro genes, complete sequence; and D-loop, partial cds	1710	1710	90%	0.0	96.89%	MF927788.1		
☒	Ursus ingressus isolate USR10 mitochondrion, partial genome	1709	1709	100%	0.0	93.25%	MN311249.1		
☒	Ursus ingressus voucher PA1 mitochondrion, partial genome	1709	1709	100%	0.0	93.25%	KX641330.1		
☒	Ursus spelaeus voucher CB_A2 mitochondrion, partial genome	1709	1709	100%	0.0	93.25%	KX641309.1		
☒	Ursus arctos isolate Perm77 mitochondrion, partial genome	2057	2057	100%	0.0	100.00%	KY419675.1		
☒	Ursus arctos isolate Perm76 mitochondrion, partial genome	2057	2057	100%	0.0	100.00%	KY419674.1		

ETAPY KONSTRUKCJI DRZEWA FILOGENETYCZNEGO

1. Wybór i zgromadzenie sekwencji do budowy drzewa
2. Przyrównanie sekwencji (alignment)

MEGA Molecular Evolutionary Genetics Analysis

features documentation feedback ask MEGA-GPT

MEGA 12: fast, filtered, substitution model selection
Filtering approach that avoids computations for sub-optimal substitution models.

Windows Graphical (GUI) MEGA 12 (64-bit) DOWNLOAD

Sequence Analyses	Statistical Methods	Powerful Visual Tools
Phylogeny Inference	Maximum Likelihood	Alignment/Trace Editor
Model Selection	Distance Methods	Tree Explorer
Dating and Clocks	Ordinary Least Squares	Data Explorers
Ancestral States	Maximum Parsimony	Legend Generator
Selection and Tests	Composite Likelihood	Gene Duplication Wizard
Sequence Alignment	Bayesian	Timetree Wizard

Copyright ©2020, J. Szyda & M. Mielczarek

ETAPY KONSTRUKCJI DRZEWA FILOGENETYCZNEGO

1. Wybór i zgromadzenie sekwencji do budowy drzewa
2. Przyrównanie sekwencji (alignment)

→ Wybranie pliku
→ Przyrównanie

ETAPY KONSTRUKCJI DRZEWA FILOGENETYCZNEGO

1. Wybór i zgromadzenie sekwencji do budowy drzewa
2. Przyrównanie sekwencji (alignment)

M12: Alignment Explorer (seqdump.fasta)

Alignment Explorer interface showing DNA sequences and alignment options.

Alignment Menu Options:

- Align by ClustalW
- Align by ClustalW (Codons)
- Align by MUSCLE
- Align by MUSCLE (Codons)
- Mark/Unmark Site
- Align Marked Sites
- Unmark All Sites
- Delete Gap-Only Sites
- Auto-Fill Gaps

→ Select all

The interface displays a table of DNA sequences and their alignment. The sequences are color-coded by nucleotide (A, G, C, T) and aligned across multiple columns. The alignment is shown in a grid format with columns representing positions and rows representing individual sequences.

ETAPY KONSTRUKCJI DRZEWA FILOGENETYCZNEGO

1. Wybór i zgromadzenie sekwencji do budowy drzewa
2. Przyrównanie sekwencji (a

M M12: Alignment Explorer (seqdump.fasta)

The screenshot shows the 'Alignment Explorer' application window. The 'Alignment' menu is open, displaying options: 'Align by ClustalW', 'Align by ClustalW (Codons)', 'Align by MUSCLE', 'Align by MUSCLE (Codons)', 'Mark/Unmark Site', 'Align Marked Sites', 'Unmark All Sites', 'Delete Gap-Only Sites', and 'Auto-Fill Gaps'. A large black button with the text '→ Select all' is overlaid on the bottom right of the menu. The background shows a table of DNA sequences and their translated protein sequences, with some sites marked with asterisks.

Species/Abbrv	DNA Sequences	Translated Protein Seq
1. KY419702.1	AAGTTTAAATC	*
2. KY419601.1	GTTTATGTAG	*
3. HQ685919.1	GTTTATGTAG	*
4. KY419702.5	AAGTTTAAATC	*

The screenshot shows the 'ClustalW Parameters' dialog box. It has two tabs: 'Alignment' and 'Matrix'. The 'Alignment' tab is active, showing settings for 'Pairwise Alignment' and 'Multiple Alignment'. The 'Pairwise Alignment' section has 'Gap Opening Penalty' set to 15.00 and 'Gap Extension Penalty' set to 6.66. The 'Multiple Alignment' section has 'Gap Opening Penalty' set to 15.00 and 'Gap Extension Penalty' set to 6.66. The 'Matrix' tab is also visible, showing settings for 'DNA Weight Matrix' (IUB), 'Transition Weight' (0.50), 'Use Negative Matrix' (OFF), 'Delay Divergent Cutoff (%)' (30), 'Keep Predefined Gaps' (unchecked), and 'Guide Tree'.

ClustalW Parameters

Alignment

Pairwise Alignment

Gap Opening Penalty: 15.00

Gap Extension Penalty: 6.66

Multiple Alignment

Gap Opening Penalty: 15.00

Gap Extension Penalty: 6.66

Matrix

DNA Weight Matrix: IUB

Transition Weight: 0.50

Use Negative Matrix: OFF

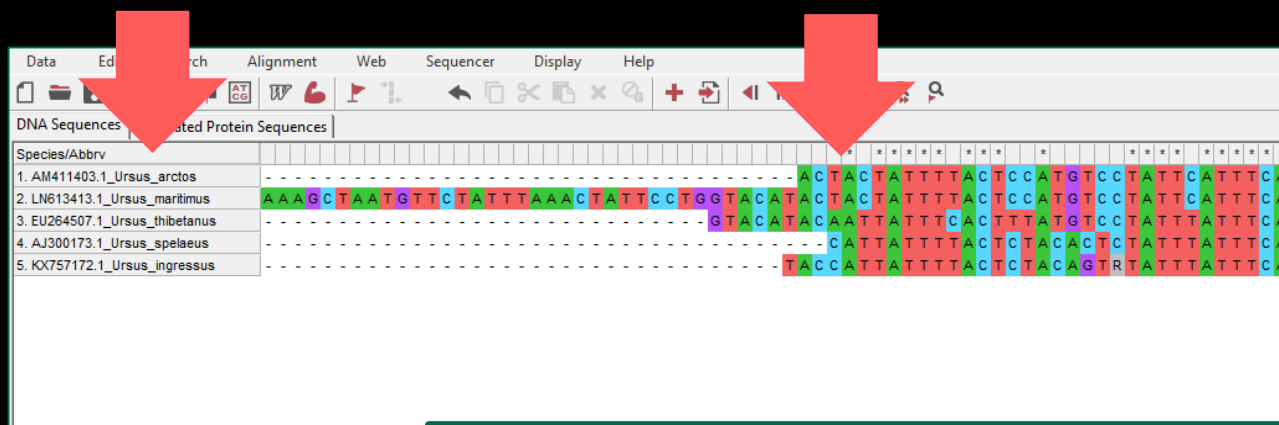
Delay Divergent Cutoff (%): 30

Keep Predefined Gaps: ☐

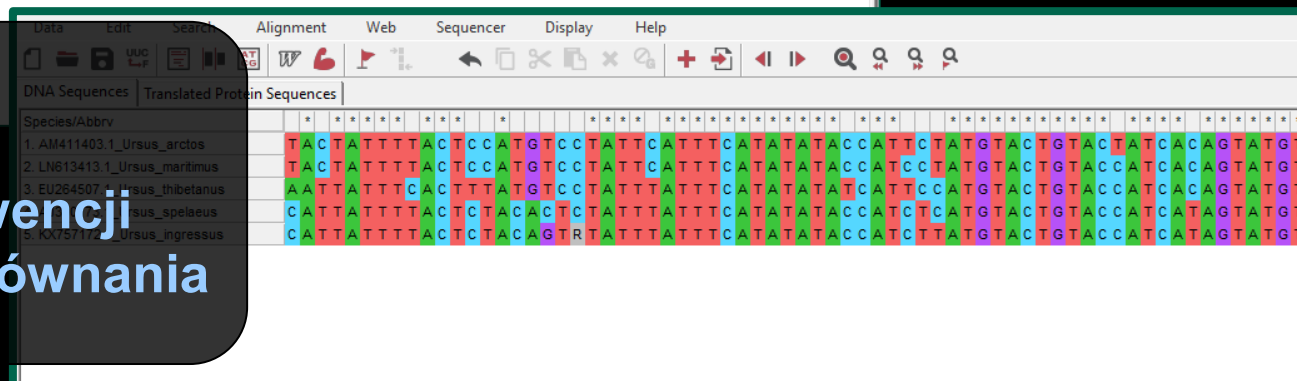
Guide Tree:

ETAPY KONSTRUKCJI DRZEWA FILOGENETYCZNEGO

1. Wybór i zgromadzenie sekwencji do budowy drzewa
2. Przyrównanie sekwencji (alignment)



- Wybranie pliku
- Przyrównanie
- Przycięcie sekwencji
- Zapisanie przyrównania w nowym pliku



ETAPY KONSTRUKCJI DRZEWA FILOGENETYCZNEGO

- 1. Wybór i zgromadzenie sekwencji do budowy drzewa**
- 2. Przyrównanie sekwencji (alignment)**
- 3. Konstrukcja drzewa za pomocą jednej z kilku metod**

ETAPY KONSTRUKCJI DRZEWA FILOGENETYCZNEGO

1. Wybór i zgromadzenie sekwencji do budowy drzewa
2. Przyrównanie sekwencji (alignment)

The screenshot shows the MEGA 12 website. At the top, the logo 'MEGA' is displayed with the text 'Molecular Evolutionary Genetics Analysis' below it. Navigation links for 'features', 'documentation', 'feedback', and 'ask MEGA-GPT' are on the right. A horizontal row of 20 small icons represents various organisms. Below this is a dark banner with the text 'MEGA 12: fast, filtered, substitution model selection' and 'Filtering approach that avoids computations for sub-optimal substitution models.' A progress indicator shows 8 circles, with the 3rd one highlighted in red. Below the banner are three dropdown menus: 'Windows', 'Graphical (GUI)', and 'MEGA 12 (64-bit)', followed by a green 'DOWNLOAD' button. A large red arrow points from the 'Graphical (GUI)' dropdown to the 'Statistical Methods' column. The main content area is divided into three columns: 'Sequence Analyses', 'Statistical Methods', and 'Powerful Visual Tools'. The 'Statistical Methods' column lists: Maximum Likelihood, Distance Methods, Ordinary Least Squares, Maximum Parsimony, Composite Likelihood, and Bayesian. The 'Powerful Visual Tools' column lists: Alignment/Trace Editor, Tree Explorer, Data Explorers, Legend Generator, Gene Duplication Wizard, and Timetree Wizard. The footer contains the copyright notice: 'Copyright ©2020, J. Szyda & M. Mielczarek'.

MEGA 12: fast, filtered, substitution model selection
Filtering approach that avoids computations for sub-optimal substitution models.

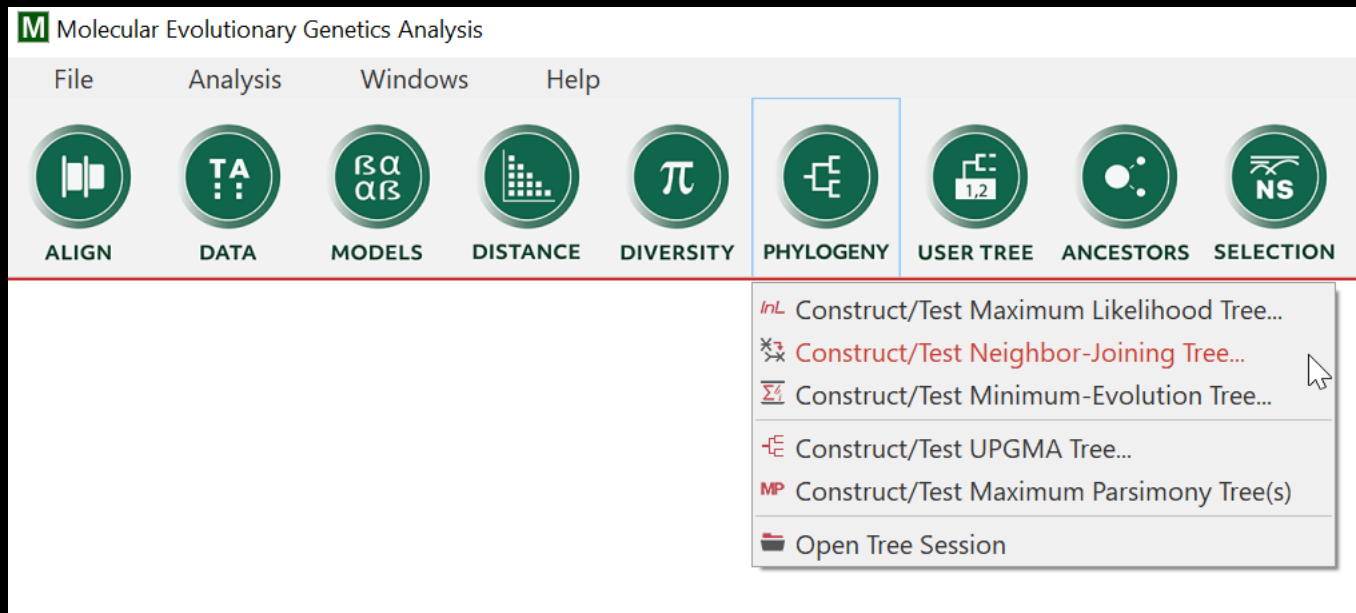
Windows Graphical (GUI) MEGA 12 (64-bit) DOWNLOAD

Sequence Analyses	Statistical Methods	Powerful Visual Tools
Phylogeny Inference	Maximum Likelihood	Alignment/Trace Editor
Model Selection	Distance Methods	Tree Explorer
Dating and Clocks	Ordinary Least Squares	Data Explorers
Ancestral States	Maximum Parsimony	Legend Generator
Selection and Tests	Composite Likelihood	Gene Duplication Wizard
Sequence Alignment	Bayesian	Timetree Wizard

Copyright ©2020, J. Szyda & M. Mielczarek

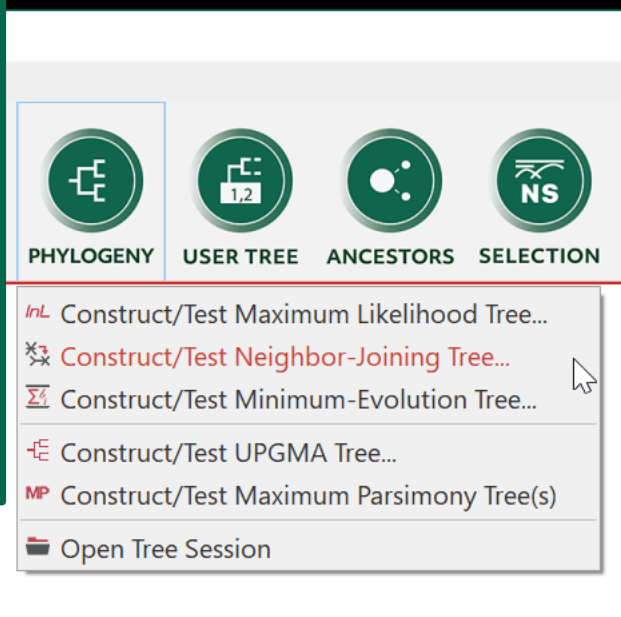
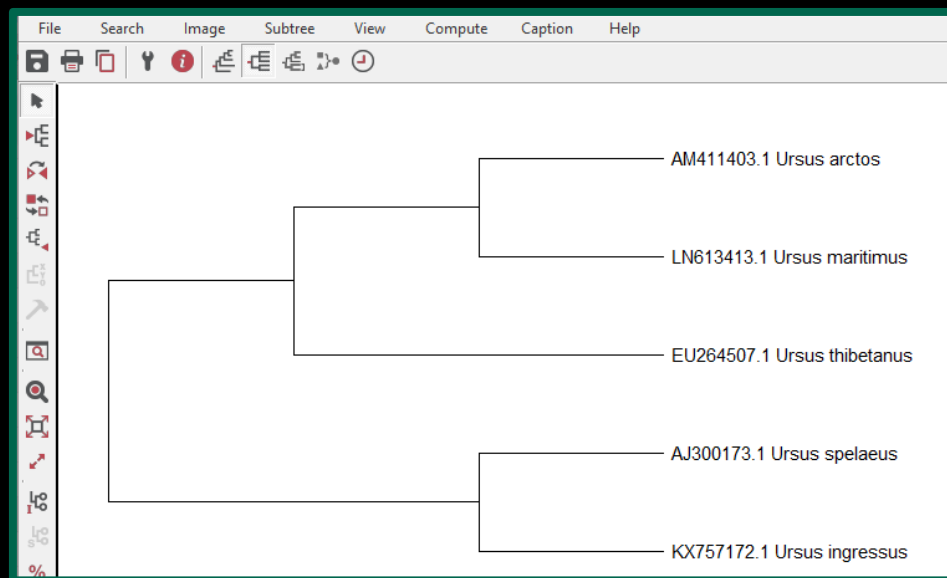
ETAPY KONSTRUKCJI DRZEWA FILOGENETYCZNEGO

1. Wybór i zgromadzenie sekwencji do budowy drzewa
2. Przyrównanie sekwencji (alignment)
3. Konstrukcja drzewa za pomocą jednej z kilku metod



ETAPY KONSTRUKCJI DRZEWA FILOGENETYCZNEGO

1. Wybór i zgromadzenie sekwencji do budowy drzewa
2. Przyrównanie sekwencji (alignment)
3. Konstrukcja drzewa za pomocą jednej z kilku



ETAPY KONSTRUKCJI DRZEWA FILOGENETYCZNEGO

1. Wybór i zgromadzenie sekwencji do budowy drzewa
2. Przyrównanie sekwencji (alignment)
3. Konstrukcja drzewa za pomocą jednej z kilku metod
4. Ocena drzewa i przedstawienie drzewa w formie odpowiedniej dla odbiorców
 1. Ocena wiarygodności drzewa
 2. Czytelna grafika

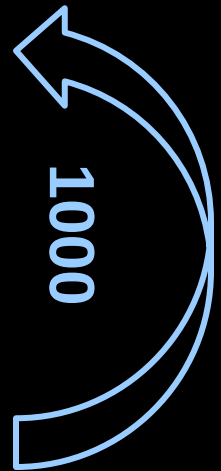
OCENA WIARYGODNOŚCI TOPOLOGII

Prawdopodobieństwo poprawności drzewa → bootstrap

STWORZENIE SZTUCZNEGO ZBIORU DANYCH
zamiana kolejności nukleotydów



STWORZENIE DRZEWA FILOGENETYCZNEGO



**OKREŚLENIE POWTARZALNOŚCI DANEGO
ROZGAŁĘZIENIA = PRAWDOPODOBIEŃSTWO**