

Filogenetyka molekularna żyjących naczelnych



Wykonała: Edyta Michalska

Przedstawię wyniki i metodologię publikacji:

1. A Molecular Phylogeny of Living Primates, Perelman P. et. al., PLOS Genetics, 2011.

- Autorzy zbadali filogenezę naczelnych w oparciu o sekwencje jądrowego DNA z 191 gatunków, w tym 186 gatunków naczelnych obejmujących 90% opisanych rodzajów naczelnych.

2. A Mitogenomic Phylogeny of Living Primates, Finstermeier K. et. al., PLOS ONE, 2013.

- Filogeneza naczelnych została zbadana w oparciu o sekwencje mitochondrialnego DNA pochodzące od 78 gatunków naczelnych.

Użyta grupa zewnętrzna (w pracy 1)

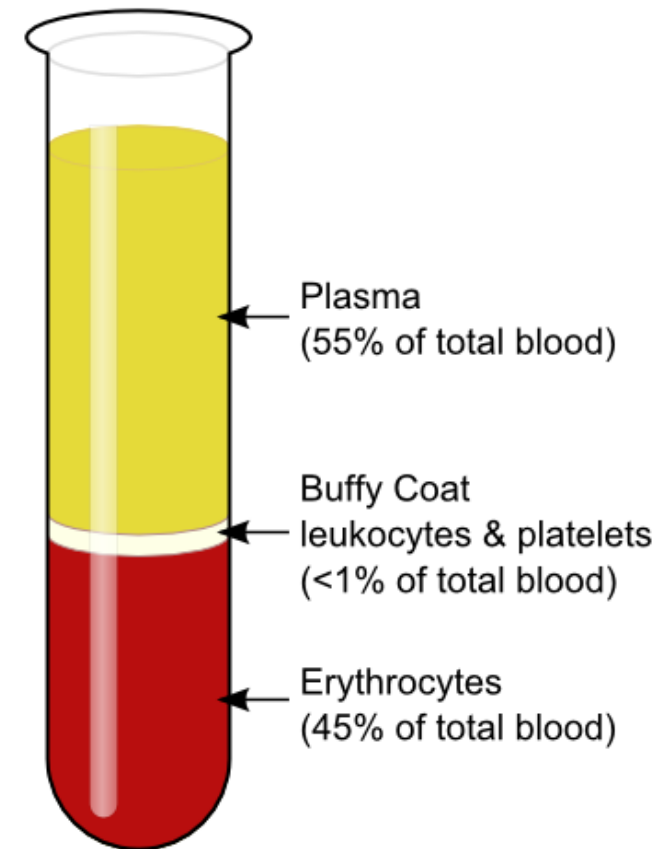
- 2 gatunki z rzędu skóroskrzydłych - lokot malajski i lotokot filipiński
- 2 gatunki z rzędu wiewióreczników - tupaja pospolita i tupaja mniejsza
- królik europejski z rzędu zajęczaków

Do ukorzenia drzewa filogenetycznego użyto królika.



Pobieranie próbek DNA, amplifikacja segmentów genów i analiza sekwencji

- DNA wyekstrahowano z próbek pełnej krwi, tzw. “buffy coat”, włosów i wymazu policzkowego.
- DNA amplifikowano w reakcji PCR przy pomocy 52 primerów, w tym zarówno używanych poprzednio jak i nowych, specjalnie zaprojektowanych primerów.
- Produkty PCR zostały zsekwencjonowane przy użyciu zestawu BigDye Terminator v1.1 cycle sequencing kit.
- Różne sekwencje dla każdego segmentu genu uliniowano przy użyciu programu MAFFT v6.
- Regiony niejednoznaczności uliniowania zidentyfikowano programem GBLOCK v. 0.91b i usunięto z dalszej analizy filogenetycznej.



Filogenetycznie pouczające pozycje (ang. phylogenetically informative sites)

- Pozycja jest filogenetycznie pouczająca w rozważanym zbiorze sekwencji jeśli występują na niej co najmniej dwa różne nukleotydy i każdy z nich występuje na tej pozycji w co najmniej 2 sekwencjach.
- Pozycje takie są pomocne przy rozstrzyganiu różnych hipotez o pokrewieństwie.

	Site								
Sequence	1	2	3	4	5	6	7	8	9
1	A	A	G	A	G	T	G	C	A
2	A	G	C	C	G	T	G	C	G
3	A	G	A	T	A	T	C	C	A
4	A	G	A	G	A	T	C	C	G
					*		*		*

Charakterystyka użytych sekwencji

Category	Sequence Sites (bp)	% Category (of 34, 927 bp)	Informative Sites (bp)	% Category (of 14, 683 bp)	% Total
Total	34927	100%	14683	100%	42.04%
Autosomes	27427	78.53%	11432	77.86%	41.69%
X Genes	4870	13.94%	1939	13.20%	39.81%
Y Genes	2630	7.53%	1312	8.94%	55.59%
Non-Coding	16371	46.87%	7878	53.63%	48.12%
Coding	14742	42.40%	5439	37.02%	36.39%
UTR	3814	10.92%	1366	9.30%	35.81%

Podział danych do konstrukcji drzew filogenetycznych

Skonstruowano drzewa filogenetyczne oddzielnie dla każdego z 6 zbiorów segmentów genów odpowiadających:

- chromosomom X i Y,
- autosomom,
- intronom,
- eksonom,
- segmentom UTR,

oraz dla połączonego zbioru wszystkich segmentów, w celu porównania zgodności topologii otrzymanych drzew.

Wybór modelu ewolucji sekwencji

- W celu wyboru najlepszego modelu ewolucji sekwencji użyto programu ModelTest v3.7 na uliniowionych sekwencjach dla każdego zbioru segmentów.
- Kryterium wyboru modelu: Akaike Information Criterion (AIC):
 $AIC = 2(-\log(\text{wiarygodność}) + \text{liczba parametrów modelu})$.
- Dla wszystkich zbiorów oprócz eksonów został wybrany model TVM+I+G. (jeden typ tranzyzji i 4 typy transwersji + niezmiennosc części pozycji + rozkład gamma tempa podstawień na pozostałych).

Data Partition	Model	Rate Matrix (AC,AG,AT,CG,CT,GT)	Gamma Shape Parameter	Nucleotide Frequency (A,G,C,T)	P-Invariant
Combined	TVM+I+G	1.0492 3.9254 0.6220 1.2887 3.9254 1.0000	0.8892	0.2901 0.2101 0.2137 0.2861	0.2359

- Dla eksonów został wybrany model TrN+I+G (1 typ tranzyzji i transwersji).

Konstrukcja drzew filogenetycznych

Drzewa filogenetyczne otrzymano przy użyciu heurystycznego przeszukiwania z różnymi kryteriami optymalności:

- największej wiarygodności (ML)
- największej parsymonii (MP)

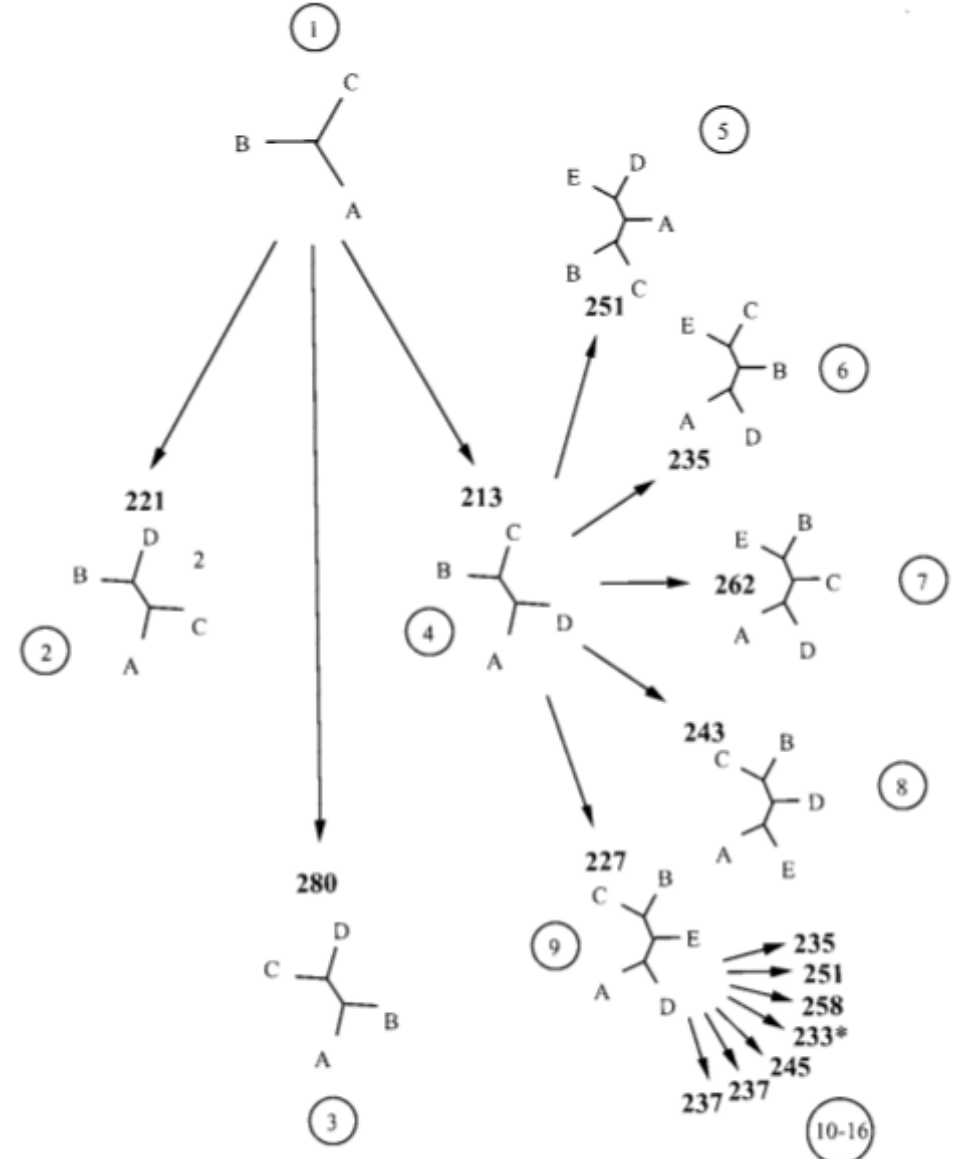
przy użyciu programów PAUP i GARLI.

Heurystyczne przeszukiwanie:

- Początkowe drzewo w obu metodach było tworzone metodą [stepwise addition](#).
- Następnie drzewa przekształcano metodą [tree-bisection-reconnection](#).

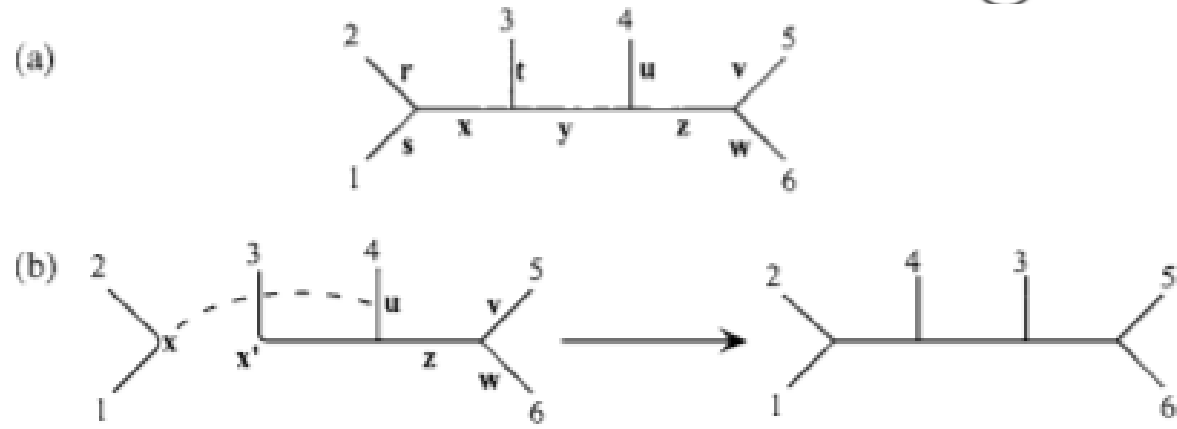
Schematy działania obu metod z książki

“The phylogenetic handbook”:



(a)

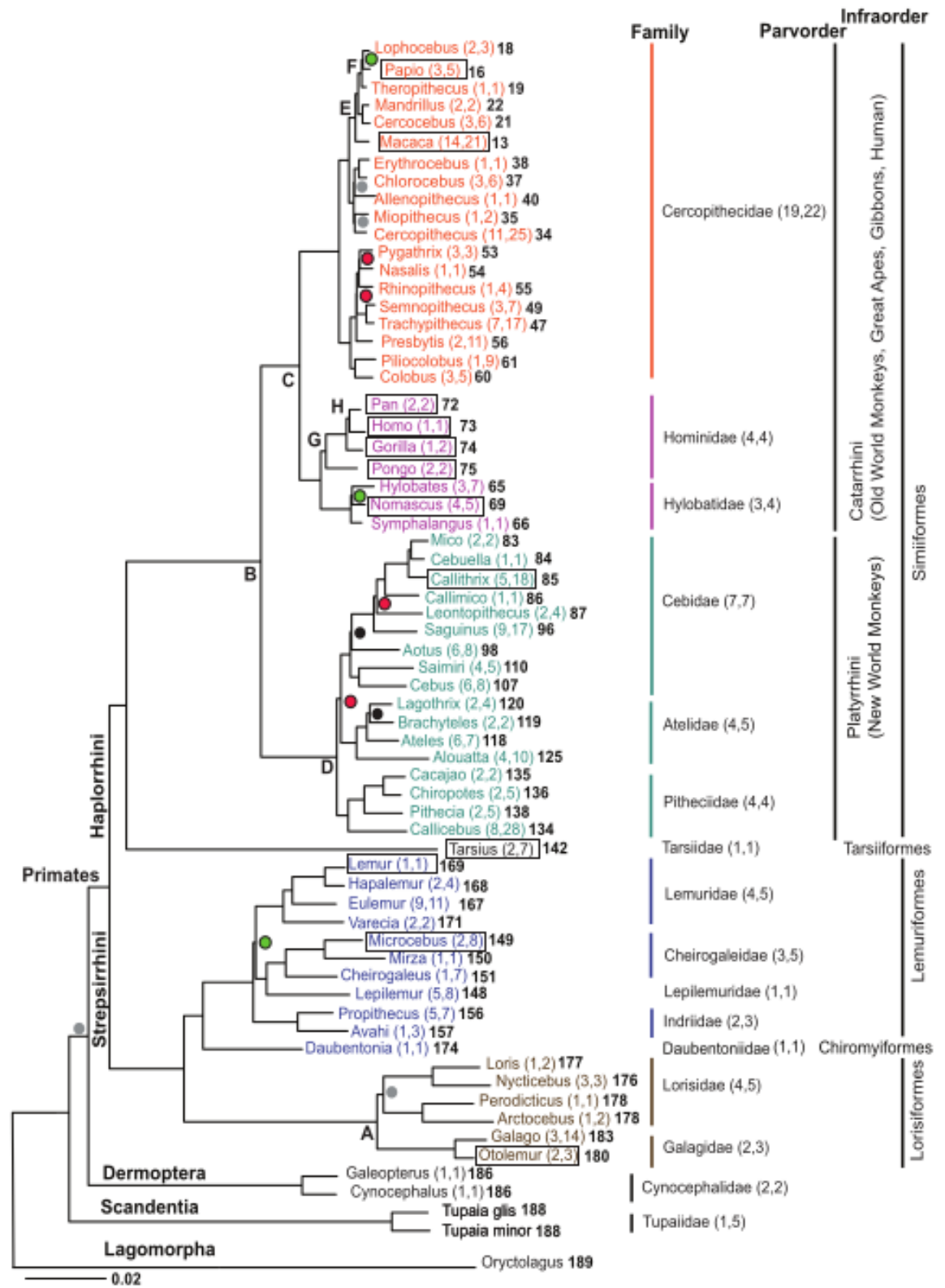
(b)



Ocena wiarygodności węzłów i porównanie wyników

Wiarygodność węzłów oceniano przy użyciu metody bootstrap, przy czym za rozstrzygnięte uznawano węzły z prawdopodobieństwem bootstrap > 50%.

- Drzewa uzyskane różnymi metodami i na różnych zbiorach segmentów genów są w dużej mierze zgodne.
- Najmniejsza liczba 3 nierozstrzygniętych węzłów wystąpiła kiedy użyto wszystkich segmentów. Ponadto, 166 z 189 węzłów miało prawdopodobieństwa bootstrap > 90%.



Analiza Bayesowska

Wykonano analizę bayesowską filogenezy i czasów dywergencji przy użyciu programu BEAST v1.5.3.

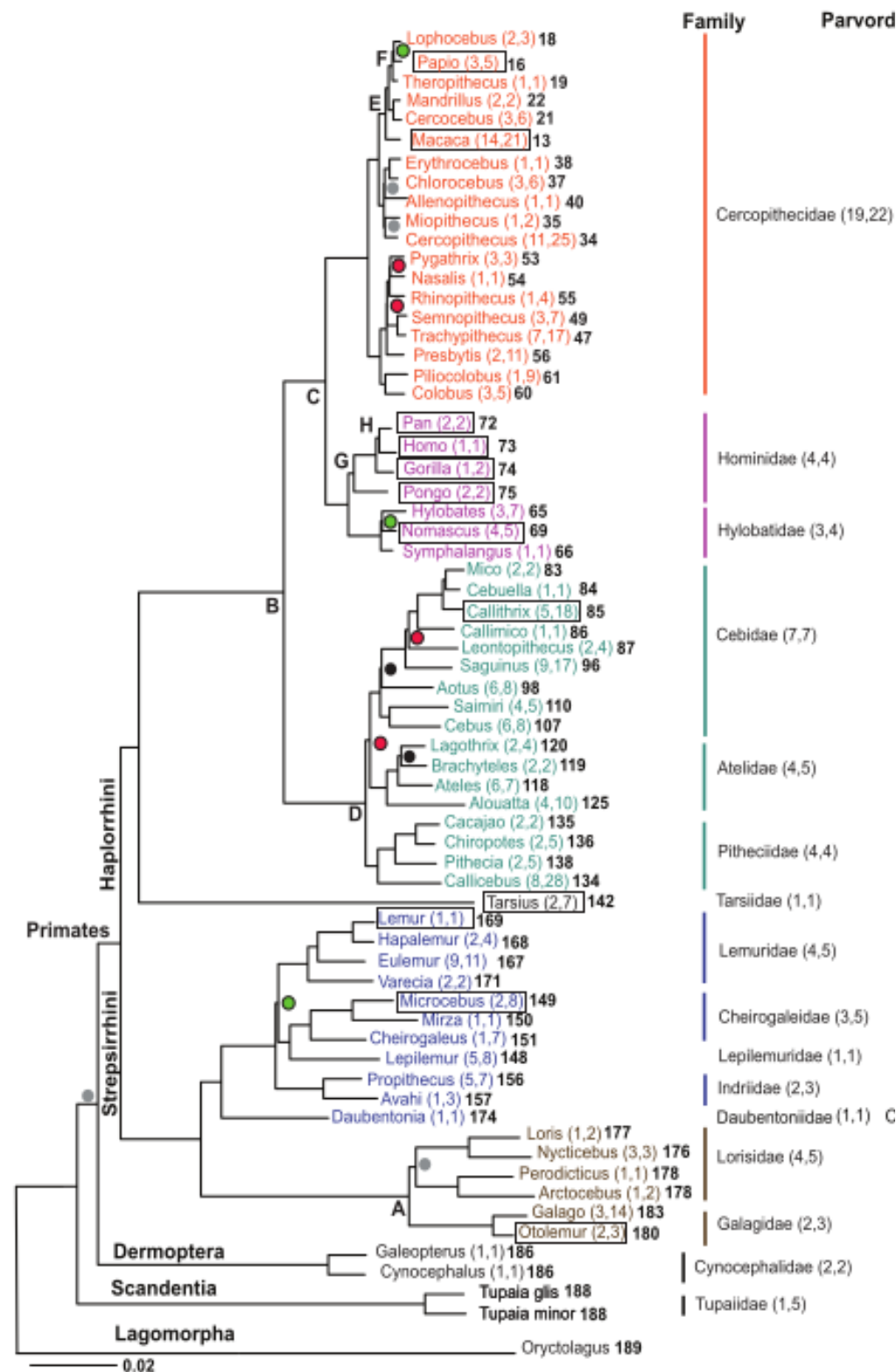
- Z powodu ograniczeń dostępnej mocy obliczeniowej, podzielono zbiór gatunków na 5 podzbiorów i w każdym wykonywano analizę oddzielnie.
- W modelach ewolucji sekwencji przyjęto jednostajne rozkłady a priori dla temp podstawień (0, 100), parametru kształtu rozkładu gamma (0, 100), i proporcji niezmiennych pozycji (0, 1).
- Przyjęto lognormalny nieskorelowany rozluźniony zegar molekularny z jednostajnym rozkładem a priori średniego tempa podstawień (0, 100) i jego odchylenia standardowego (0,10).
- Zastosowano proces Yule'a jako rozkład a priori drzew filogenetycznych oraz użyto metody UPGMA do konstrukcji początkowego drzewa dla metody MCMC.

Punkty kalibracyjne

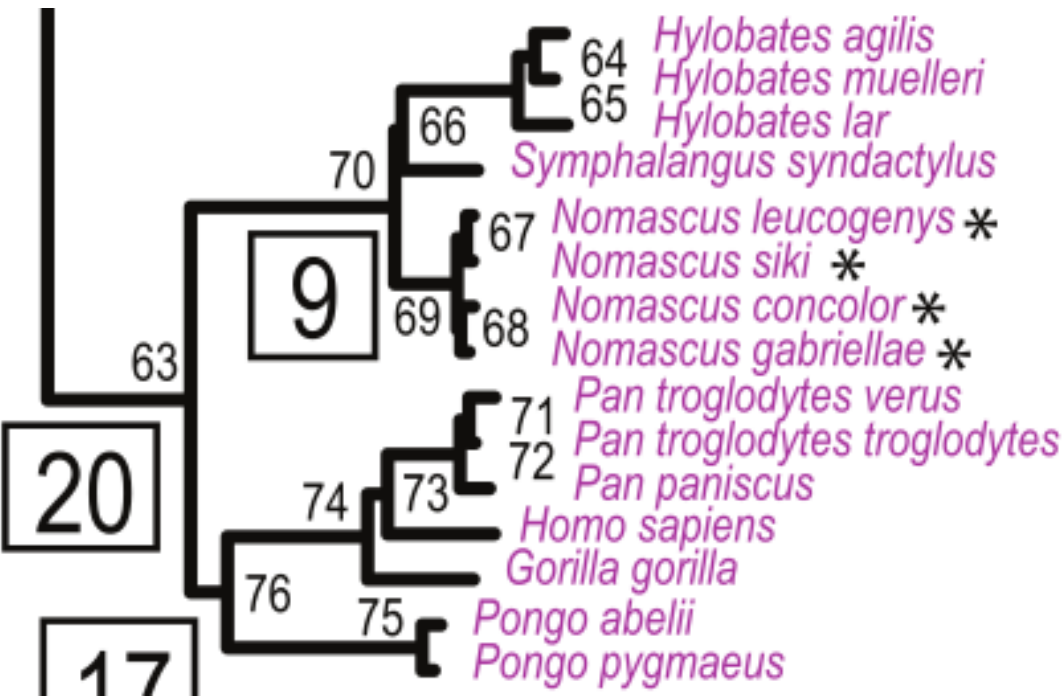
Żeby otrzymać rozkłady a posteriori czasów dywergencji użyto 9 punktów kalibracyjnych w postaci normalnych rozkładów a priori dla węzłów A-H.

Przyjęto też normalny rozkład a priori ze średnią 90.0 milionów lat temu i odchyleniem standardowym 6 milionów lat dla najniedawniejszego przodka wszystkich naczelnych.

Wybrane punkty kalibracyjne opierają się na datach znanych skamieniałości.



Mały człekokształtne



Node Fig. 2	Date (MY)	Date 95% HPD		Rate	Rate 95% HPD		Branch Length (ML)*	Branch Length (MP)*	Indels per Branch*	Node Support (ML)	Node Support (MP)	Posterior Probability
63	20.32	16.59	24.22	4.61	2.59	7.12	54.22	190	4	100	100	1.00
64	1.86	0.98	2.85	4.94	2.22	8.14	6.34	19	0	79	78	1.00
65	3.26	1.95	4.54	8.70	5.39	12.55	43.89	137	1	100	100	1.00
66	Note C	-	-	NoteC	-	-	2.78	Note B	0	100	Note A	Note C
67	0.56	0.23	0.94	4.78	1.99	7.85	3.50	7	0	98	96	1.00
68	0.92	0.42	1.47	4.37	1.69	7.69	1.34	8	0	64	75	1.00
69	1.34	0.80	1.98	3.48	2.25	4.88	23.66	12	2	100	100	1.00
70	8.93	6.58	11.52	6.73	4.73	9.04	77.38	235	3	100	100	1.00
71	1.24	0.66	1.90	4.47	1.67	7.53	3.43	19	0	100	97	1.00
72	2.17	1.28	3.21	6.14	4.03	8.55	27.46	91	2	100	100	1.00
73	6.60	5.40	7.96	4.87	2.02	8.20	7.57	27	0	100	99	1.00
74	8.30	6.58	10.07	6.54	4.21	9.35	53.10	177	3	100	100	1.00
75	1.31	0.61	2.20	4.80	3.70	6.08	73.78	231	1	100	100	1.00
76	16.52	13.45	19.68	4.18	1.84	7.34	13.96	61	3	100	100	1.00

Rozstrzygnięcie pewnych spornych wczesnych zdarzeń z ewolucji naczelnych

- Skóroskrzydłe są rzędem ssaków najbliższym naczelnym.
- Wyrakokształtne są siostrzanym taksonem małpokształtnych i tworzą wraz z nimi wyższe naczelne.
- Palczakowate i lemurokształtne tworzą takson monofiletyczny, co sugeruje że ich wspólny przodek raz skolonizował Madagaskar.
- Została wyjaśniona filogeneza rodzin małp Nowego Świata: sakowatych, czepiakowatych i płaksowatych.
- Najniedawniejszy przodek naczelnych żył około 87 milionów lat temu prawdopodobnie w Azji.



Wyrak filipiński



Palczak madagaskarski



Lemur katta

A Mitogenomic Phylogeny of Living Primates. Użyte dane molekularne.

- Autorzy zsekwencjonowali 32 nowe genomy mitochondrialne.
- 13 nowych genomów mitochondrialnych należy do małp Nowego Świata a 10 do lemurowych.
- Z bazy GenBank pobrano dodatkowe 51 mitochondrialnych sekwencji genomowych naczelnych oraz 4 nienaczelne użyte jako grupa zewnętrzna - tupaja północna, mysz domowa, szczur wędrowny i dziobak australijski.
- Zbiór danych reprezentuje wszystkie 16 rodzin naczelnych, 57 z 78 rodzajów, oraz 78 z 480 gatunków.

Metody analizy

- Uliniowane sekwencje do rekonstrukcji filogenetyzy przygotowano na 4 sposoby (m. in. opcjonalnie zastępując nukleotydy C i T przez Y ze względu na duże różnice w zawartości C/T u różnych gatunków, usuwając insercje/delecje/słabo uliniowane fragmenty, czy ograniczając się do obszarów kodujących białka).
- Drzewa filogenetyczne konstruowano metodą maximum likelihood (ML) i bayesowską korzystając z programów GARLI i MrBayes.
- Model ewolucji sekwencji dla każdego zbioru sekwencji wybierano w oparciu o [Bayesian information criterion](#).

$$\text{BIC} = -2 \log(\text{wiarygodnosc}) + K \log(n),$$

gdzie K – liczba parametrów modelu , n – długość uliniowienia.

- Czasy dywergencji szacowano metodą bayesowską z lognormalnym nieskorelowanym rozluźnionym zegarem molekularnym oraz punktami kalibracyjnym z wykorzystaniem programu BEAST, analogicznie jak w pierwszej pracy.

Wyniki

- Użycie metody ML i bayesowskiej oraz różnych zbiorów danych dało takie same drzewa filogenetyczne.
- Wiarygodność węzłów obliczona metodą bootstrap dla metody ML lub jako prawdopodobieństwo a posteriori dla metody bayesowskiej była zazwyczaj niższa dla zbiorów danych używających typów nukleotydów AGY niż AGTC.
- Na części wspólnej drzewo filogenetyczne jest w dużej mierze zgodne z tym z wcześniejszej pracy, ale występują pewne rozbieżności. Niektóre z nich można wyjaśnić hybrydyzacją różnych gatunków.
- W porównaniu do wcześniejszych prac otrzymano stosunkowo późną datę wystąpienia najniedawniejszego wspólnego przodka wszystkich naczelnych od 66 do 69 mln lat temu, co sugeruje że dywergencja istniejących naczelnych rozpoczęła się blisko granicy kreda-trzeciorzęd.