

# PHYLOGENETIC RECONSTRUCTION

Topik Khusus 4: Bioinformatika dan AI

Kelompok Cytosine

# LATAR BELAKANG

Rekonstruksi filogenetik adalah metode untuk memetakan hubungan evolusi antar spesies menggunakan data genetik, morfologi, atau molekuler guna memahami pola diversifikasi dan kekerabatan organisme berdasarkan nenek moyang bersama. Pendekatan ini menggunakan algoritma seperti **Maximum Parsimony**, **Maximum Likelihood**, dan **Bayesian Inference** untuk mengidentifikasi pohon evolusi yang paling sesuai dengan data. Teknologi seperti Next Generation Sequencing (NGS) serta perangkat lunak RAxML, MrBayes, dan ASTRAL telah meningkatkan kemampuan untuk merekonstruksi filogeni secara akurat, dengan prospek ke depan semakin mengandalkan data genomik lengkap dan model yang dapat menangani variasi intra-spesies dan fenomena evolusi kompleks lainnya.

# RUMUSAN MASALAH

01

Apa saja metode populer yang digunakan untuk inferensi pohon filogenetik, dan bagaimana prinsip dasar dari masing-masing metode tersebut?

02

Bagaimana metode komputasional lanjutan seperti concatenation dan coalescence phylogeny dapat digunakan untuk menganalisis dataset filogenetik yang lebih kompleks?

03

Bagaimana implementasi berbagai metode inferensi pohon filogenetik dapat dilakukan dalam lingkungan bahasa pemrograman R dan python?

04

Bagaimana evaluasi/interpretasi terhadap hasil pohon filogenetik dapat dilakukan?

# TUJUAN

01

Menjelaskan berbagai metode populer yang digunakan dalam inferensi pohon filogenetik, termasuk metode berbasis jarak, Maximum Parsimony (MP), Maximum Likelihood (ML), dan Bayesian Inference (BI).

02

Menguraikan metode komputasional lanjutan seperti concatenation dan coalescence phylogeny dalam inferensi pohon filogenetik.

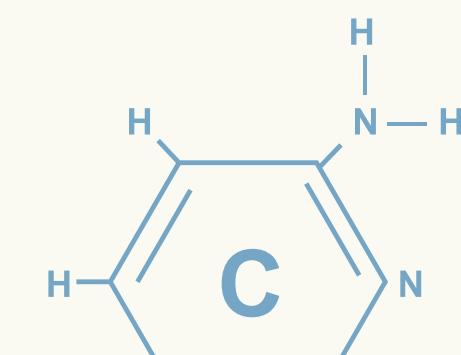
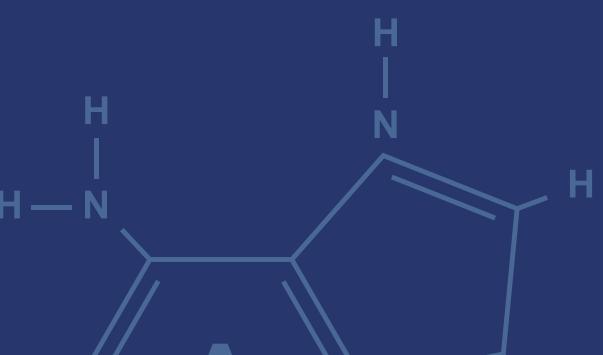
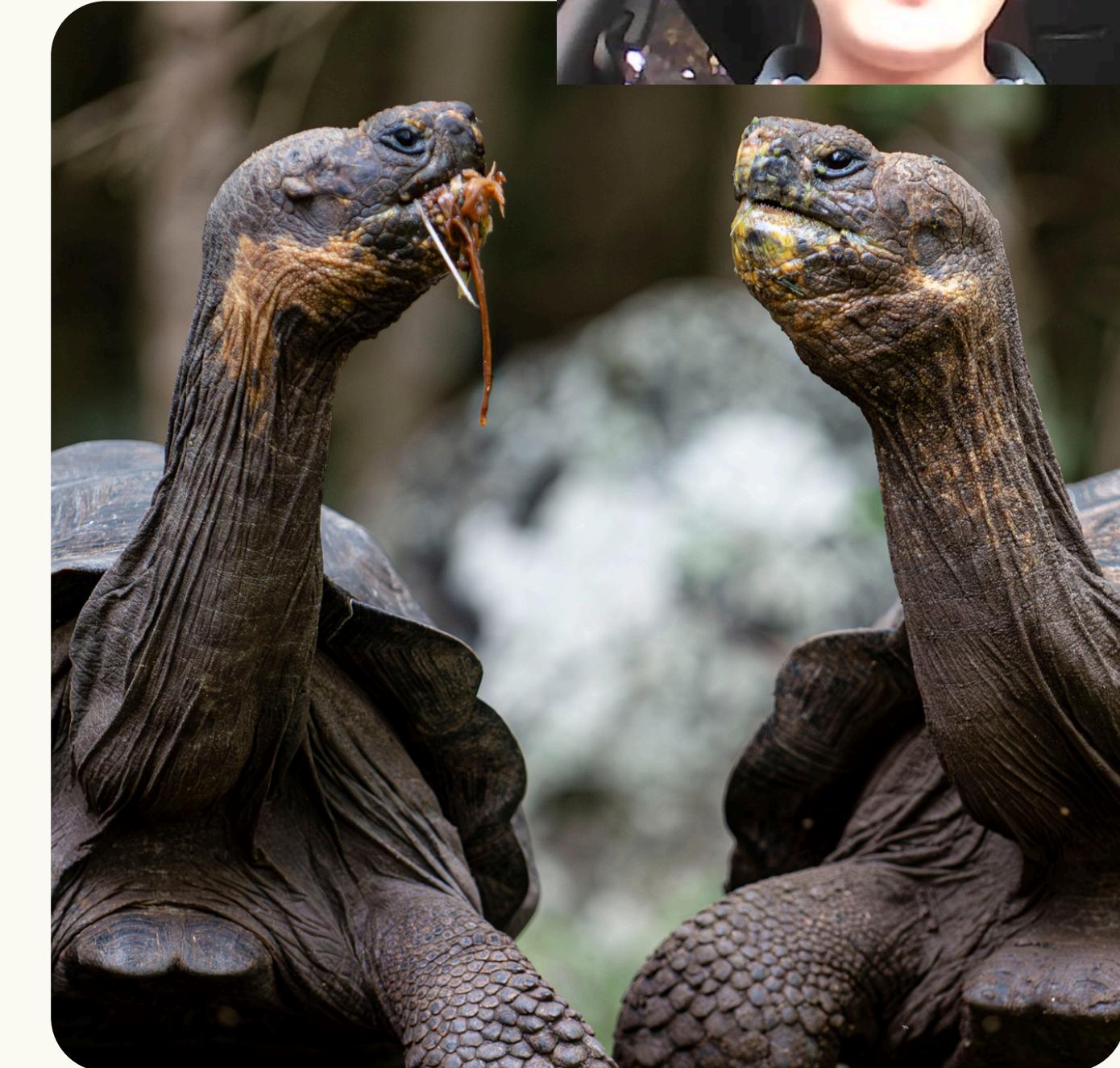
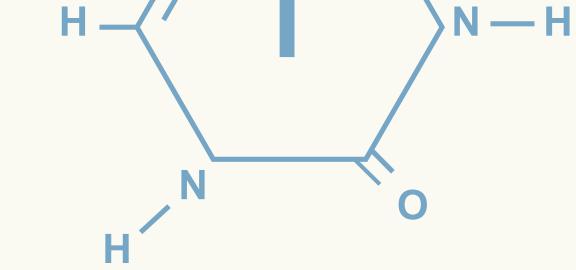
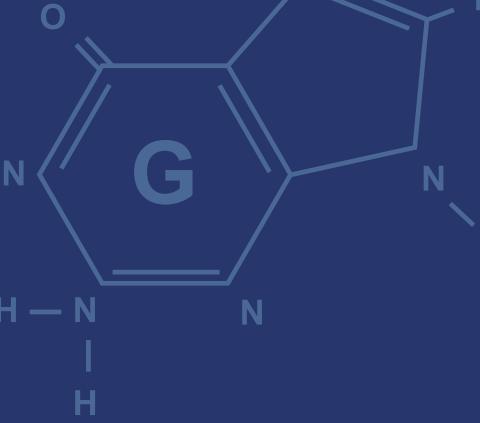
03

Mengimplementasikan metode inferensi pohon filogenetik dalam lingkungan bahasa pemrograman R dan python, dengan menggunakan paket-paket yang relevan untuk melakukan analisis data filogenetik.

04

Mengevaluasi atau melakukan interpretasi terhadap hasil pohon filogenetik.

# PHYLOGENETIC RECONSTRUCTION





# APA ITU PHYLOGENETIC RECONSTRUCTION?

*Phylogenetic reconstruction* adalah proses untuk mengidentifikasi hubungan evolusi antar spesies atau kelompok organisme.

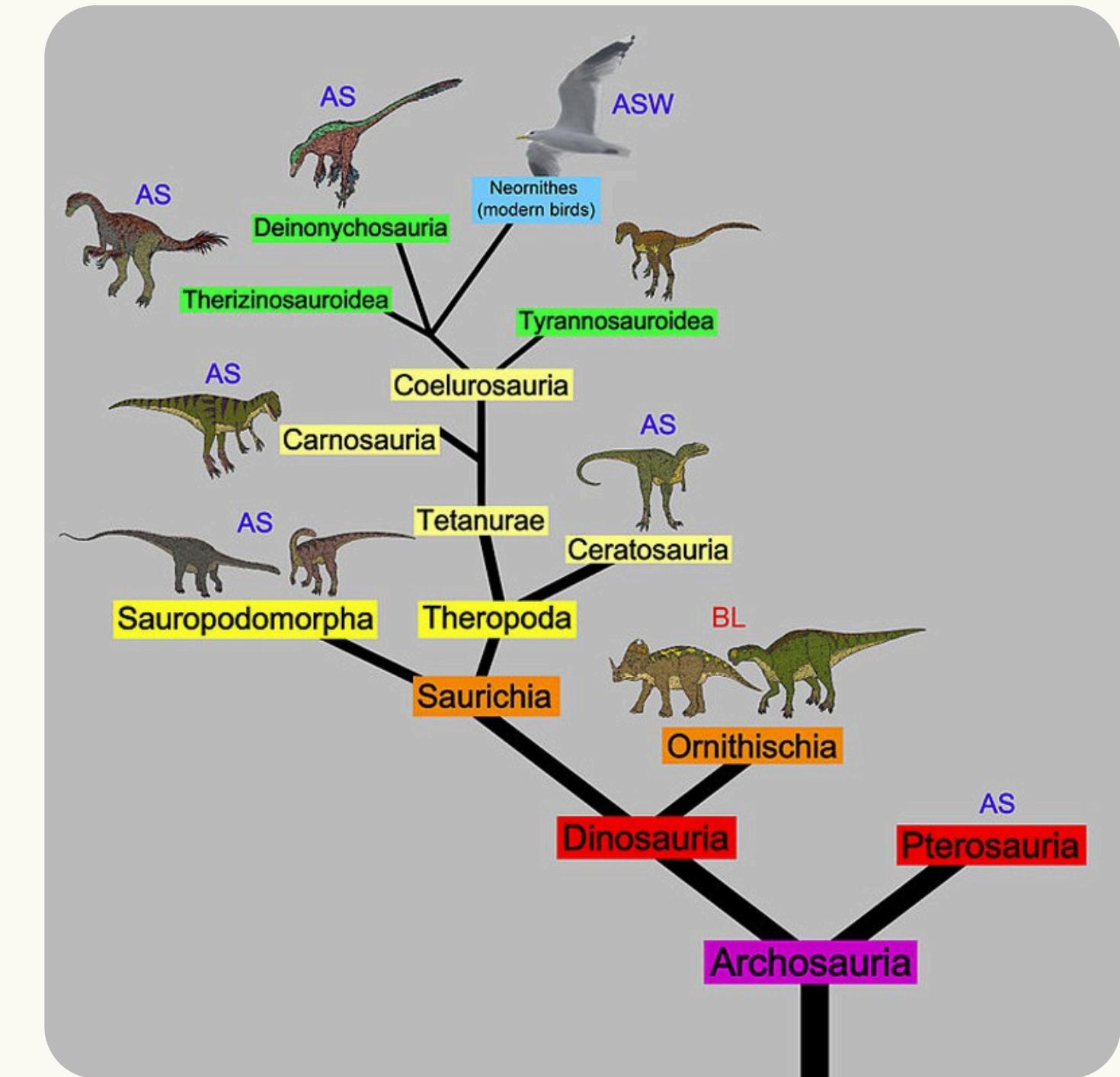
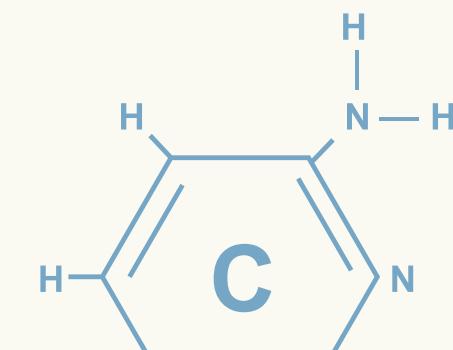
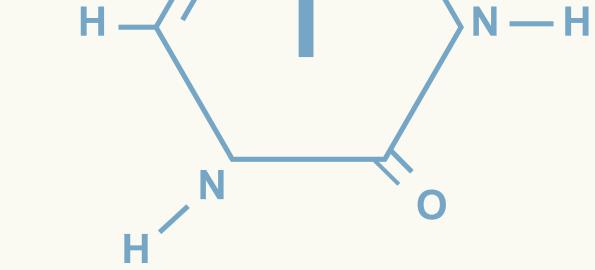
Proses ini melibatkan penggunaan data genetik, morfologis, atau karakteristik lainnya untuk membangun pohon filogenetik yang menggambarkan bagaimana spesies berhubungan satu sama lain berdasarkan nenek moyang bersama.



## ALASAN

- 
1. Klasifikasi taksonomi menjadi lebih mudah dilakukan dengan memahami hubungan evolusi
  2. Memberikan wawasan tentang bagaimana spesies beradaptasi dan berkembang seiring waktu
  3. Penemuan pola evolusi dalam patogen atau tanaman dapat membantu dalam pengembangan vaksin dan varietas tanaman yang lebih baik
  4. Mengidentifikasi spesies yang lebih rentan, sehingga upaya konservasi dapat lebih difokuskan untuk melindungi keanekaragaman hayati

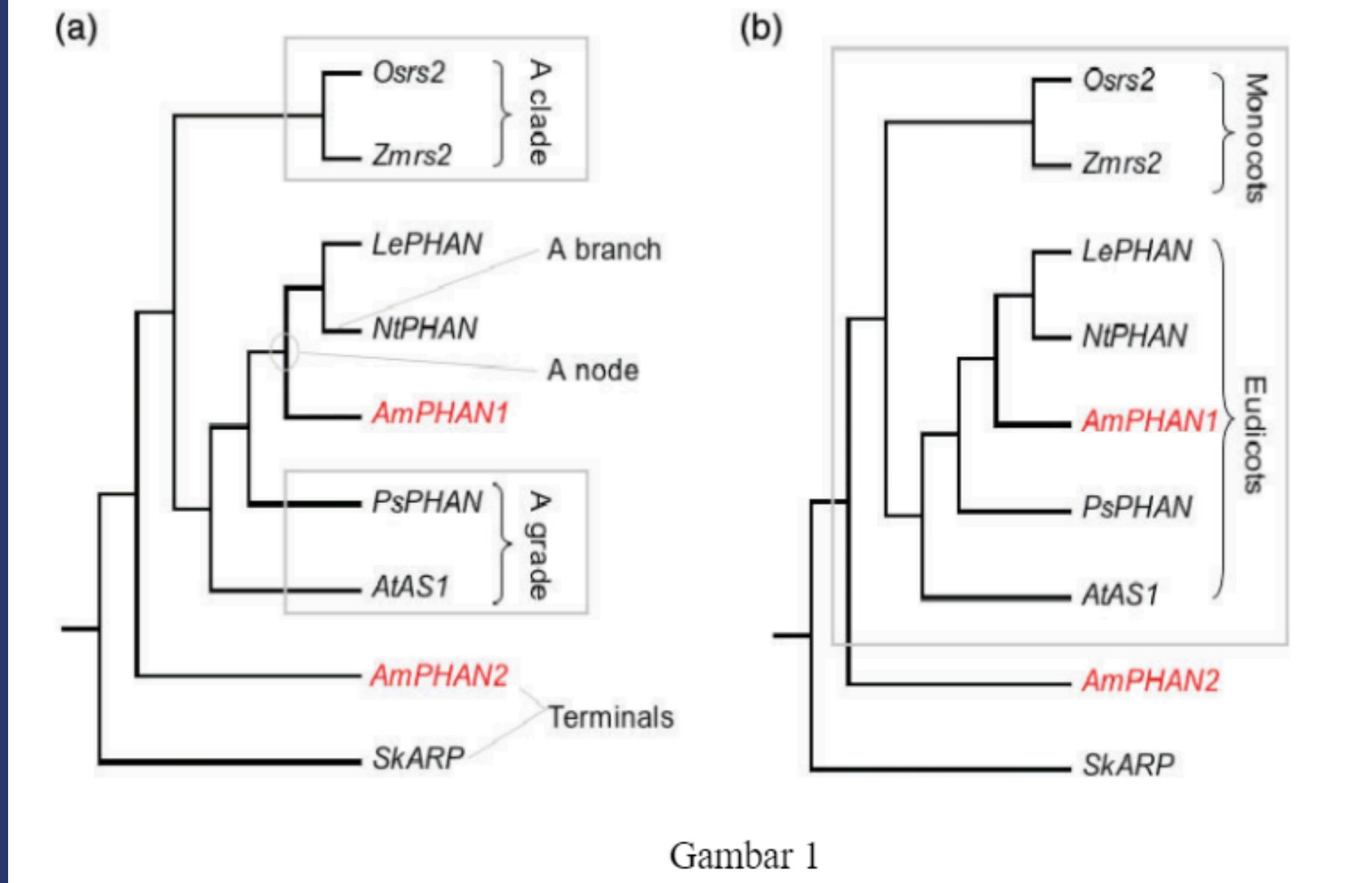
# PHYLOGENETIC TREE



**Pohon filogenetik** adalah representasi grafis dari hubungan evolusi antar organisme, yang bertujuan untuk menggambarkan dan memvisualisasikan bagaimana spesies berevolusi dari nenek moyang yang sama serta bagaimana mereka saling terkait.

Tiga jenis hubungan utama (Hennig, 1966):

1. **Monofili:** ancestor dan semua keturunannya
2. **Parafili:** ancestor dan sebagian dari keturunannya
3. **Polifili:** organisme yang tidak terkait tetapi memiliki karakteristik yang serupa

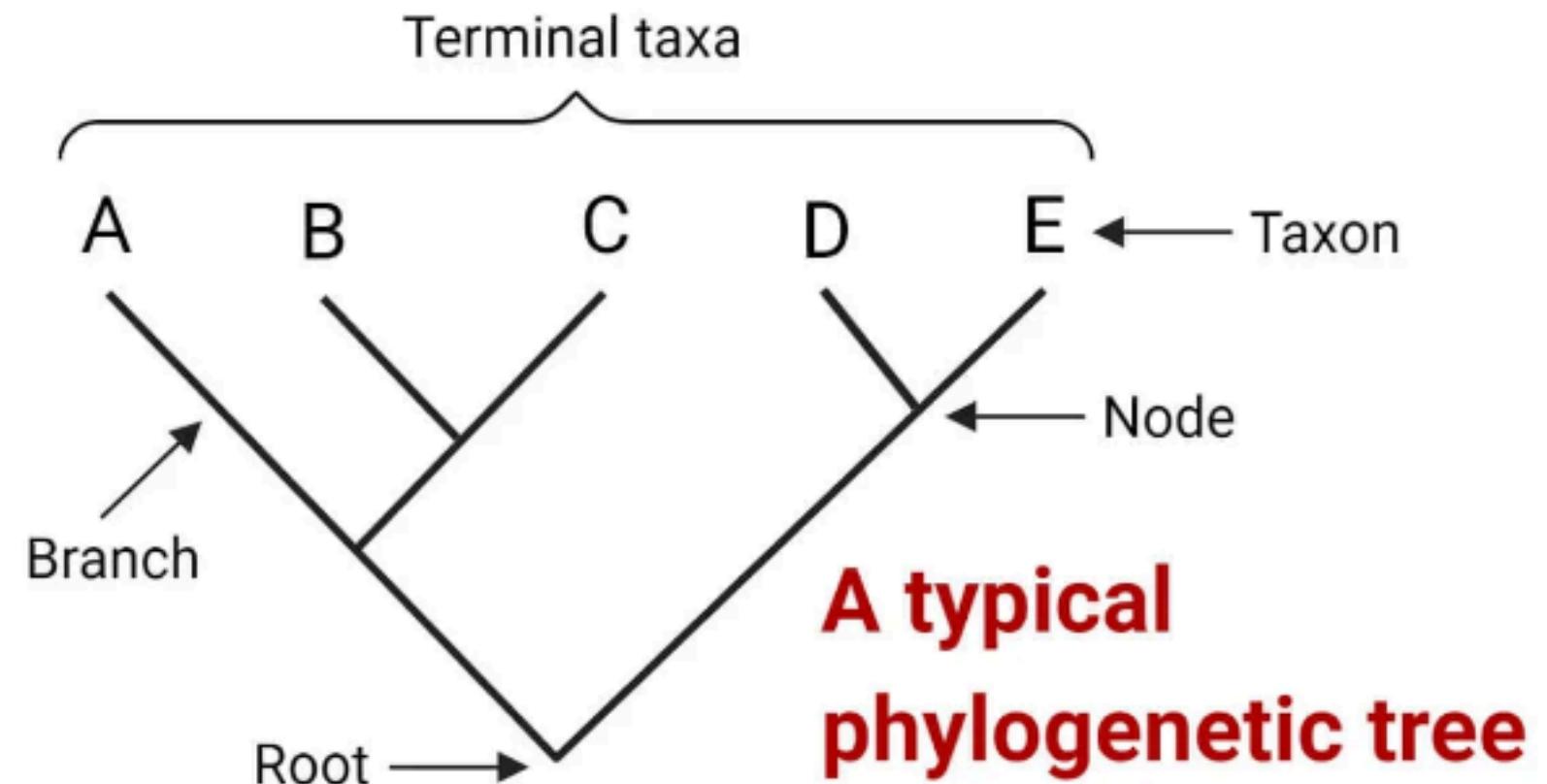


Gambar 1

Hubungan ini direpresentasikan sebagai diagram percabangan, atau pohon, dengan cabang-cabang yang dihubungkan oleh simpul dan mengarah ke terminal di ujung pohon

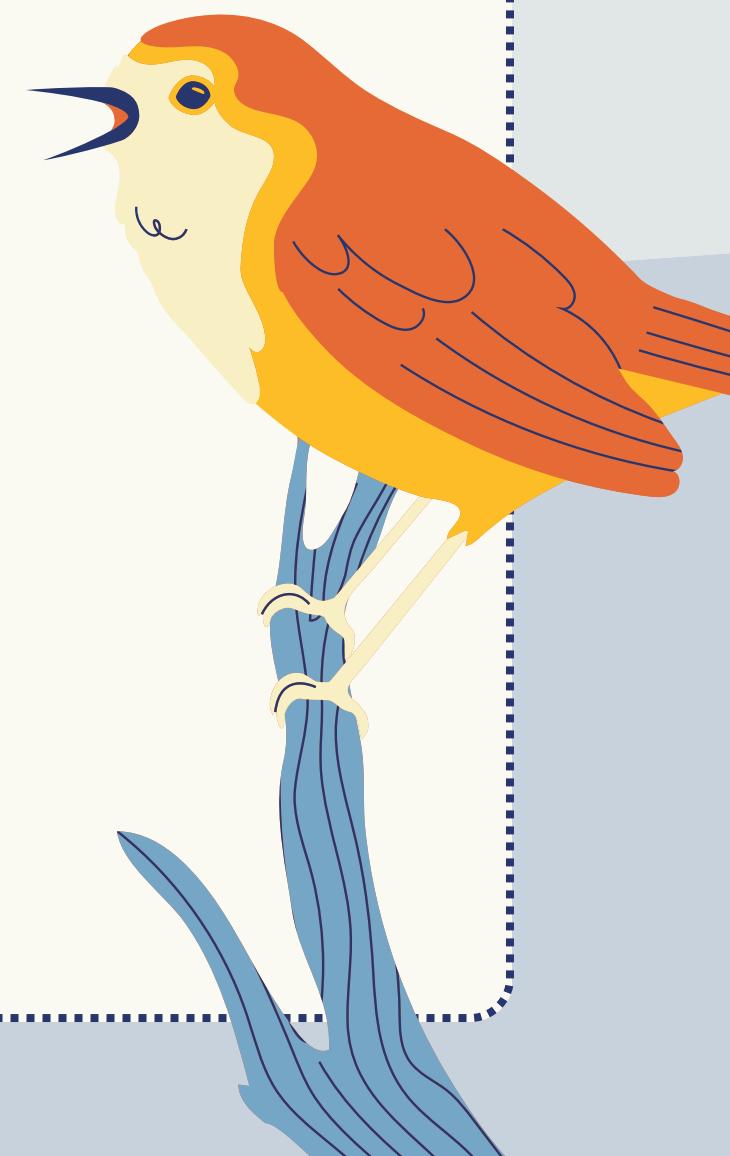
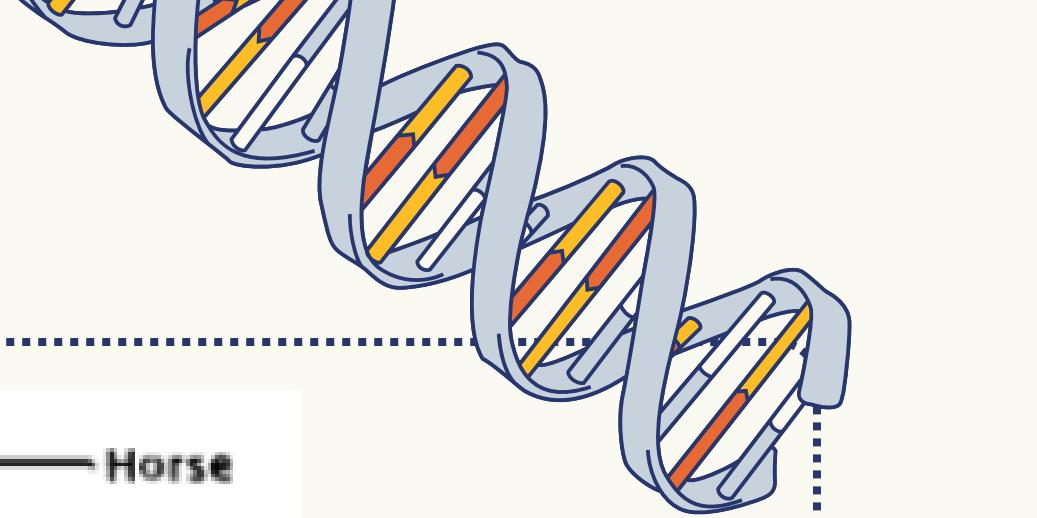
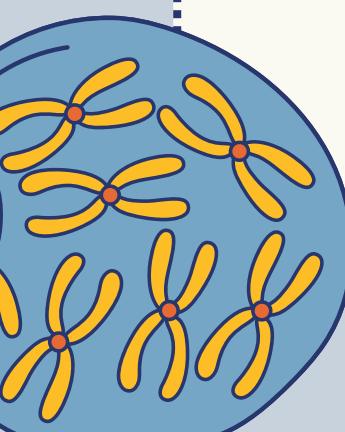
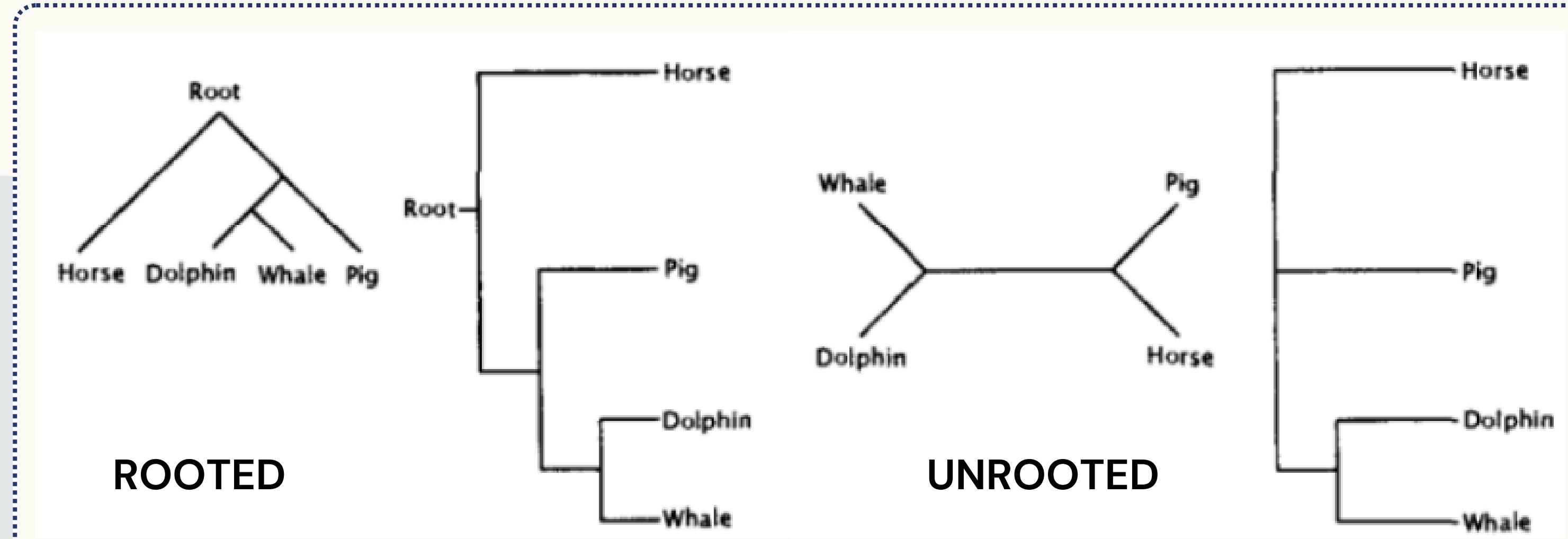
# STRUKTUR POHON FILOGENETIK

- 1. Node:** titik di mana satu cabang pohon bercabang menjadi dua atau lebih, yang mewakili titik divergensi di mana spesies baru muncul dari nenek moyang bersama
- 2. Tip:** ujung cabang yang mewakili spesies atau kelompok organisme yang ada saat ini, dengan setiap tip menunjukkan spesies hidup atau punah yang telah dianalisis
- 3. Root:** titik asal dari mana semua cabang bercabang, menunjukkan nenek moyang tertua dari semua spesies dalam pohon dan memberikan konteks temporal pada evolusi



Gambar 2

# STRUKTUR POHON FILOGENETIK



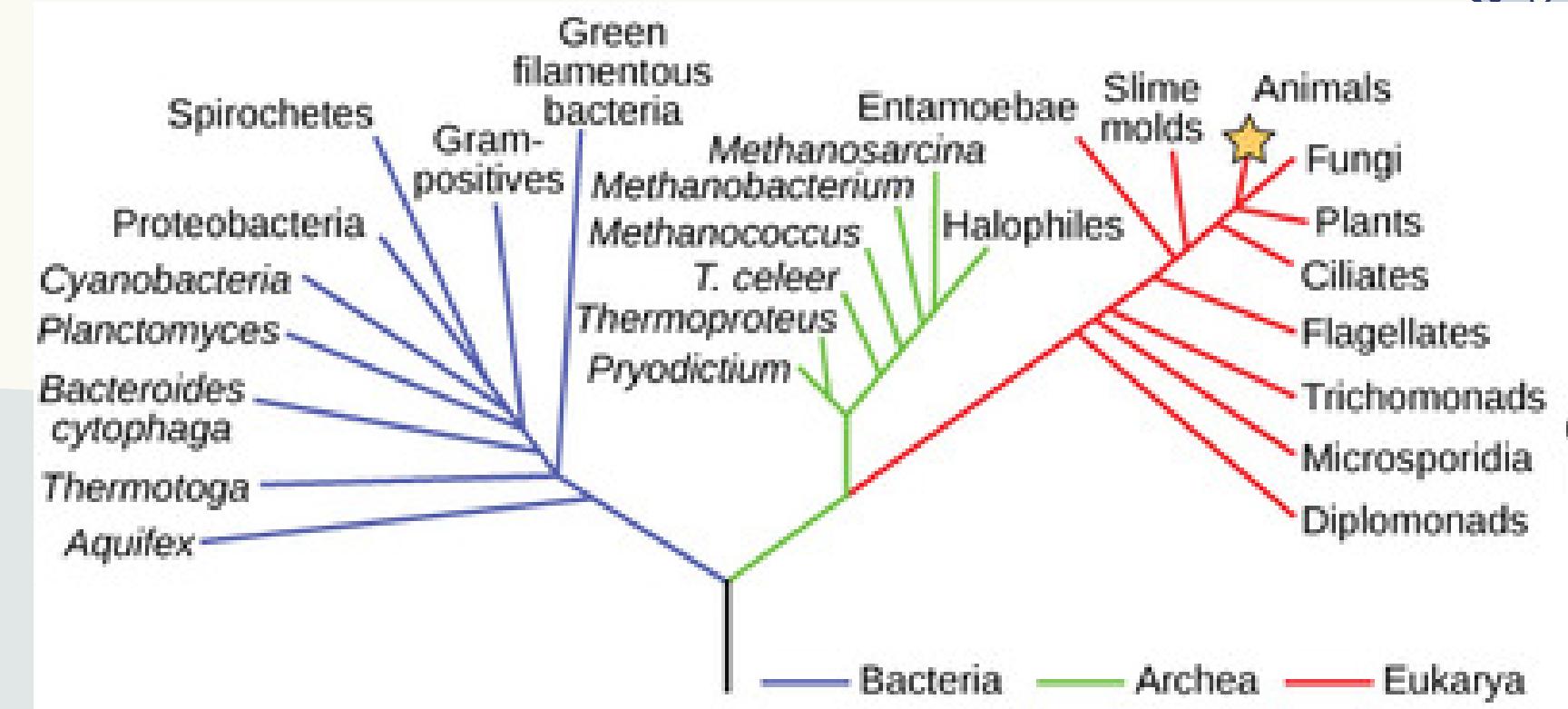
# STRUKTUR POHON FILOGENETIK

## Rooted Tree

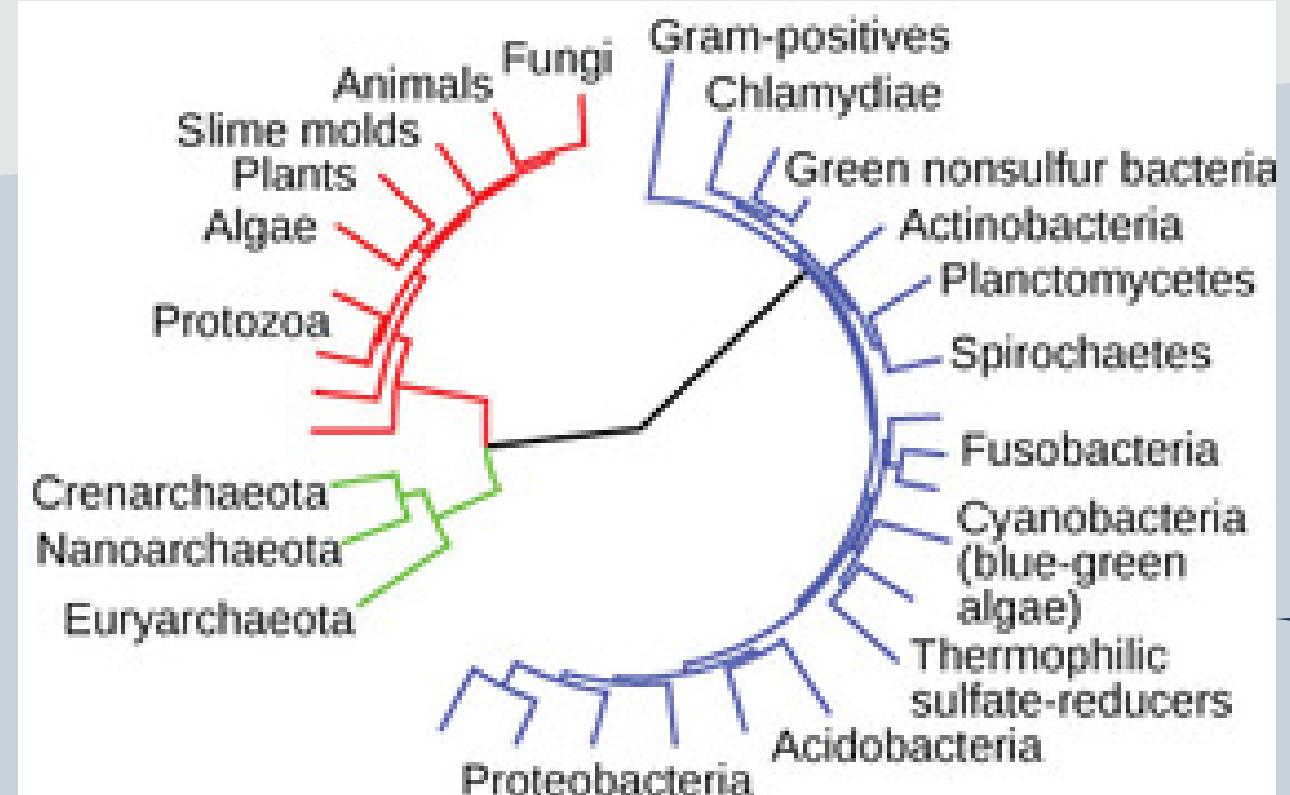
- Memiliki titik akar yang mewakili nenek moyang bersama.
- Memberi konteks temporal yang jelas,

## Unrooted Tree

- Menggambarkan hubungan evolusi antar spesies tanpa menentukan nenek moyang paling awal atau arah evolusi.
- Hanya menunjukkan hubungan relatif antar spesies tanpa memberikan urutan temporal atau asal evolusi.



(a) Rooted phylogenetic tree



(b) Unrooted phylogenetic tree

# JENIS POHON FILOGENETIK

## Pohon Taksonomi

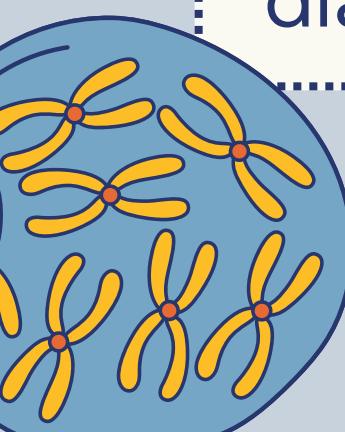
mewakili klasifikasi taksonomi dari organisme berdasarkan karakteristik morfologis dan genetik, sering digunakan dalam sistemmatika untuk mengidentifikasi dan mendeskripsikan spesies

## Pohon Genetik

menggambarkan hubungan antar spesies berdasarkan data genetik, sering kali dari analisis sekuens DNA, memberikan informasi lebih mendetail tentang variasi genetik di antara spesies

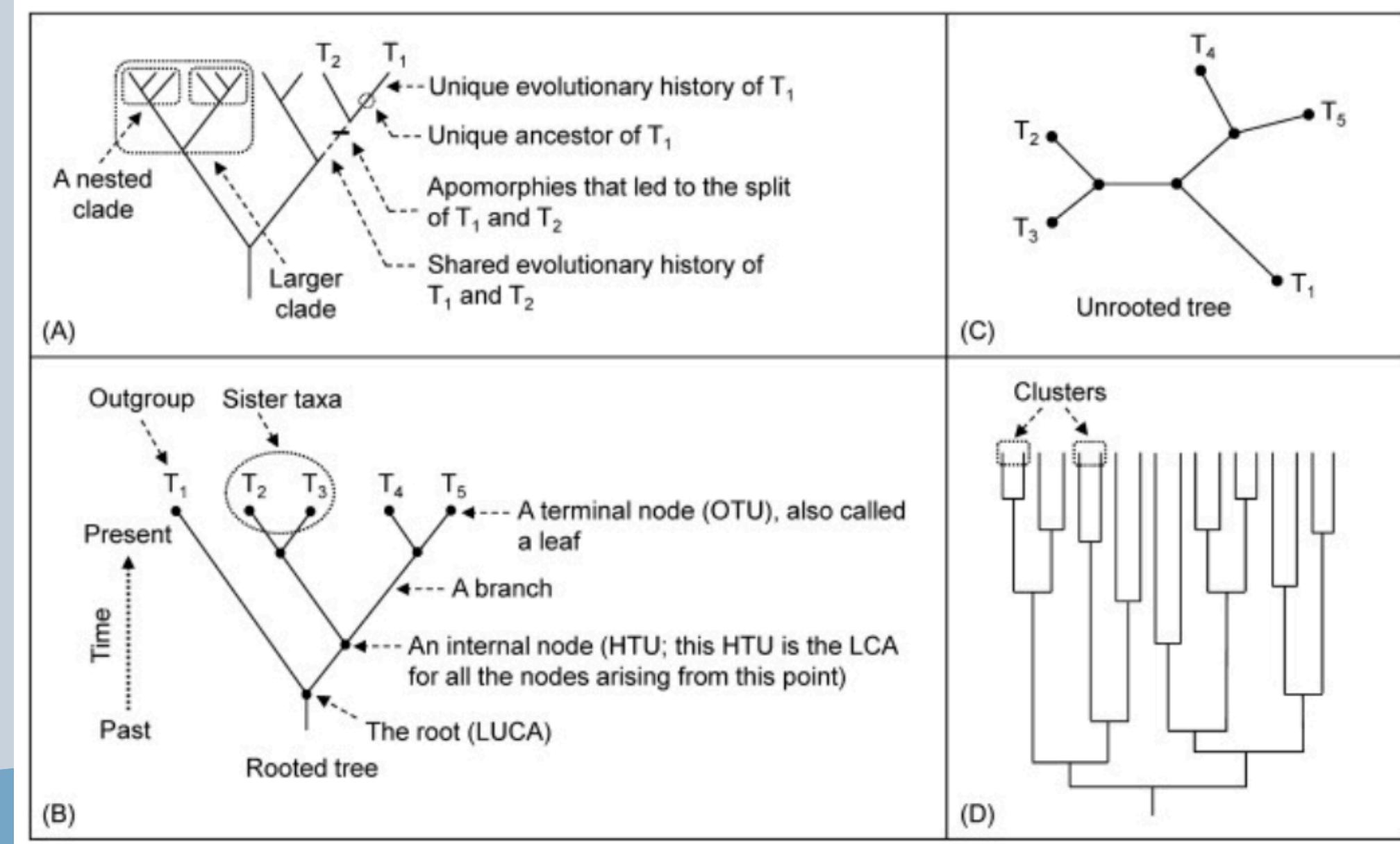
## Pohon Morfologis

didasarkan pada karakter morfologis yang terlihat, seperti bentuk, ukuran, dan struktur fisik organisme, berguna dalam konteks fosil atau spesies yang tidak dapat dianalisis secara genetik



# POHON FILOGENETIK

Cara membaca pohon filogenetik:

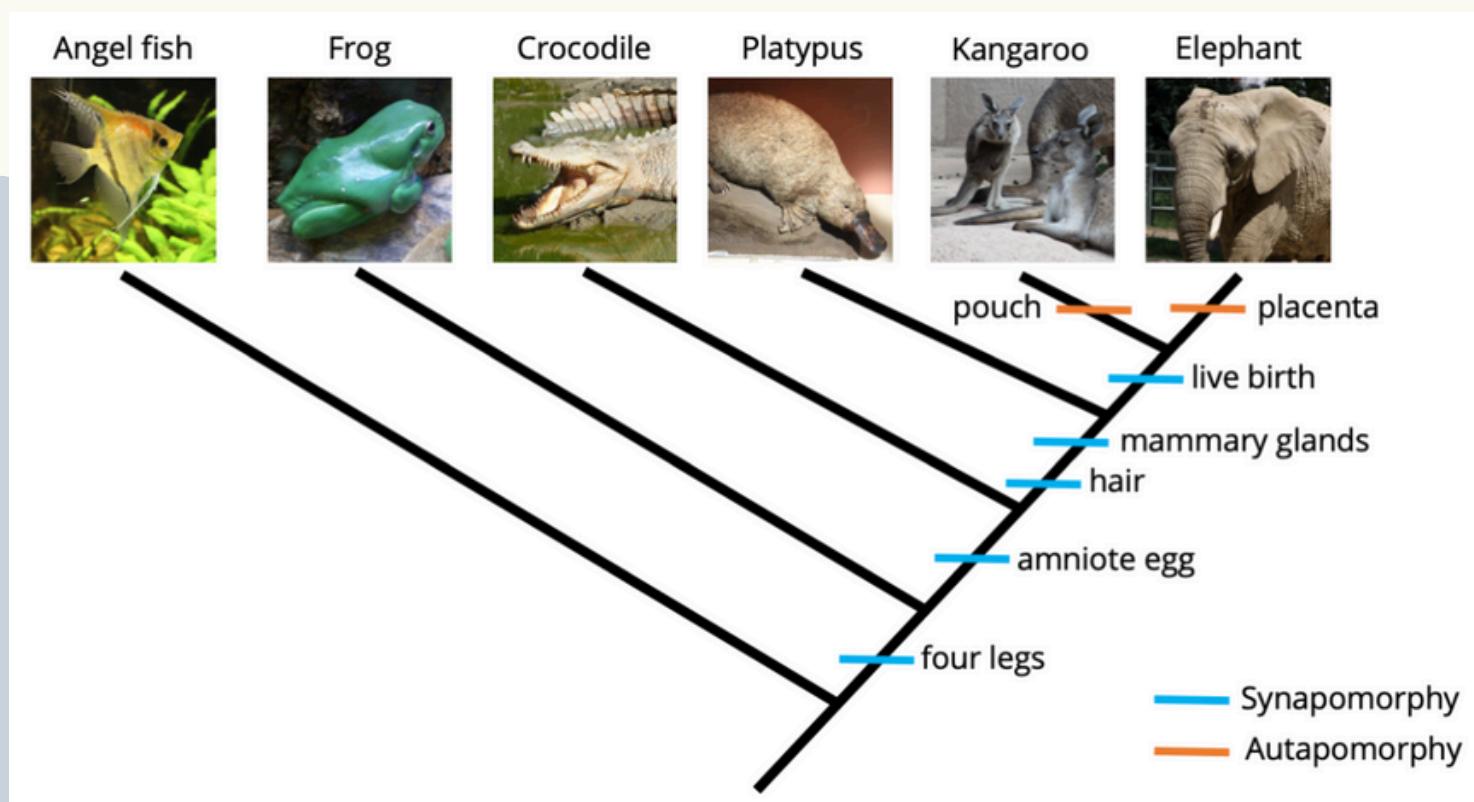


Phylogenetic Tree

Sumber: Supratim Choudhuri, in Bioinformatics for Beginners, 2014



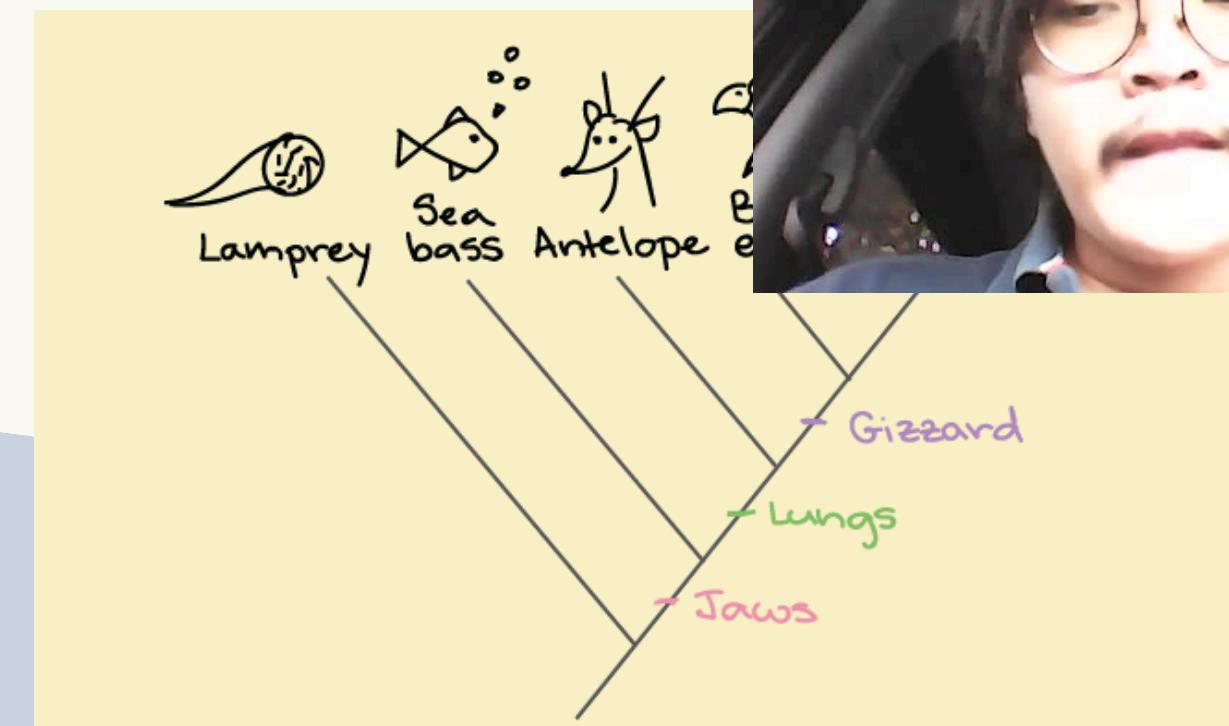
## Pohon 1



Sumber: Jonathan R. Hendricks is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

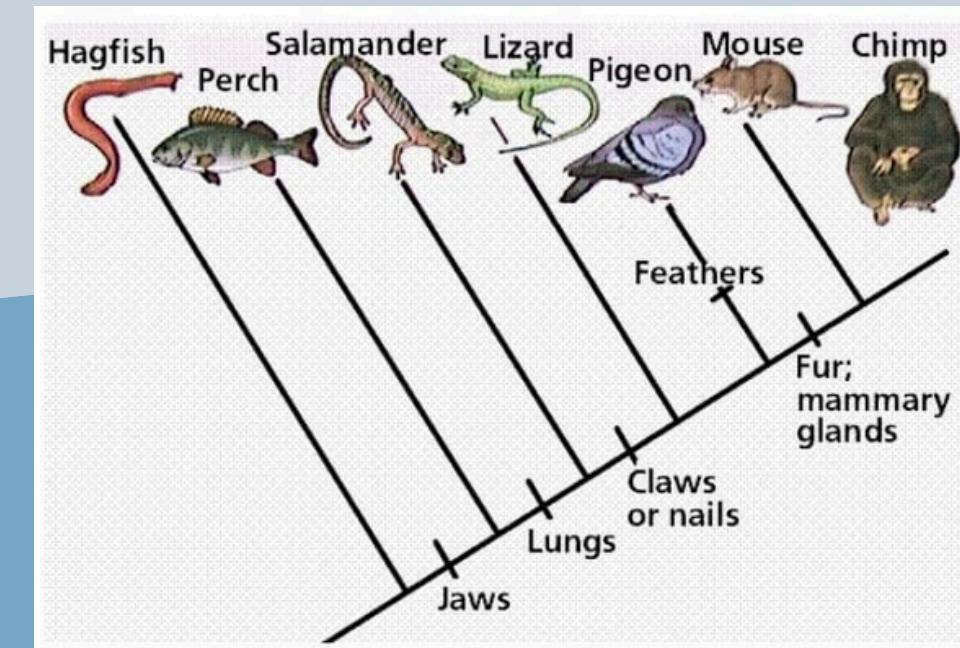
Menunjukkan hubungan antara spesies ingroup dengan ***synapomorphies*** yang digambarkan di setiap cabang. Synapomorphies ini menunjukkan karakter turunan, seperti "empat kaki" yang dimiliki oleh katak, buaya, platipus, kanguru, dan gajah. Karakter khusus seperti "rambut" dan "kelenjar susu" ada pada mamalia, sedangkan kanguru dan gajah berbagi synapomorphy "melahirkan".

## Pohon 2



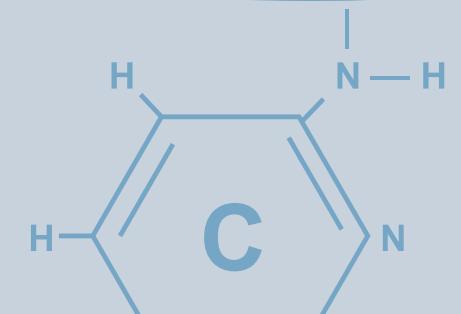
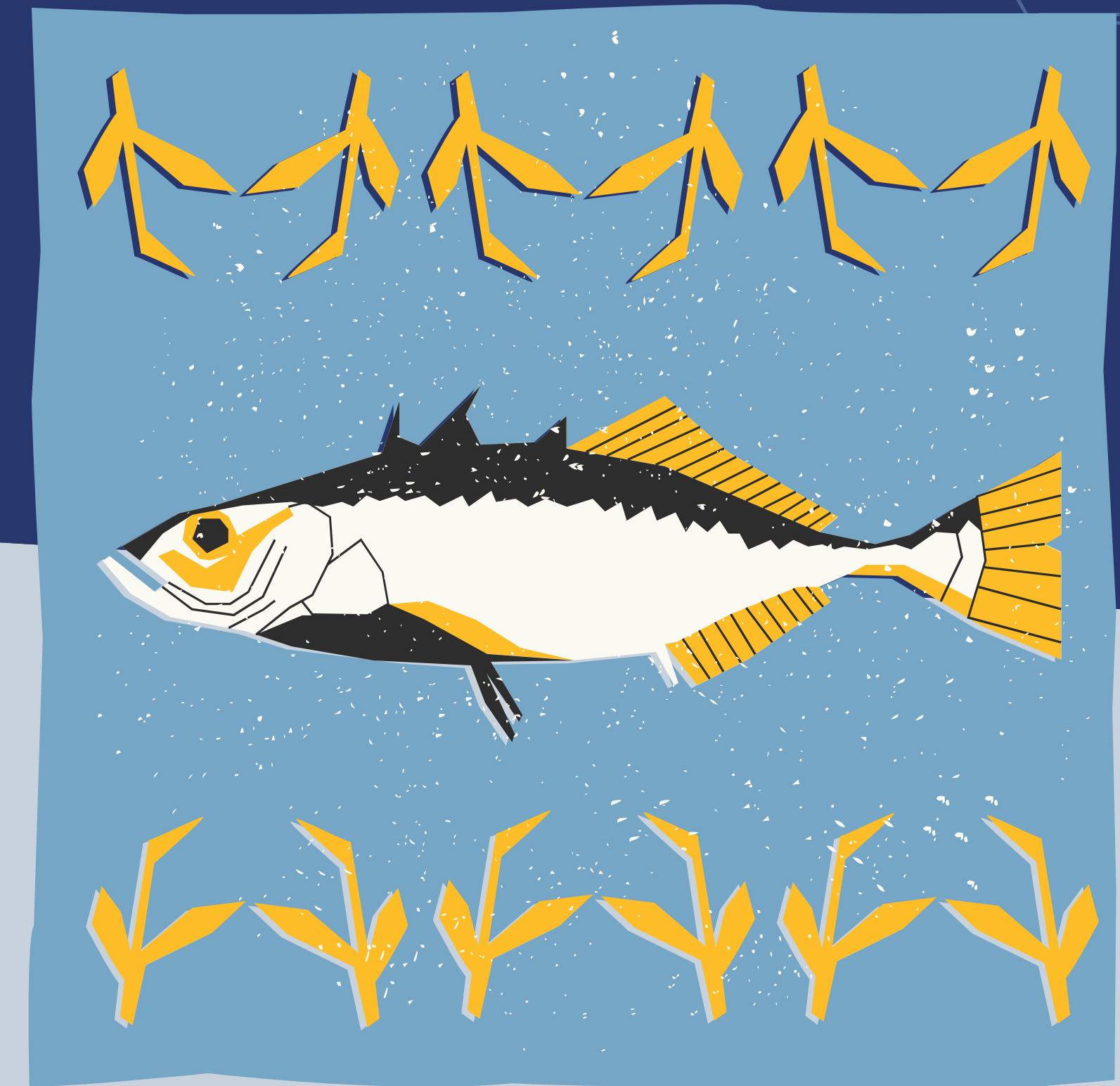
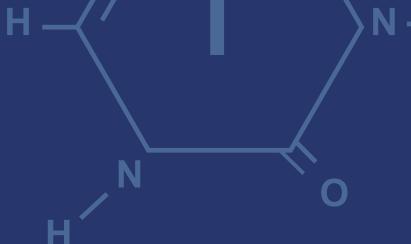
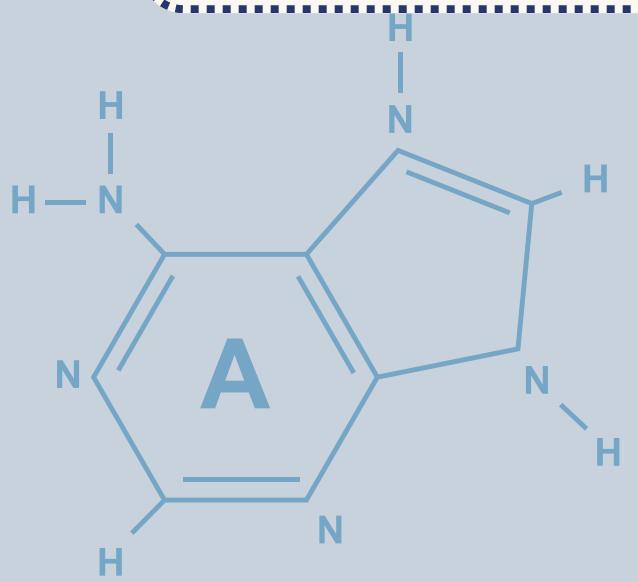
Sumber: Building a phylogenetic tree – Khan Academy

## Pohon 3

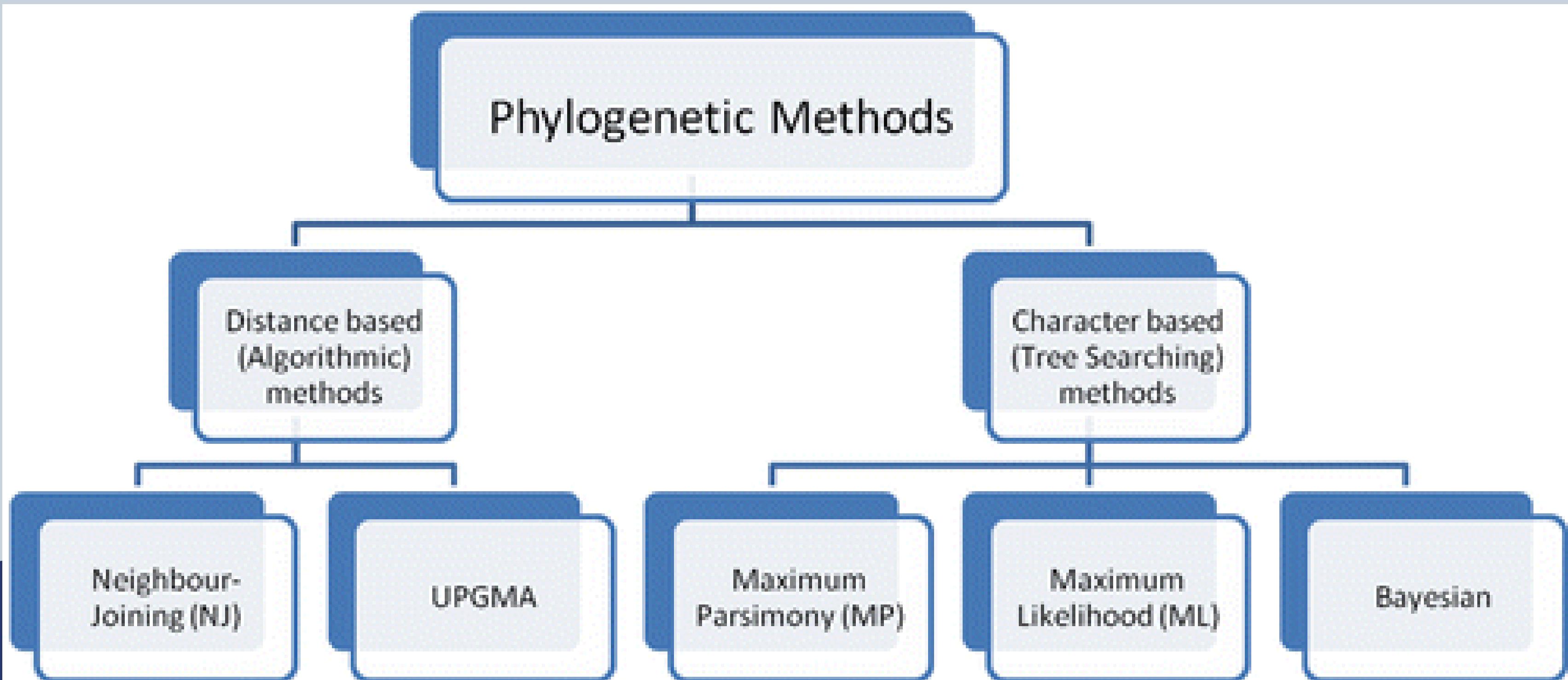


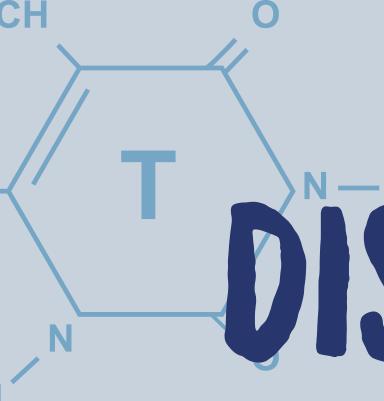
Sumber: <https://www.edrawsoft.com/article/what-is-phylogenetic-tree.html>

# METODE POPULER UNTUK MEMBANGUN PHYLOGENETIC TREES



# PHYLOGENETIC TREE METHODS

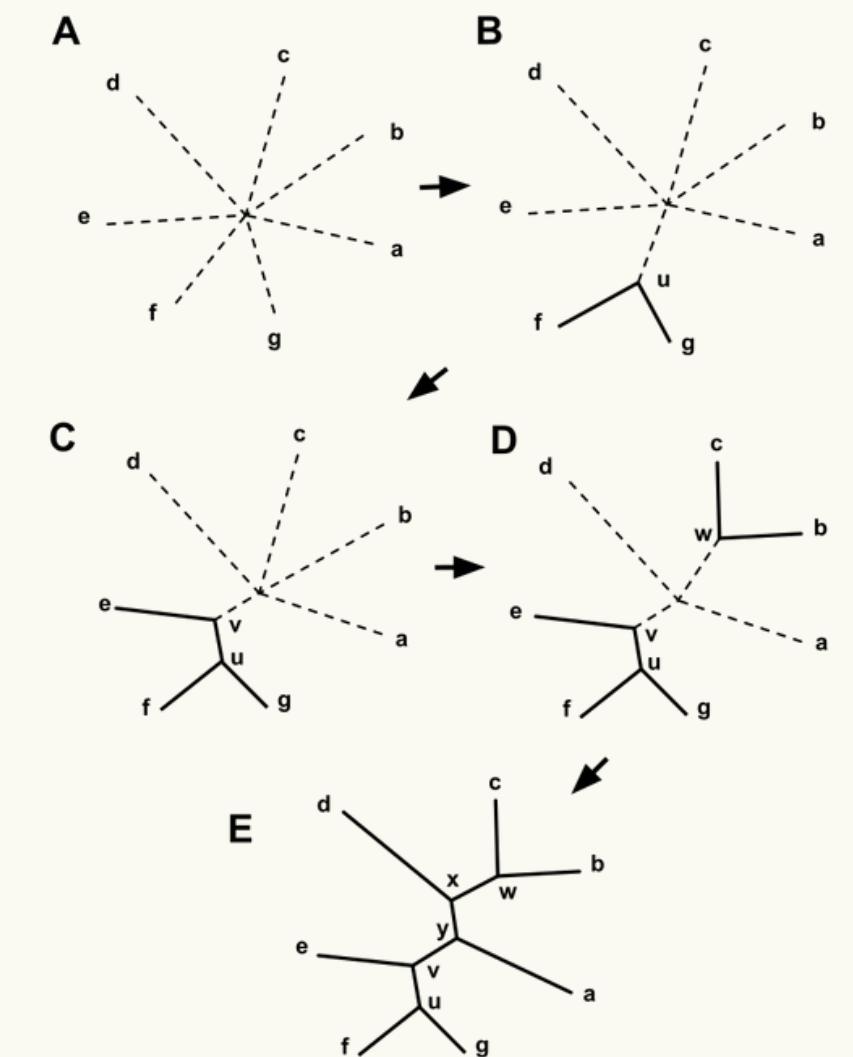
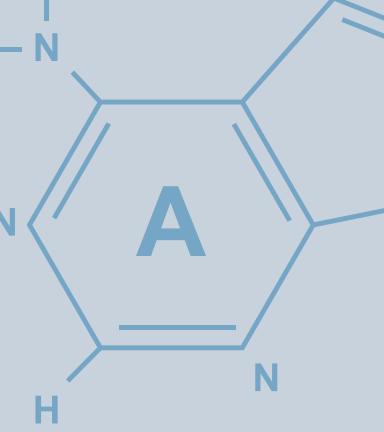




# DISTANCE BASED METHOD

Metode mengubah matriks fitur molekuler menjadi matriks jarak, lalu mengelompokkan spesies berdasarkan jarak tersebut. Contoh metode ini adalah neighbor-joining (NJ) dan UPGMA. NJ, diperkenalkan pada 1987, adalah algoritma pengelompokan yang menyusun pohon berdasarkan matriks jarak awal.

Metode NJ lebih cepat dan lebih akurat dalam merekonstruksi pohon filogenetik, terutama untuk kumpulan data besar, namun bisa kehilangan informasi ketika divergensi sekuens besar.



# CONTOH UPGMA

Akan kita coba buat pohon filogeni menggunakan metode distance-based (UPGMA). Misal akan kita cari pohon filogeni dari empat hewan berikut:

Antilope cervicapra: ATGTTCGTTA

Rupicapra sp. mtE1: TTTCGCGCTGA

Capra hircus: CCCTGGTTTG

Ovis aries: GATGCCAACCC

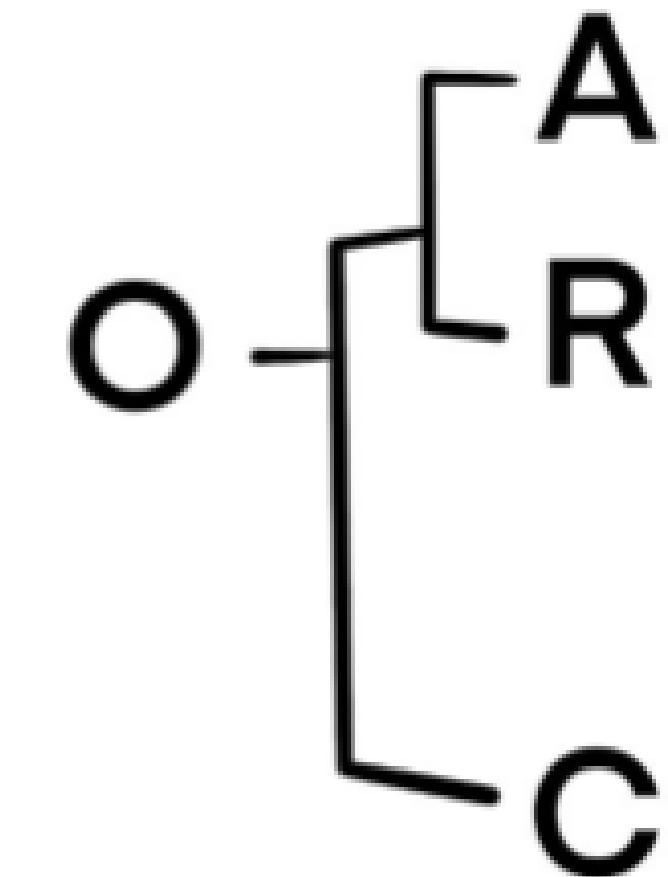
Masukkan nilai mismatch ke dalam matriks

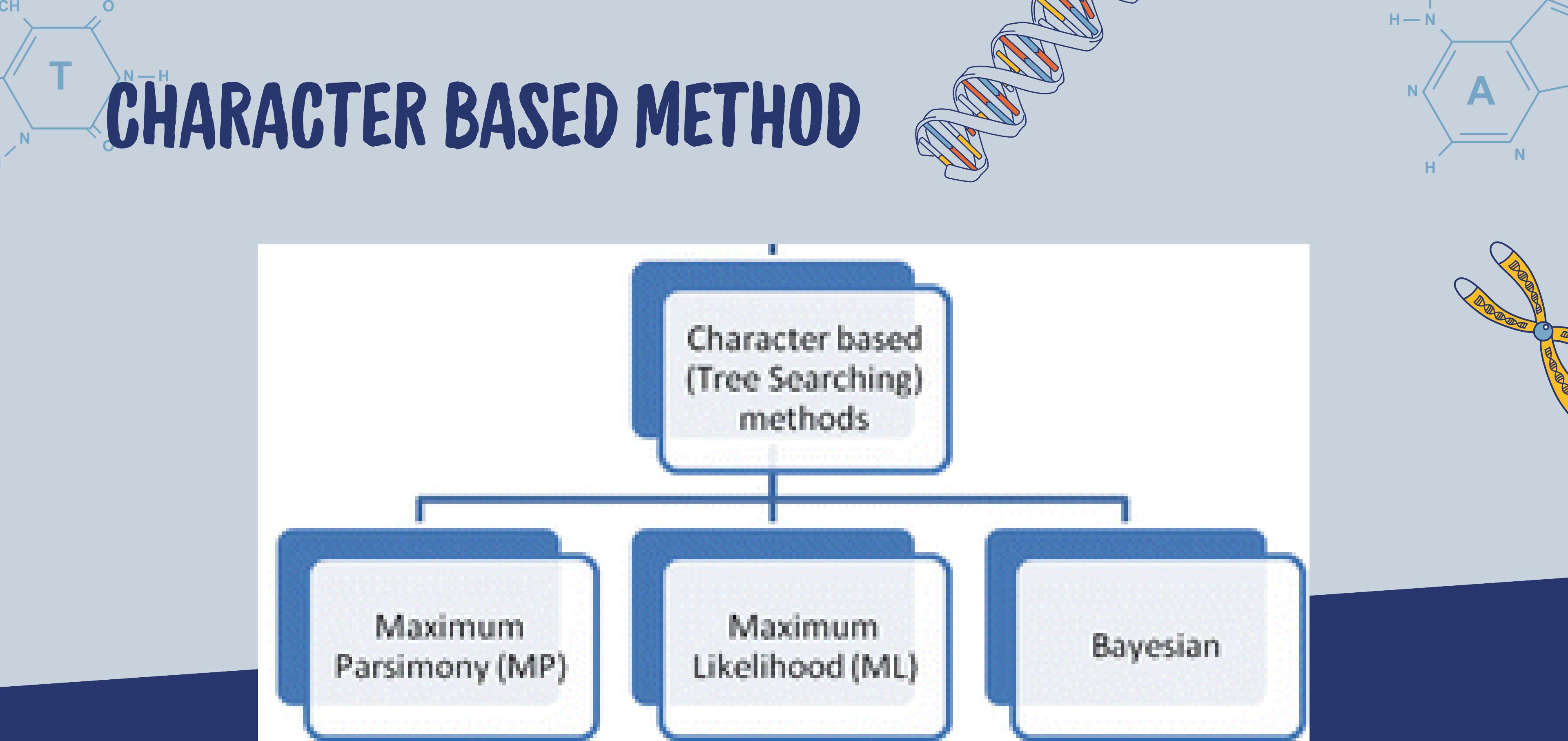
	A	R	C	O
A		7	7	9
R			7	8
C				10
O				

# CONTOH UPGMA

Pilih pasangan dengan mismatch terkecil. Misal kita pilih R dengan A. Maka R dan A memiliki hubungan yang dekat. Kemudian coret dari matriks. Hitung mismatch antara A/R dengan C, yaitu  $((A-C)+(R-C))/2 = (7+7)/2 = 7$ . Gunakan cara yang sama untuk menghitung mismatch A/R dengan O.

	A/R	C	O
A/R		7	8,5
C			10
O			





# MAXIMUM PARSIMONY METHOD

- Menggunakan prinsip Occam's razor,
- Berfokus pada meminimalkan jumlah langkah evolusi.
- Diperkenalkan oleh James S. Farris dan Walter M. Fitch pada 1970–1971.
- Bekerja dengan menggunakan situs informatif dalam urutan DNA, yaitu situs yang memiliki setidaknya dua nukleotida berbeda yang muncul dalam dua atau lebih urutan.
- Mencari semua topologi pohon yang mungkin dan menghitung jumlah substitusi minimum untuk menemukan pohon yang paling parsimonious

## KELEBIHAN

Kesederhanaan dan intuisi yang tinggi, sehingga mudah dipahami dan diterapkan

## KEKURANGAN

Untuk data yang besar MP kurang efisien karena menghasilkan jumlah taksa yang lebih besar sehingga meningkatkan kompleksitas topologi pohon dan algoritma.

# MAXIMUM PARSIMONY METHOD

## KARAKTER SEQUENCE

Karakter sequence dalam Metode MP dapat berbentuk ekspresi suatu gen (binary → 0 = tidak ada gen tertentu, 1 = ada gen tertentu) atau berbentuk posisi basa nitrogen dalam Multiple Sequence Alignment (MSA), yaitu A,C,T,G.

Sequences	a	b	c	d	e	f
1	G	G	G	G	G	G
2	G	G	G	A	G	T
3	G	G	A	T	A	G
4	G	A	T	C	A	T

	A	B
T1	1	1
T2	1	0
T3	0	1
T4	0	0

# MAXIMUM PARSIMONY METHOD

Metode MP tidak memiliki rumus atau model khusus seperti *distance based method* atau *maximum likelihood method* yang memerlukan model beserta asumsinya.

Perhitungan Metode MP didasarkan pada *informative sites*. *Informative sites* diperlukan untuk urutan spesies/taxa sebelum konstruksi pohon.

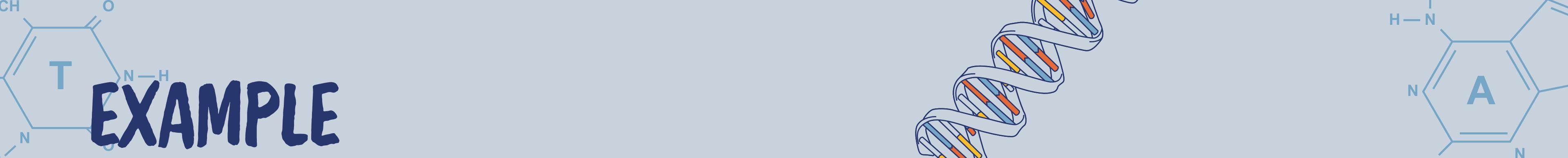
Kriteria situs informatif:

**Mengandung variasi karakter antarspesies** menunjukkan variasi untuk memisahkan spesies atau kelompok.

**Mengandung setidaknya dua karakter yang muncul lebih dari sekali**

misal 0 dan 1 atau dua nukleotida yang berbeda yang muncul setidaknya dua kali di antara spesies.

**Menyebabkan skor parsimony yang berbeda untuk setiap pohon**



# EXAMPLE

Dipilih empat spesies berbeda dari tabel NCBI di pembahasan 2.2 yang secara morfologi mirip, lalu kita akan membandingkan karakter di setiap sekuens DNanya. Berikut langkah-langkahnya:

**Diambil empat spesies berbeda dengan karakternya masing-masing**

*Antilope cervicapra* : ATGTTCGTTA

*Rupicapra sp. mtE1* : TTTCGCGCTGA

*Capra hircus* : CCCTGGTTTG

*Ovis aries* : GATGCCAACCC

*Antilope cervicapra*



*Rupicapra sp mE1*



*Capra hircus*



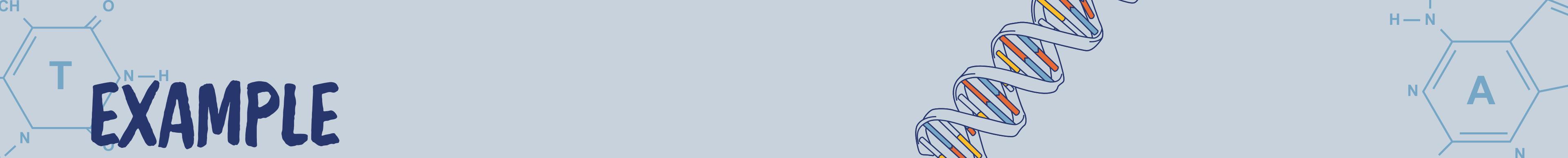
*Ovis aries*



**EXAMPLE**

Menyusun matriks/tabel karakter

Spesies	Site 1	Site 2	Site 3	Site 4	Site 5	Site 6	Site 7	Site 8	Site 9	Site 10
<i>Antilope cervicapra</i>	A	T	G	T	T	C	G	T	T	A
<i>Rupicapra sp. mtEl</i>	T	T	C	G	C	G	C	T	G	A
<i>Capra hircus</i>	C	C	C	T	G	G	T	T	T	G
<i>Ovis aries</i>	G	A	T	G	C	C	A	C	C	C



# EXAMPLE

## Menentukan *informative sites*

Kita punya dua situs informasi yaitu **Site 4** dan **Site 6** karena Site 4 memiliki variasi karakter (T dan G) yang muncul lebih dari sekali (T pada *Antilope cervicapra* dan *Capra hircus*; G pada *Rupicapra sp. mtE1* dan *Ovis aries*).

Sedangkan Site 6 memiliki variasi karakter (C dan G) yang muncul lebih dari sekali (C pada *Antilope cervicapra* dan *Ovis aries*; G pada *Rupicapra sp. mtE1* dan *Capra hircus*). Hal ini berarti, ada dua pohon yang terbentuk dari masing-masing *informative sites*.

**EXAMPLE**

Spesies	Site 1	Site 2	Site 3	Site 4	Site 5	Site 6	Site 7	Site 8	Site 9	Site 10
<i>Antilope cervicapra</i> <i>a</i>	A	T	G	T	T	C	G	T	T	A
<i>Rupicapra sp. mtEl</i>	T	T	C	G	C	G	C	T	G	A
<i>Capra hircus</i>	C	C	C	T	G	G	T	T	T	G
<i>Ovis aries</i>	G	A	T	G	C	C	A	C	C	C

CH  
O  
T  
N—H

# EXAMPLE

**Membentuk kemungkinan pohon**

$$n!! \equiv \begin{cases} n \cdot (n-2) \dots 5 \cdot 3 \cdot 1 & n > 0 \text{ odd} \\ n \cdot (n-2) \dots 6 \cdot 4 \cdot 2 & n > 0 \text{ even} \\ 1 & n = -1, 0. \end{cases}$$

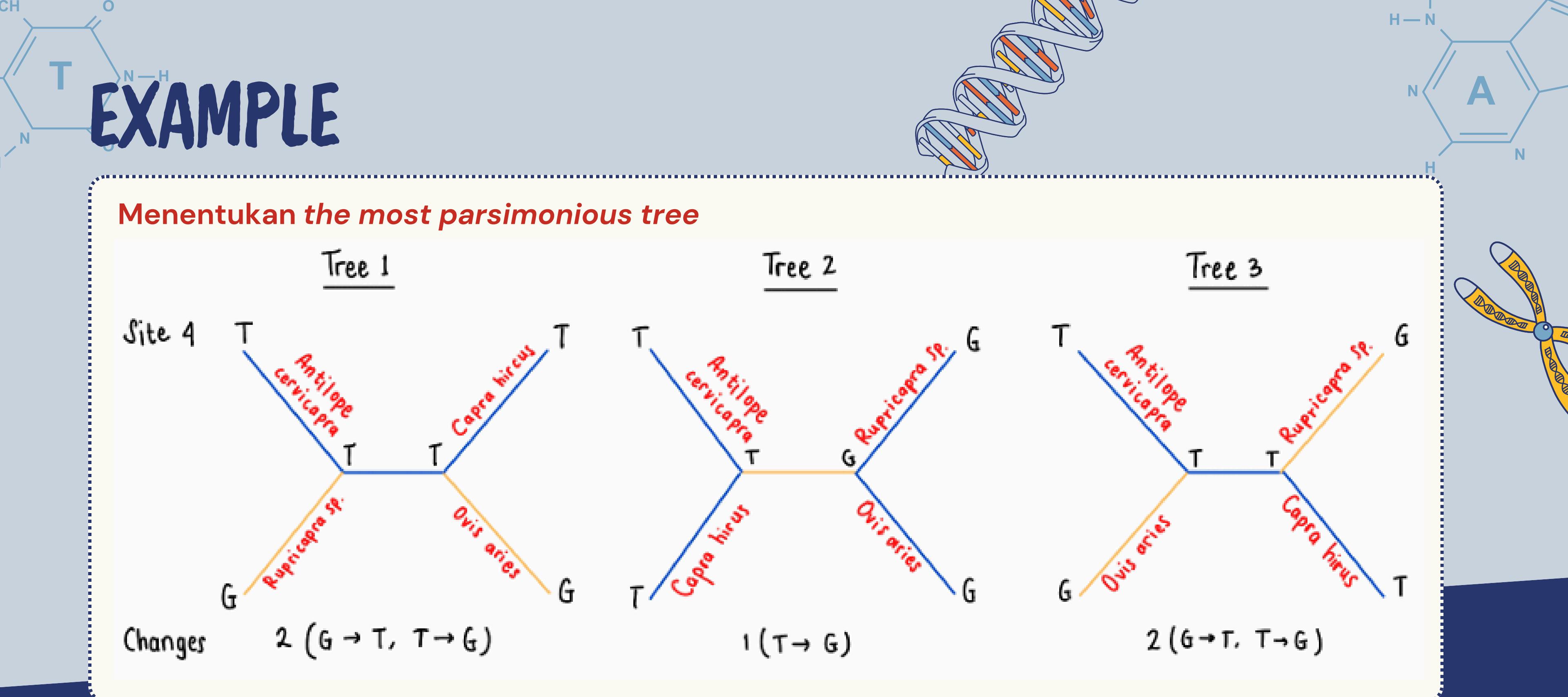
dengan  $n = 2k - 5$  dan  $k$  adalah banyaknya spesies atau taxa

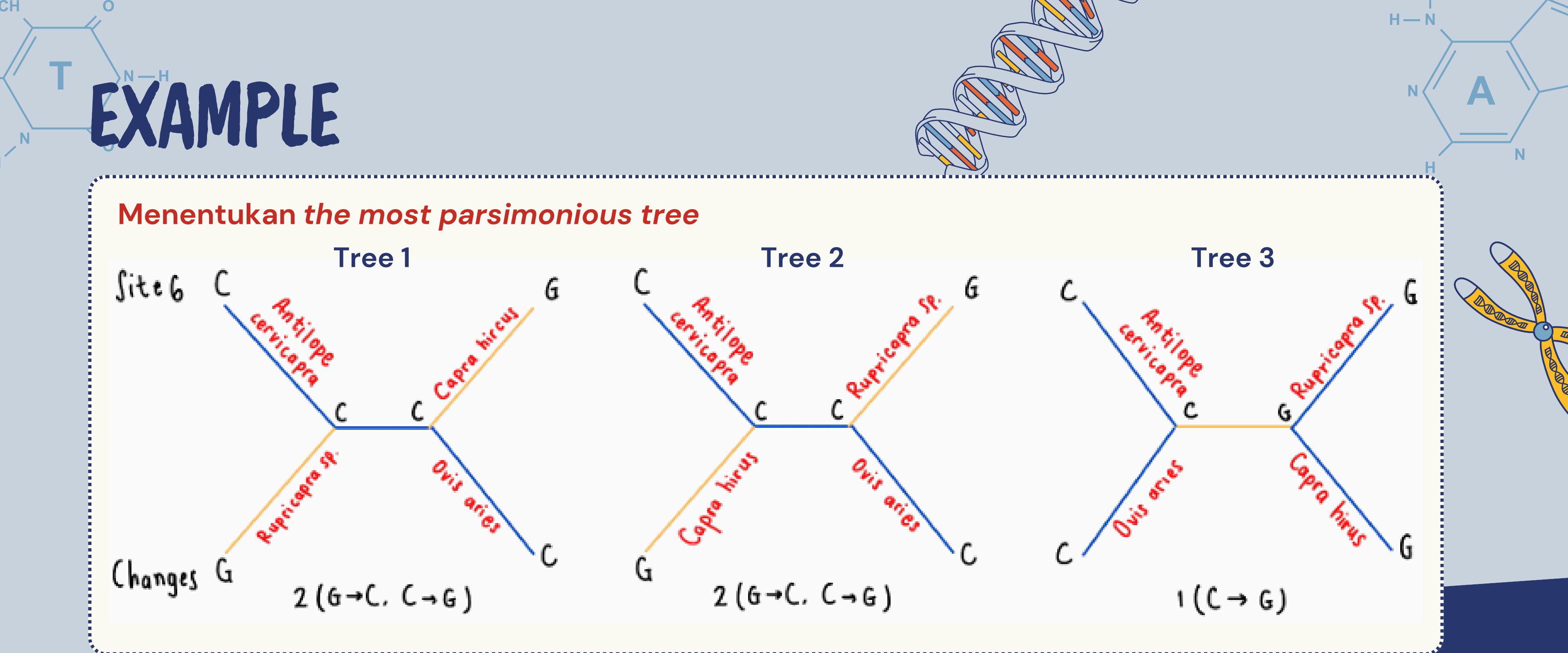
Kita punya  $k = 4$  dan  $n = 3$ , jadi kemungkinan pohnnya ada tiga.

**Pohon 1:** ((Antilope cervicapra, Rupicapra sp. mtE1), (Capra hircus, Ovis aries))

**Pohon 2:** ((Antilope cervicapra, Capra hircus), (Rupicapra sp. mtE1, Ovis aries))

**Pohon 3:** ((Antilope cervicapra, Ovis aries), (Rupicapra sp. mtE1, Capra hircus))



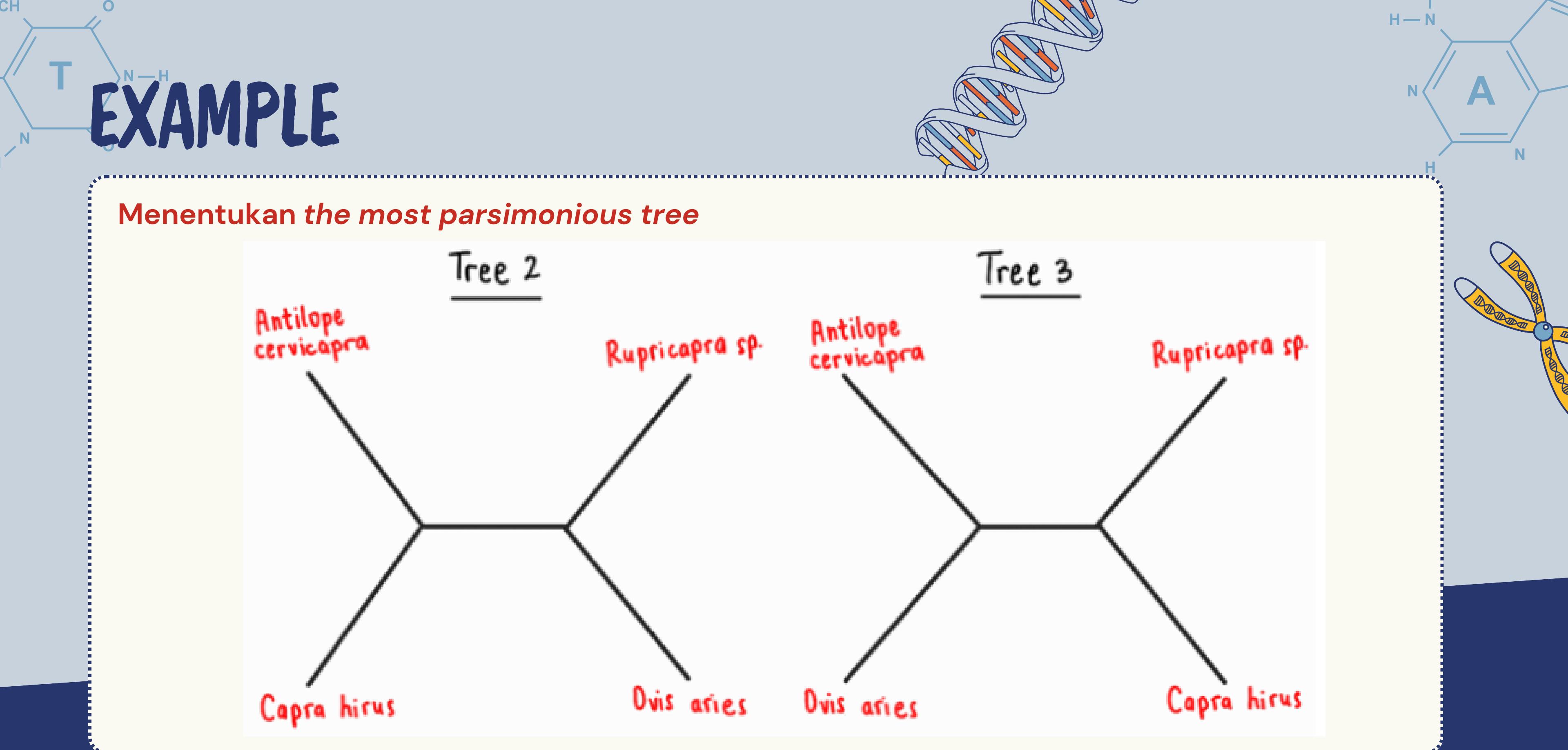


	Tree 1	Tree 2	Tree 3
Sum	4	3	3

# EXAMPLE

**Menentukan *the most parsimonious tree***

Dari perhitungan manual tersebut didapatkan total perubahan di setiap kemungkinan pohon. Pohon 1 memiliki total perubahan sebanyak 4, pohon 2 memiliki total perubahan sebanyak 3 dan pohon 3 memiliki total perubahan sebanyak 3. Ada dua pohon yang memiliki perubahan karakter paling sedikit, yaitu pohon 2 ((*Antilope cervicapra*, *Capra hirus*) dengan (*Rupicapra sp.*, *Ovis aries*)) dan pohon 3 ((*Antilope cervicapra*, *Ovis aries*) dengan (*Rupicapra sp.*, *Capra hirus*)). Sehingga kita dapatkan pohon yang paling parsimonious adalah pohon 2 dan pohon 3.



C=CN1C=CC2=C1C(=O)N(C2=O)N T

CN1C=CC2=C1C(=O)N(C2=O)N A

# EXAMPLE

Menentukan *the most parsimonious tree*



Tree 2





Tree 3



CN1C=CC2=C1C(=O)N(C2=O)N

Nc1ccccc1

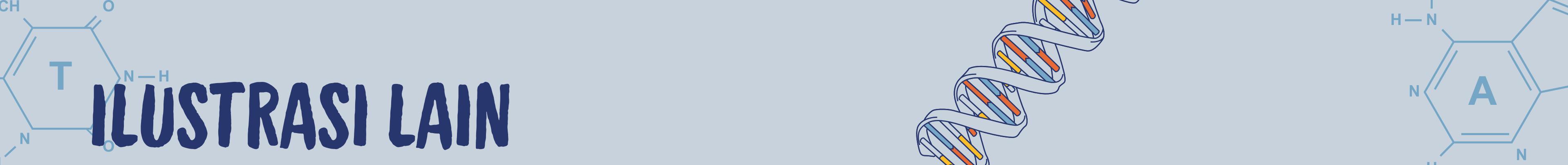
CN1C=CC2=C1C(=O)N(C2=O)N

Nc1ccccc1

# EXAMPLE

**Menentukan *the most parsimonious tree***

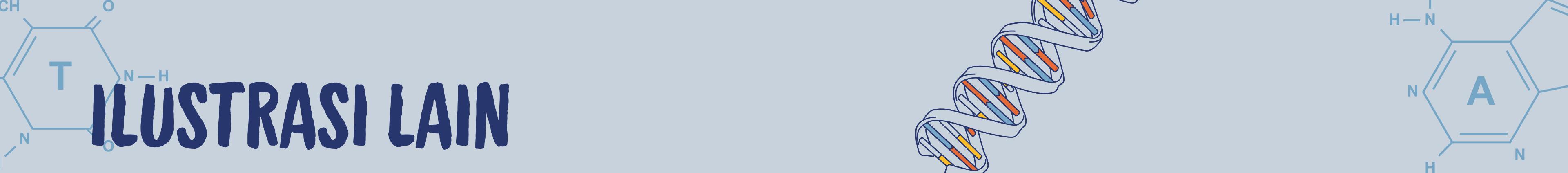
Beberapa program filogenetik akan memilih satu pohon berdasarkan urutan input atau aturan internal jika terdapat beberapa pohon dengan skor yang sama, sedangkan program lain, seperti PAUP (*Phylogenetic Analysis Using Parsimony*) atau MEGA (*Molecular Evolutionary Genetics Analysis*), akan mengidentifikasi bahwa terdapat lebih dari satu pohon yang sama-sama parsimonious dan memberikan opsi untuk melihat atau memilih di antara pohon-pohon tersebut.



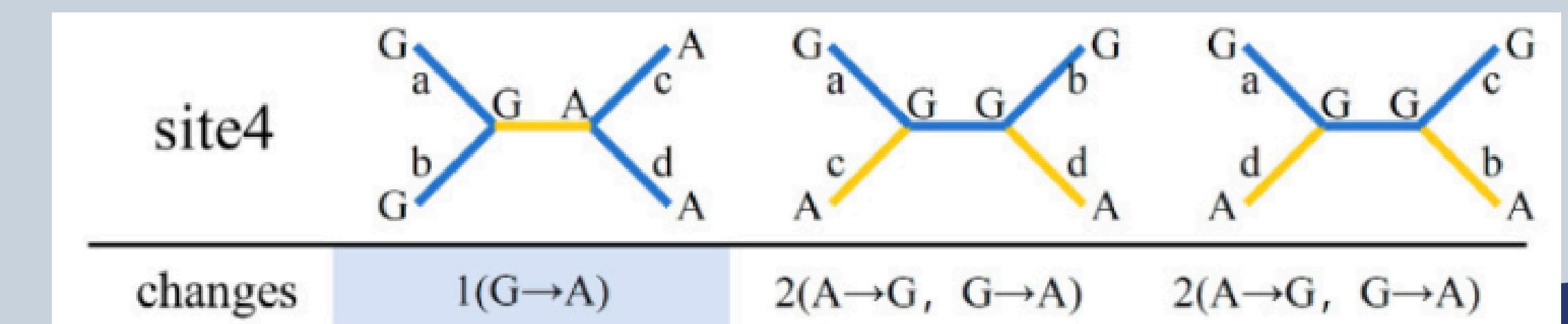
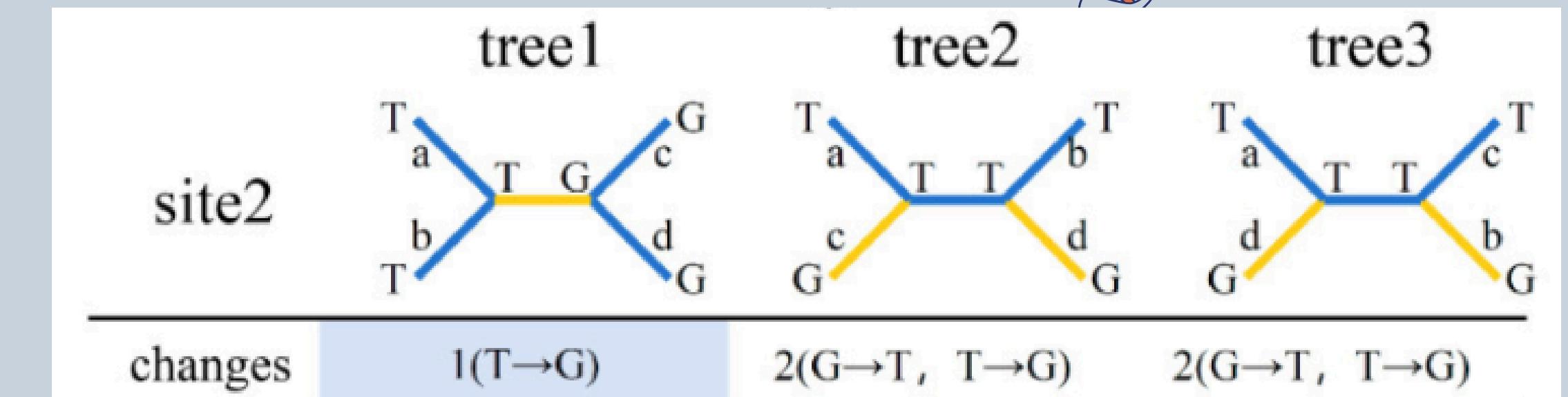
# ILUSTRASI LAIN

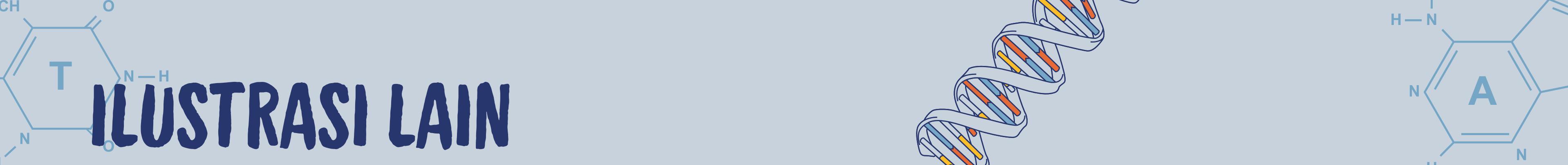
## Informative Sites

	site1	site2	site3	site4	site5	site6	site7
sequence-a	A	T	T	G	C	C	A
sequence-b	A	T	C	G	A	C	T
sequence-c	A	G	T	A	A	C	A
sequence-d	T	G	T	A	A	C	T



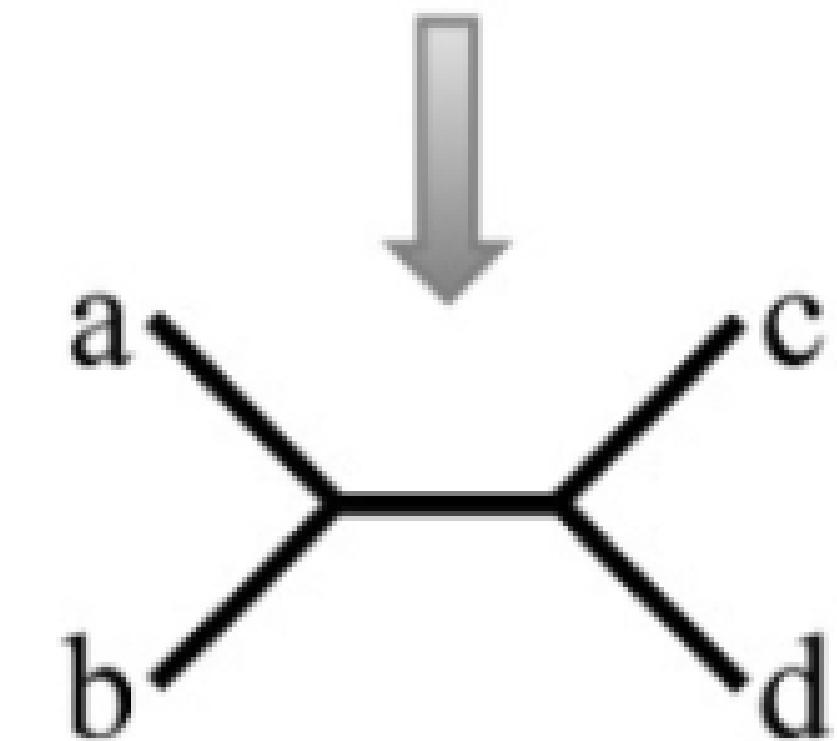
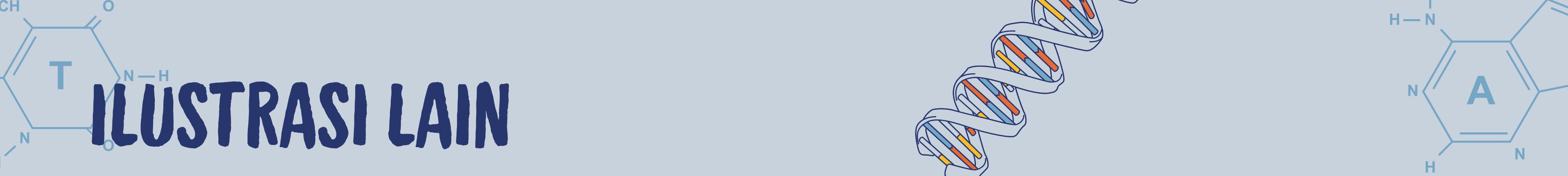
# ILUSTRASI LAIN





site 7			
changes	2(T→A, A→T)	1(A→T)	2(T→A, A→T)

sum	4	5	6
-----	---	---	---



the most parsimonious tree — tree 1

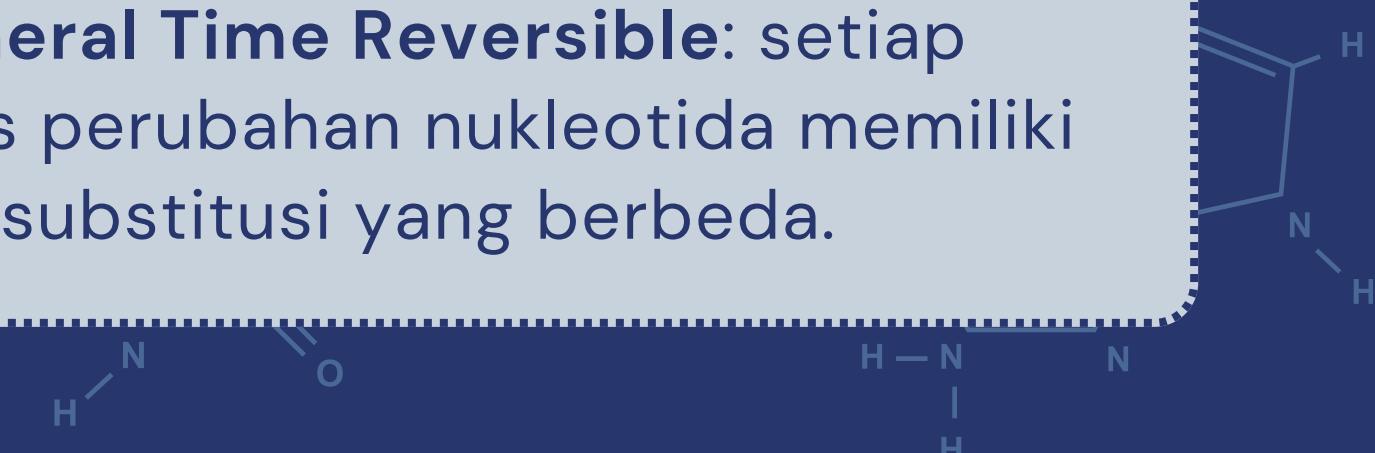
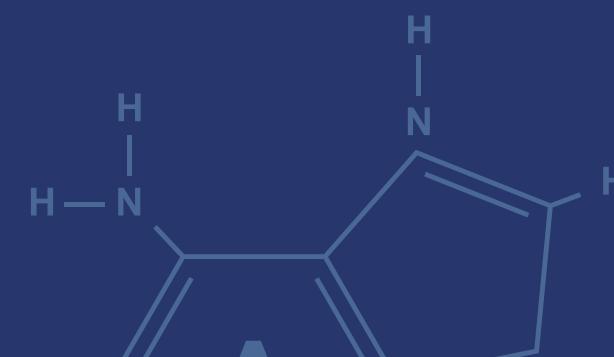
# MAXIMUM LIKELIHOOD METHOD

- Maximum Likelihood (ML) adalah metode berbasis probabilitas.
- Tujuan: menemukan pohon filogenetik "terbaik" untuk menjelaskan sequence yang diamati.
- Berdasarkan **model evolusi molekuler** yang memodelkan perubahan nukleotida.

Memodelkan **laju dan probabilitas perubahan** dari satu nukleotida selama evolusi.

Yang paling umum digunakan:

- **Jukes-Cantor (1969):** mengasumsikan semua substitusi nukletioda terjadi pada laju yang sama.
- **Kimura 2-Parameter (1980):** mempertimbangkan perbedaan laju transisi dan transversi.
- **General Time Reversible:** setiap jenis perubahan nukleotida memiliki laju substitusi yang berbeda.



# PENDEKATAN MAXIMUM LIKELIHOOD

**Starting point:** Data observasi dan model probabilitas dari data tersebut.

## Contoh

Data: hasil lempar koin 10 kali, yaitu 7 heads, 3 tails

Model: koin memiliki probabilitas  $p$  untuk heads,  $1 - p$  untuk tails

Probabilitas mengamati  $h$  heads dalam  $n$  kali lemparan:

$$P(\text{h heads}) = \binom{n}{h} p^h (1 - p)^{n-h}$$

**Goal:** Estimasi terbaik dari nilai parameter  $p$  berdasarkan observasi.

Likelihood (Model) = Probability (Data | Model)

# MAXIMUM LIKELIHOOD PADA FILOGENI

**Data:** sequence alignment

**Parameter Model:** frekuensi nukleotida, laju substitusi nukleotida, *tree topology*, dan panjang cabang pohon filogeni

- Pilih nilai awal untuk setiap parameter, hitung *likelihood*
- Ubah nilai parameter sedikit demi sedikit dengan tujuan *likelihood* semakin baik
- Ulangi sampai ditemukan yang maksimum

**Hasil:**

- ML mengestimasi *tree topology*
- ML mengestimasi panjang cabang
- ML mengestimasi parameter lainnya
- Mengukur seberapa baik model sesuai (*fit*) dengan data (*likelihood*)

# PROBABILITAS DARI SATU KOLOM DAN SELURUH PADA ALIGNMENT

**Asumsi:** telah diberikan *tree topology* dan parameter lainnya

A	T	G	G	A	T	T	C	A
A	T	G	G	T	T	T	-	A
A	C	G	G	A	T	T	-	A
A	G	G	G	T	T	T	-	A

Kolom pada *alignment* berisi **nukleotida yang homolog**.

Artinya: posisi yang sesuai di antara sekuens DNA yang berbeda menunjukkan nukleotida yang dianggap berevolusi dari nenek moyang yang sama.

$$L_{(j)} = \text{Prob} \left( \begin{array}{c} T \\ \diagup \quad \diagdown \\ A & A \\ \diagdown \quad \diagup \\ T & G \end{array} \right) + \text{Prob} \left( \begin{array}{c} T \\ \diagup \quad \diagdown \\ C & A \\ \diagdown \quad \diagup \\ T & G \end{array} \right)$$

+ ...

$$+ \text{Prob} \left( \begin{array}{c} T \\ \diagup \quad \diagdown \\ T & T \\ \diagdown \quad \diagup \\ T & G \end{array} \right)$$

- Probabilitas harus dijumlahkan dari seluruh kombinasi mungkin *ancestral nucleotides*
- Pada gambar disamping terdapat 2 node internal, sehingga ada 16 kombinasi mungkin

$$L = L_{(1)} \cdot L_{(2)} \cdots L_{(N)} = \prod_{j=1}^N L_{(j)}$$

- Probabilitas dari masing-masing kolom dikalikan untuk mendapat probabilitas keseluruhan *alignment*

# PENERAPAN MAXIMUM LIKELIHOOD METHOD DENGAN METODE MANUAL

## Data:

1. ATGTCGTTA = *Antilope cervicapra*
2. TTTCGCGCTGA = *Rupicapra sp.*
3. CCCTGGTTG = *Capra hircus*
4. GATGCCAAC = *Ovis aries*

**Model:** Jukes-Cantor

**Asumsi:** semua substitusi memiliki peluang yang sama dan mutasi terjadi secara acak

$$d = -\frac{3}{4} \log(1 - \frac{4}{3}p)$$

$d$  = jarak antar sekuens

$p$  = proporsi perbedaan antar sekuens

## Menghitung jarak setiap pasangan sekuen yang mungkin dengan rumus dari model Jukes-Cantor

A T G T T C G T T A	<i>Antilope cervicapra</i>
T T C G C G C T G A	<i>Rupicapra sp.</i>
$p = \frac{7}{10} = 0.7$ $d = -\frac{3}{4} \log(1 - \frac{4}{3}(0.7)) \approx 2.031$	

T T C G C G C T G A	<i>Rupicapra sp.</i>
C C C T G G T T T G	<i>Capra hircus</i>
$p = \frac{7}{10} = 0.7$ $d = -\frac{3}{4} \log(1 - \frac{4}{3}(0.7)) \approx 2.031$	

A T G T T C G T T A	<i>Antilope cervicapra</i>
C C C T G G T T T G	<i>Capra hircus</i>
$p = \frac{7}{10} = 0.7$ $d = -\frac{3}{4} \log(1 - \frac{4}{3}(0.7)) \approx 2.031$	

T T C G C G C T G A	<i>Rupicapra sp.</i>
G A T G C C A A C C	<i>Ovis aries</i>
$p = \frac{8}{10} = 0.8$ $d = -\frac{3}{4} \log(1 - \frac{4}{3}(0.8)) \approx TD \text{ (tidak terdefinisi)}$	

A T G T T C G T T A	<i>Antilope cervicapra</i>
G A T G C C A A C C	<i>Ovis aries</i>
$p = \frac{9}{10} = 0.9$ $d = -\frac{3}{4} \log(1 - \frac{4}{3}(0.9)) \approx TD \text{ (tidak terdefinisi)}$	

C C C T G G T T T G	<i>Capra hircus</i>
G A T G C C A A C C	<i>Ovis aries</i>
$p = \frac{10}{10} = 1$ $d = -\frac{3}{4} \log(1 - \frac{4}{3}(0.7)) \approx TD \text{ (tidak terdefinisi)}$	

## Matriks jarak dari spesies yang ada

	<i>Antilope cervicapra</i>	<i>Rupicapra sp.</i>	<i>Capra hircus</i>	<i>Ovis aries</i>
<i>Antilope cervicapra</i>	0	2.031	2.031	-
<i>Rupicapra sp.</i>	2.031	0	2.031	-
<i>Capra hircus</i>	2.031	2.031	0	-
<i>Ovis aries</i>	-	-	-	0

Dari data jarak yang diketahui, penghitungan selanjutnya hanya akan difokuskan pada jarak antara sekuens yang terdefinisi:

- *Antilope cervicapra* – *Rupicapra sp.* : 2.031
- *Antilope cervicapra* – *Capra hirus* : 2.031
- *Rupicapra sp.* – *Capra hircus* : 2.031

*Ovis aries* akan dikeluarkan dari pembentukan pohon ini karena perbedaan sekuens terlalu besar.

Dari data jarak yang diketahui, penghitungan selanjutnya hanya akan difokuskan pada jarak antara sekuens yang terdefinisi:

- *Antilope cervicapra* – *Rupicapra* sp. : 2.031
- *Antilope cervicapra* – *Capra hirus* : 2.031
- *Rupicapra* sp. – *Capra hircus* : 2.031

*Ovis aries* akan dikeluarkan dari pembentukan pohon ini karena perbedaan sekuens terlalu besar.

**Terdapat 10 site di setiap pasangan spesies dengan 7 posisi berbeda dan 3 posisi yang sama.**

### Likelihood dari Model Jukes-Cantor

Diketahui  $d = 2.031$

$$P(\text{change}) = \frac{1}{4} - \frac{1}{4}e^{-4d/3} \approx 0.2335$$

$$P(\text{no change}) = \frac{1}{4} + \frac{3}{4}e^{-4d/3} \approx 0.2995$$

Log-likelihood untuk setiap pasangan:

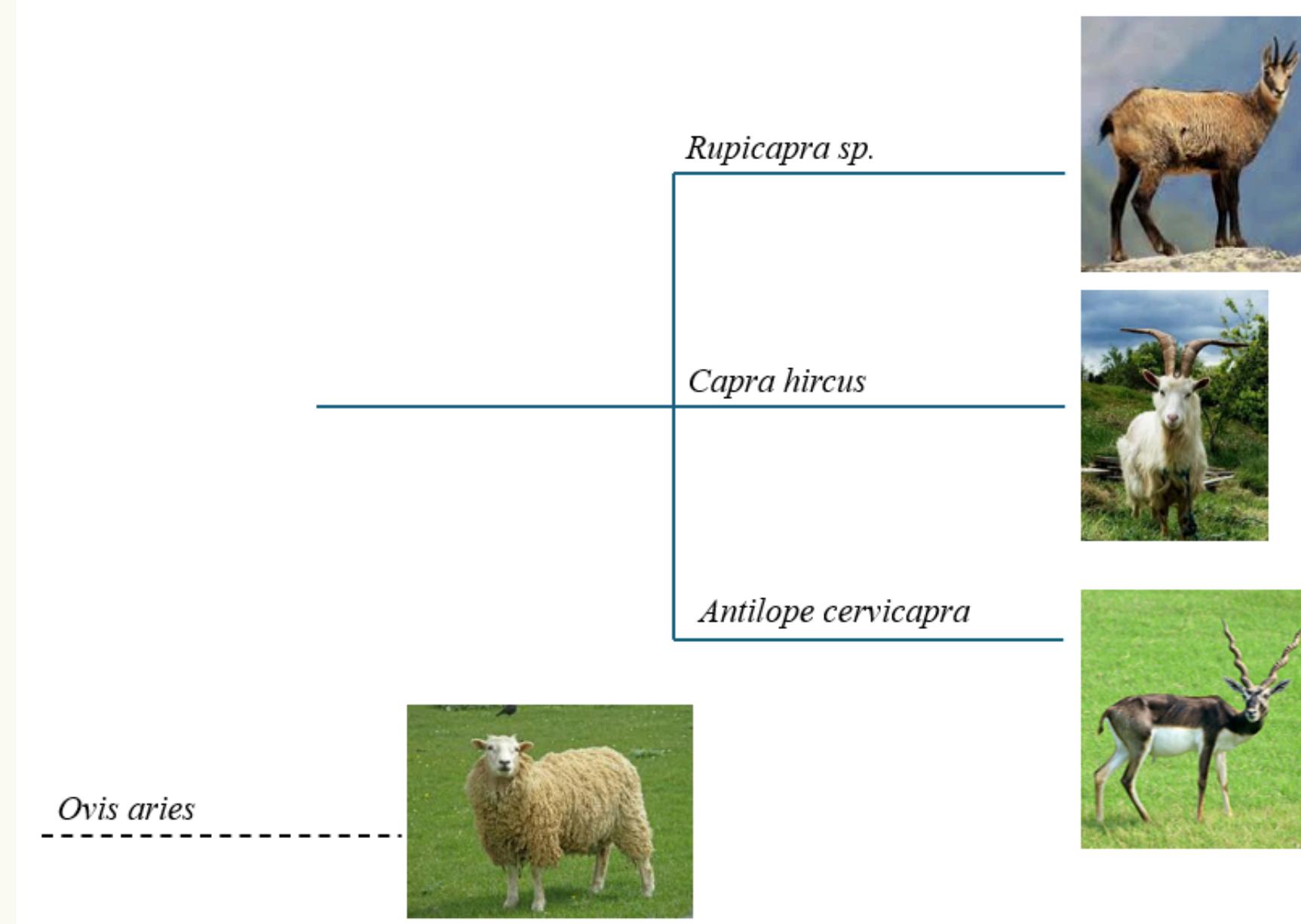
$$\ln L = (7 \times \ln(0.2335)) + (3 \times \ln(0.2995))$$

$$\ln L = -10.171 + (-3.618) = -13.789$$

Total untuk ketiganya:

$$\text{Total log-likelihood} = 3 \times (-13.789) = -41.367$$

Berdasarkan data jarak yang sama di antara tiga pasangan spesies dan asumsi bahwa nilai-nilai tersebut menunjukkan tingkat divergensi yang serupa. Pohon paling sederhana yang dapat dihasilkan:



- **Node pusat (root):** Merepresentasikan leluhur bersama dari ketiga spesies, dari mana semua cabang berawal.
- **Cabang yang simetris:** Menggambarkan bahwa jarak evolusi antara setiap pasangan spesies relatif sama (2.031), sehingga pohon berbentuk bintang adalah representasi paling sederhana dari hubungan tersebut.

# INFERENSI BAYESIAN METHOD



Bayesian Inference Method adalah pendekatan statistik yang digunakan untuk merekonstruksi pohon filogenetik dengan menggabungkan informasi awal atau asumsi tertentu (dikenal sebagai prior) dengan data sekuens yang tersedia. Metode ini didasarkan pada Teorema Bayes, yang memperbarui keyakinan kita terhadap suatu model atau hipotesis setelah mempertimbangkan bukti atau data baru.

Konsep dasar Bayesian inference:

- Prior Probability (Prior):** Probabilitas awal suatu hipotesis sebelum ada data baru. Prior ini didasarkan pada pengetahuan awal atau asumsi yang ada.
- Likelihood:** Kemungkinan data atau bukti baru terjadi jika hipotesis yang diajukan benar.
- Posterior Probability (Posterior):** Probabilitas yang diperbarui dari hipotesis setelah memperhitungkan data baru. Posterior ini diperoleh dengan mengalikan prior dengan likelihood, lalu menormalkan hasilnya.

# INFERENSI BAYESIAN METHOD



Secara matematis, berikut adalah rumus Bayesian Inference

$$P(H|D) = \frac{P(D|H)P(H)}{P(D)}$$

$P(H|D)$  adalah posterior probability atau probabilitas hipotesis H setelah mengamati data D.

$P(D|H)$  adalah likelihood, yaitu probabilitas data D diberikan hipotesis H.

$P(H)$  adalah prior probability dari hipotesis H

$P(D)$  adalah probabilitas total dari data D

# CONTOH KASUS DENGAN METODE BAYES



Antilope cervicapra:	ATGTTCGTTA
Rupicapra sp. mtE1:	TTCGCGCTGA
Capra hircus:	CCCTGGTTTG
Ovis aries:	GATGCCAACCC

Langkah-langkah :

- **Menentukan hipotesis (H)**

Sebagai contoh, kita ingin mengetahui dua spesies tertentu memiliki hubungan nenek moyang yang sama. misalkan kita ingin mengetahui probabilitas bahwa Antilope cervicapra dan Rupicapra sp. mtE1 memiliki nenek moyang bersama. Hipotesis kita adalah bahwa "Antilope cervicapra dan Rupicapra sp. mtE1 berasal dari nenek moyang yang sama."

# CONTOH KASUS DENGAN METODE BAYES



Antilope cervicapra: ATGTTCGTTA  
Rupicapra sp. mtE1: TTTCGCGCTGA  
Capra hircus: CCCTGGTTTG  
Ovis aries: GATGCCAACCC

- Menentukan Likelihood ( $P(D|H)$ )

$$P(D|H) = \frac{\text{Jumlah posisi yang sama}}{\text{Total panjang}} = \frac{3}{10} = 0,3$$

- Menentukan Prior Probability ( $P(H)$ )

$$P(H) = \frac{1}{6} \approx 0,1667$$

- Menghitung probabilitas Total Total ( $P(D)$ )

$$P(D) = \frac{0.3+0.1+0+0.1+0.1+0.2}{6} = \frac{2}{15} \approx 0.13$$

# CONTOH KASUS DENGAN METODE BAYES



Antilope cervicapra: ATGTTCGTTA  
Rupicapra sp. mtE1: TTTCGCGCTGA  
Capra hircus: CCCTGGTTTG  
Ovis aries: GATGCCAACCC

$$P(H|D) = \frac{P(D|H)P(H)}{P(D)}$$

- Menghitung Posterior Probability ( $P(H|D)$ )

$$P(H | D) = \frac{0.3 \cdot 0.1667}{0.13} = \frac{5001}{13000} \approx 0.3846$$

## Interpetasi Hasil

Probabilitas bahwa Antilope cervicapra dan Rupicapra sp. mtE1 memiliki nenek moyang bersama berdasarkan sekuen yang diberikan adalah sekitar 0.385 atau 38.5%.

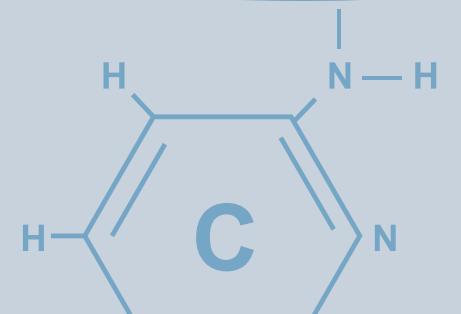
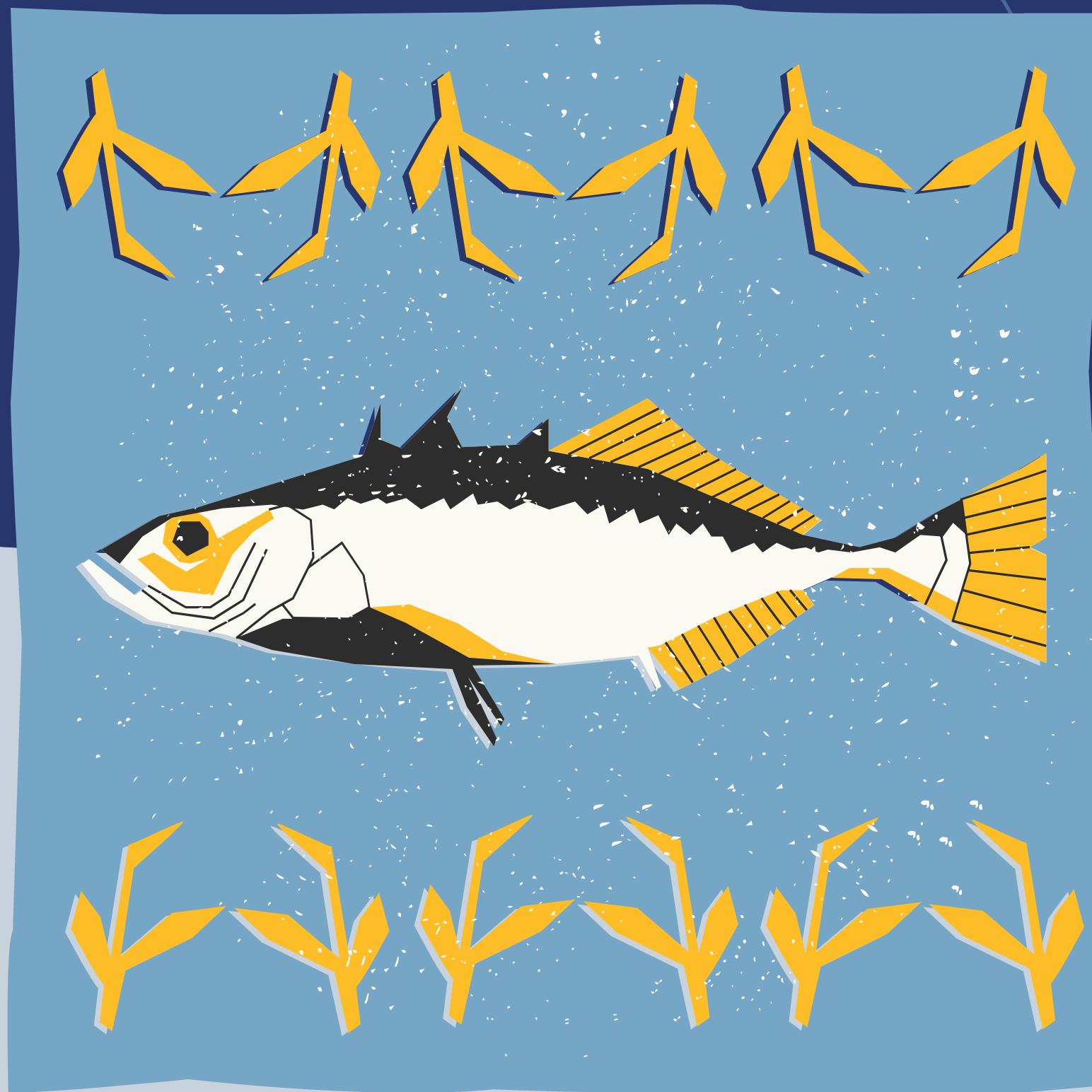
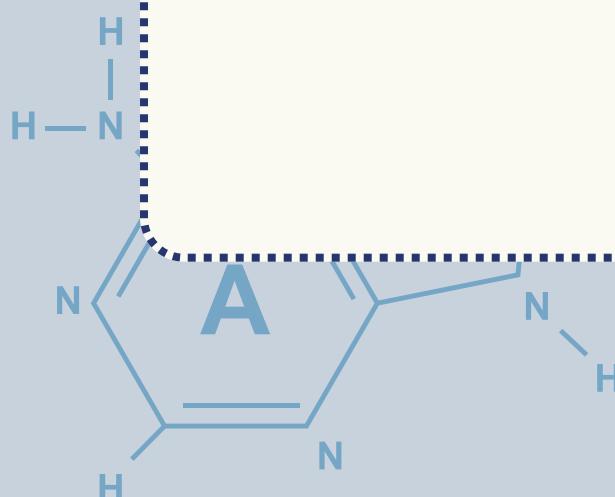
# KELEBIHAN DAN KEKURANGAN BAYES



Bayesian inference memungkinkan pembaruan probabilitas hipotesis berdasarkan data baru, menggabungkan informasi awal (prior) yang sangat berguna dalam situasi data kecil dan penuh ketidakpastian. Metode ini fleksibel, mampu menangani berbagai model kompleks, dan memiliki aplikasi luas dalam banyak bidang. Namun, Bayesian inference rentan terhadap subjektivitas dalam pemilihan prior, membutuhkan komputasi yang intensif, dan sering kali sulit dipahami oleh pengguna tanpa latar belakang statistik yang kuat.

## METODE INTEGRATIF KOMPUTASI TINGKAT LANJUT

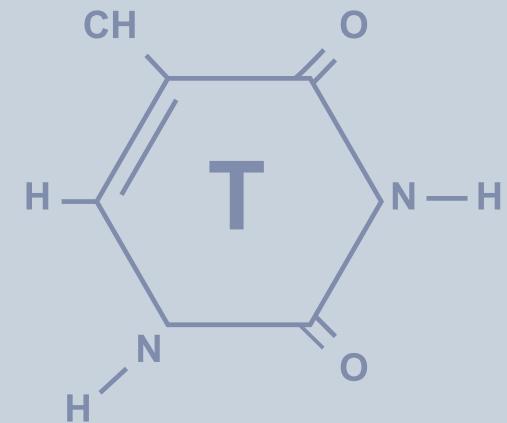
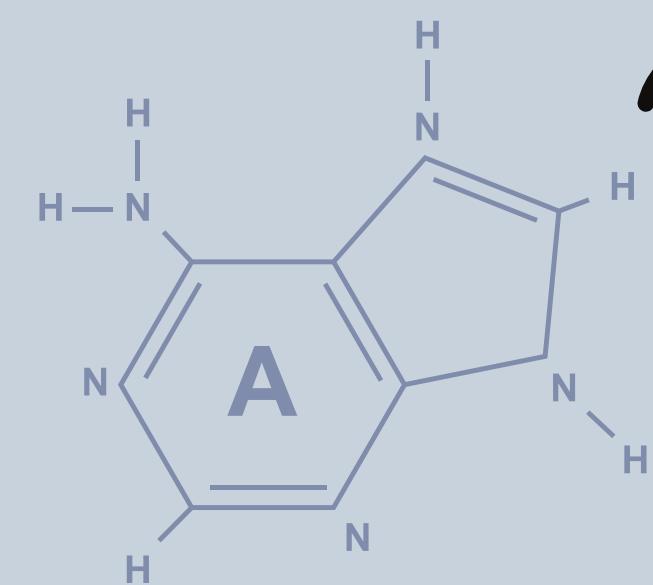
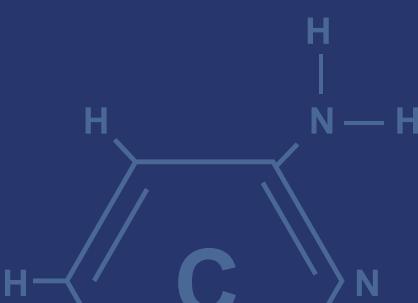
1. Metode Filogeni Konkatenasi
2. Metode Filogeni Koalesensi





## Metode Filogeni Konkatenasi

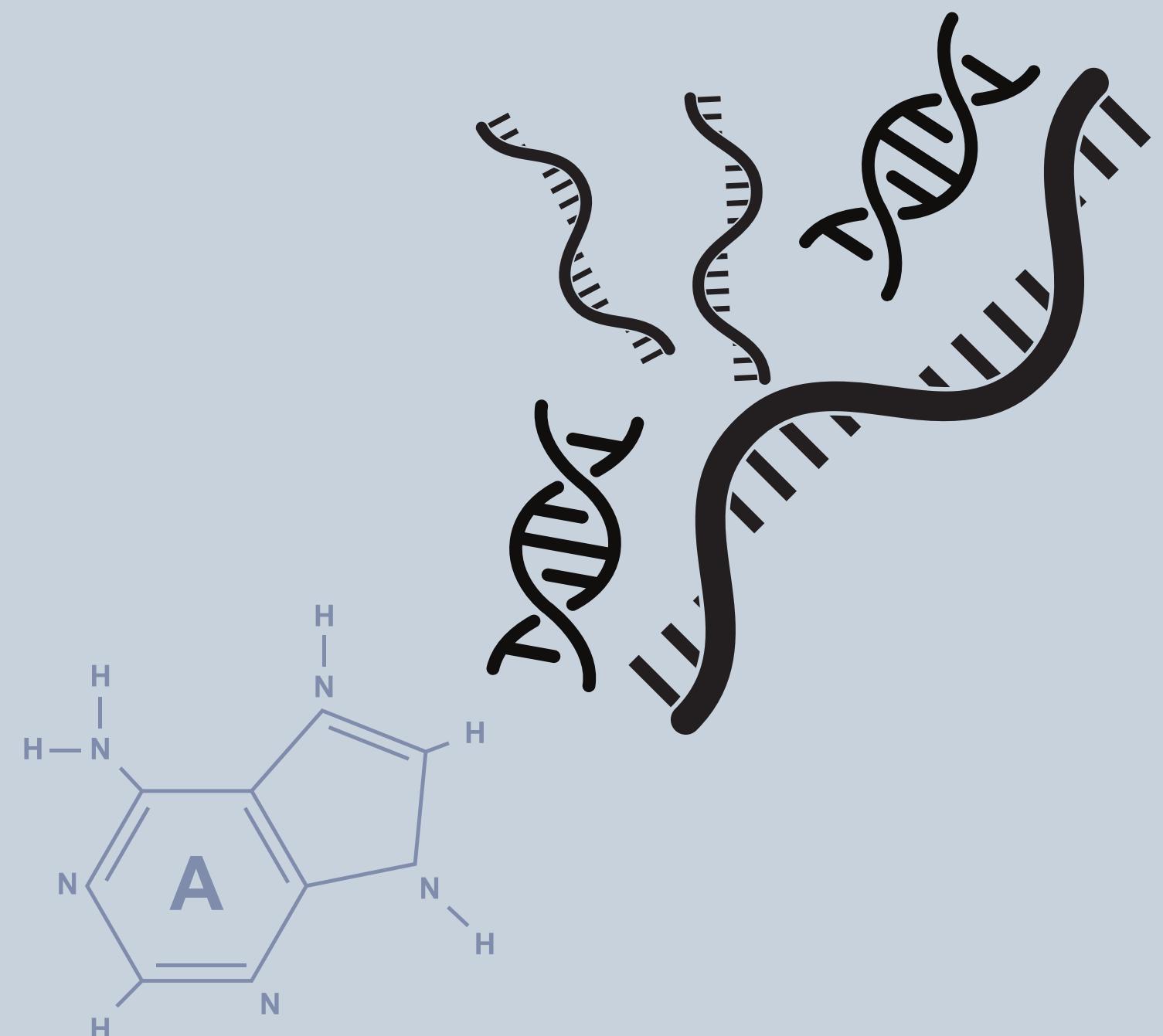
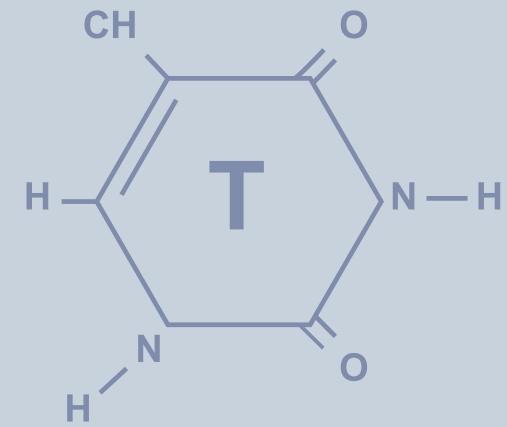
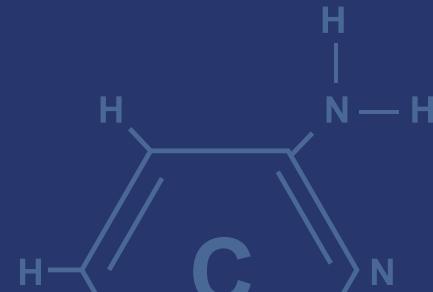
- Teknik yang digunakan untuk menyusun pohon filogenetik dengan menggabungkan data dari beberapa gen atau genom secara keseluruhan menjadi satu dataset besar. Memiliki asumsi bahwa semua gen atau lokus yang ada dalam dataset berbagi sejarah evolusi yang sama.
- Pendekatan melibatkan penjajaran urutan dari gen yang berbeda, lalu menggabungkannya ke dalam satu dataset tunggal.
- Keunggulan utama yaitu kemampuannya untuk menghasilkan pohon yang lebih terperinci.
- Keterbatasan paling utama adalah asumsi bahwa semua gen yang dikonkatenasi berbagi sejarah evolusi yang sama. Biaya komputasi untuk mengkonkatenasi dataset genom besar juga cukup besar.



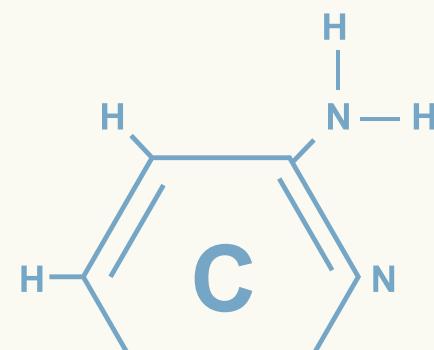
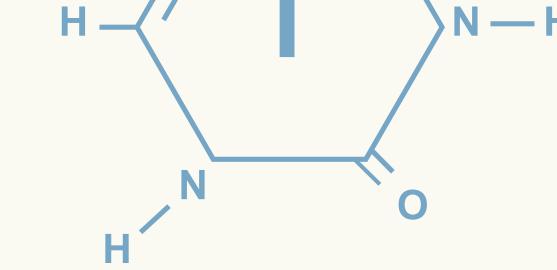
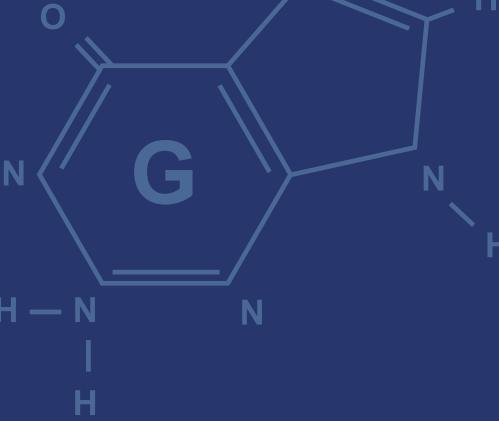


## Metode Filogeni Koalesensi

- Metode ini memanfaatkan konsep koalesensi, yaitu model dari genetic drift yang melacak sejarah genealogi alel dalam populasi kembali ke nenek moyang bersama yang paling baru.
- Fokus utamanya adalah menginferensi pohon spesies langsung dari pohon gen, daripada mengasumsikan bahwa semua lokus berbagi sejarah yang sama.
- Keunggulannya yaitu mereka memodelkan variasi alami dalam proses evolusi di seluruh genom, memberikan pandangan yang lebih cermat dan akurat tentang hubungan spesies.
- Kelemahannya yaitu metode ini membutuhkan komputasi yang intensif, dengan algoritma yang canggih dan daya pemrosesan tinggi untuk menangani banyak pohon gen yang terlibat.



# IMPLEMENTASI PROGRAM DI R DAN PYHTON



# DISTANCE BASED METHOD

```
setwd("D:/R_study/Constructing_phylogenetic_trees/")
library(adegenet)
# reading alignments with the fasta format and outputting a DNAbin object
dna_K00927 <- fasta2DNAbin("./data/ntK00927.fas")

## method_1: NJ
#install.packages("ape")
library(ape)
# computing a matrix of pairwise distances from DNA sequences
dist <- dist.dna(dna_K00927)
# reconstructing a NJ tree from a distance matrix with possibly missing values
treenj <- njs(dist)
treenj <- root(treenj, outgroup = "ece", r = TRUE)
# bootstrap analysis
bstrees <- boot.phylo(treenj, dna_K00927,
                      FUN = function(x) nj(dist.dna(x)),
                      trees = FALSE)
# Visualizing the NJ tree
plot(treenj)
# adding bootstrap values
drawSupportOnEdges(bstrees)
```

```
# Memuat library yang diperlukan
library(ape)
library(phangorn)

# Menyimpan data sekuens DNA dalam format FASTA
sequences <- ">Canis lupus\nATATGGTGAG\n>Lama pacos\nGTTAATGTAG\n>Antilope cervicapra\nATGTCGTTA\n>Babirusa\nCTTGTGTTGAA"

# Menyimpan ke file FASTA
writeLines(sequences, "sequences.fasta")

# Membaca data sekuens DNA dari file
dna_sequences <- read.dna("sequences.fasta", format = "fasta")

# Mengonversi ke objek phyDat
phydata <- as.phyDat(dna_sequences)

set.seed(42)

# Mencari pohon parsimonius menggunakan metode ratchet
parsimony_tree <- pratchet(phydata, trace = 0)

# Menghitung panjang cabang
parsimony_tree <- acctran(parsimony_tree, phydata)

# Menghapus cabang internal yang terlalu kecil
parsimony_tree <- di2multi(parsimony_tree)

# Menyimpan dan menampilkan pohon
write.tree(parsimony_tree, file = "parsimony_tree.nwk")
plot(parsimony_tree)
```

# MAXIMUM PARSIMONY METHOD

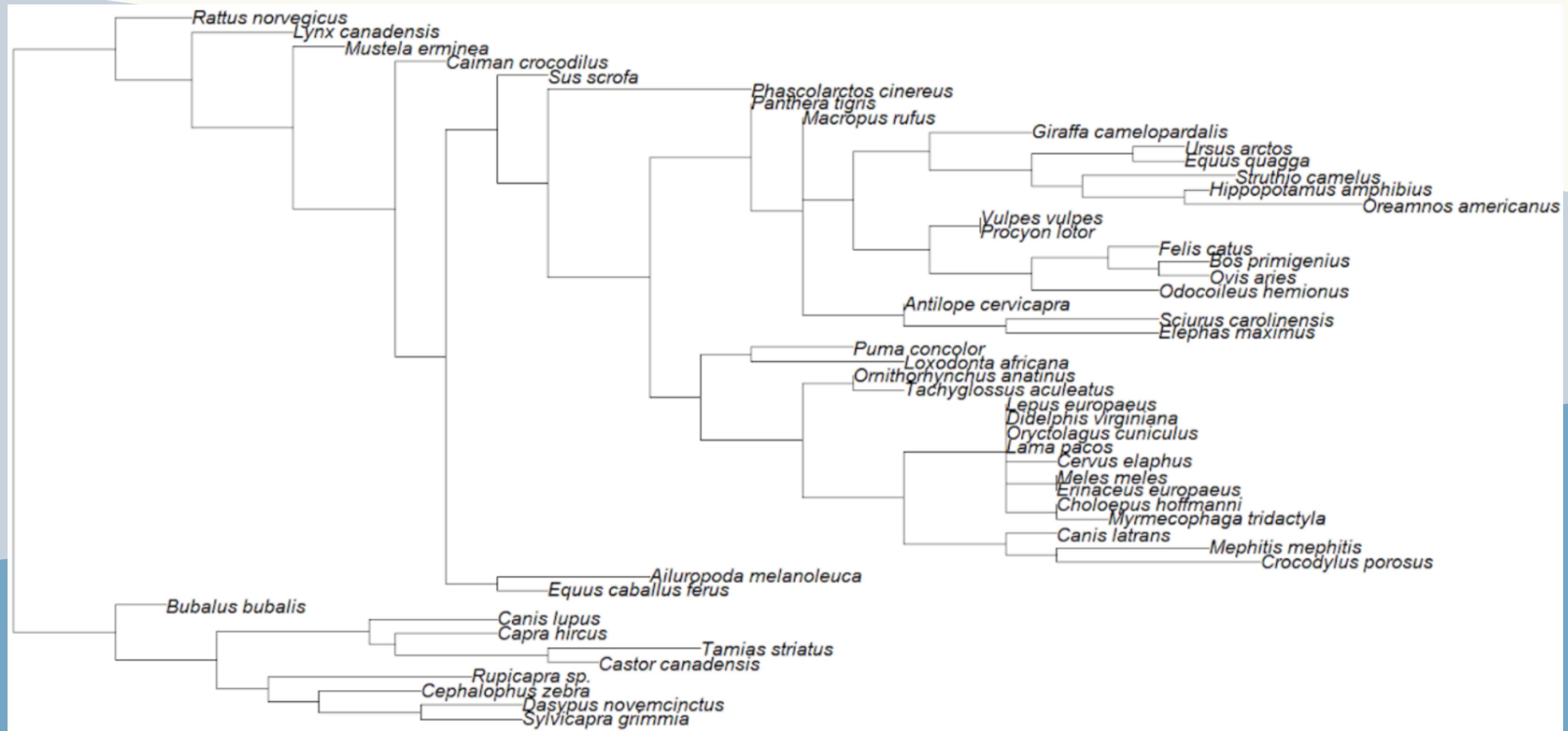
# MAXIMUM PARSIMONY METHOD

## Penjelasan Code

Kode ini menggunakan paket **ape** dan **phangorn** dalam R untuk melakukan analisis filogenetik pada data sekuens DNA. Data sekuens disimpan dalam format FASTA, di mana setiap sekuens ditandai dengan nama spesies. Kode menyimpan data ke file FASTA ("sequences.fasta") dan membacanya kembali sebagai objek R menggunakan fungsi `read.dna` dari paket **ape**, yang memungkinkan manipulasi dan analisis sekuens DNA. Selanjutnya, data sekuens dikonversi menjadi objek **phyDat** melalui **phangorn** agar siap untuk analisis filogenetik, termasuk perhitungan parsimony dan pembangkitan pohon.

Analisis selanjutnya mencakup pencarian pohon yang paling parsimonious menggunakan metode **ratchet**, dilanjutkan dengan perhitungan panjang cabang pohon menggunakan `acctran`. Cabang-cabang internal yang sangat kecil kemudian diubah menjadi polytomy untuk menyederhanakan pohon dan mengurangi noise. Akhirnya, pohon yang dihasilkan disimpan dalam format Newick ("parsimony\_tree.nwk") menggunakan `write.tree`, dan divisualisasikan melalui `plot` untuk menunjukkan hubungan antar spesies dalam bentuk grafis.

# POHON PARSIMOIIOUS MP METHOD



# MAXIMUM PARSIMONY METHOD

## Penjelasan Pohon

Pohon filogenetik dari analisis *Maximum Parsimony* menggambarkan hubungan evolusi antar spesies berdasarkan sekuen DNA, di mana simpul mewakili nenek moyang bersama dan cabang menunjukkan kedekatan evolusi. Spesies yang terletak lebih dekat pada cabang, seperti *Canis lupus* (serigala) dan *Canis latrans* (coyote), memiliki hubungan evolusi lebih dekat dibanding spesies lain, sedangkan cabang lebih panjang menunjukkan lebih banyak perbedaan genetik atau waktu evolusi yang lebih lama, misalnya antara *Elephas maximus* (gajah Asia) dan *Loxodonta africana* (gajah Afrika). Nilai dukungan *bootstrap* yang tinggi pada cabang menunjukkan keandalan hubungan tersebut. Pohon ini berguna dalam memahami hubungan evolusi, mengidentifikasi spesies yang mirip, serta membantu dalam bidang konservasi, ekologi, dan biologi molekuler, memberikan wawasan tentang interaksi dan perkembangan spesies seiring waktu.

# MAXIMUM LIKELIHOOD METHOD PADA PYTHON

## Import Library

```
1 !pip install biopython
```

Adalah library Python yang dirancang untuk analisis bioinformatika dan biologi komputasi.

```
from Bio import Phylo, AlignIO
from Bio.Phylo.TreeConstruction import DistanceCalculator, DistanceTreeConstructor
from Bio.Align import MultipleSeqAlignment
from Bio.Seq import Seq
from Bio.SeqRecord import SeqRecord
import matplotlib.pyplot as plt
```

< Bio.AlignIO > :

untuk membaca format sekuens yang akan dianalisis

< Bio. Phylo > :

untuk menggambar pohon dan bekerja dengan objek pohon filogenetik

< Bio.Pyhlo.TreeConstruction > :

untuk melakukan konstruksi pohon dengan menggunakan metode Maximum Likelihood

## Mengonversi Data DNA

```
sequences = [SeqRecord(Seq(seq), id=name) for name, seq in sequences.items()]
alignment = MultipleSeqAlignment(sequences)
```

Data DNA yang telah diberikan dikonversi menjadi Multiple Sequence Alignment. Hal ini diperlukan untuk menghitung jarak antar sekuens.

## Menghitung Jarak

```
calculator = DistanceCalculator("identity")
distance_matrix = calculator.get_distance(alignment)
```

Penghitungan jarak seharusnya memakai model Jukes-Cantor, tetapi pada kode ini, digunakan 'identity' (perbandingan langsung). Metode Jukes-Cantor dapat diterapkan pada PhyML.

## Membangun Pohon

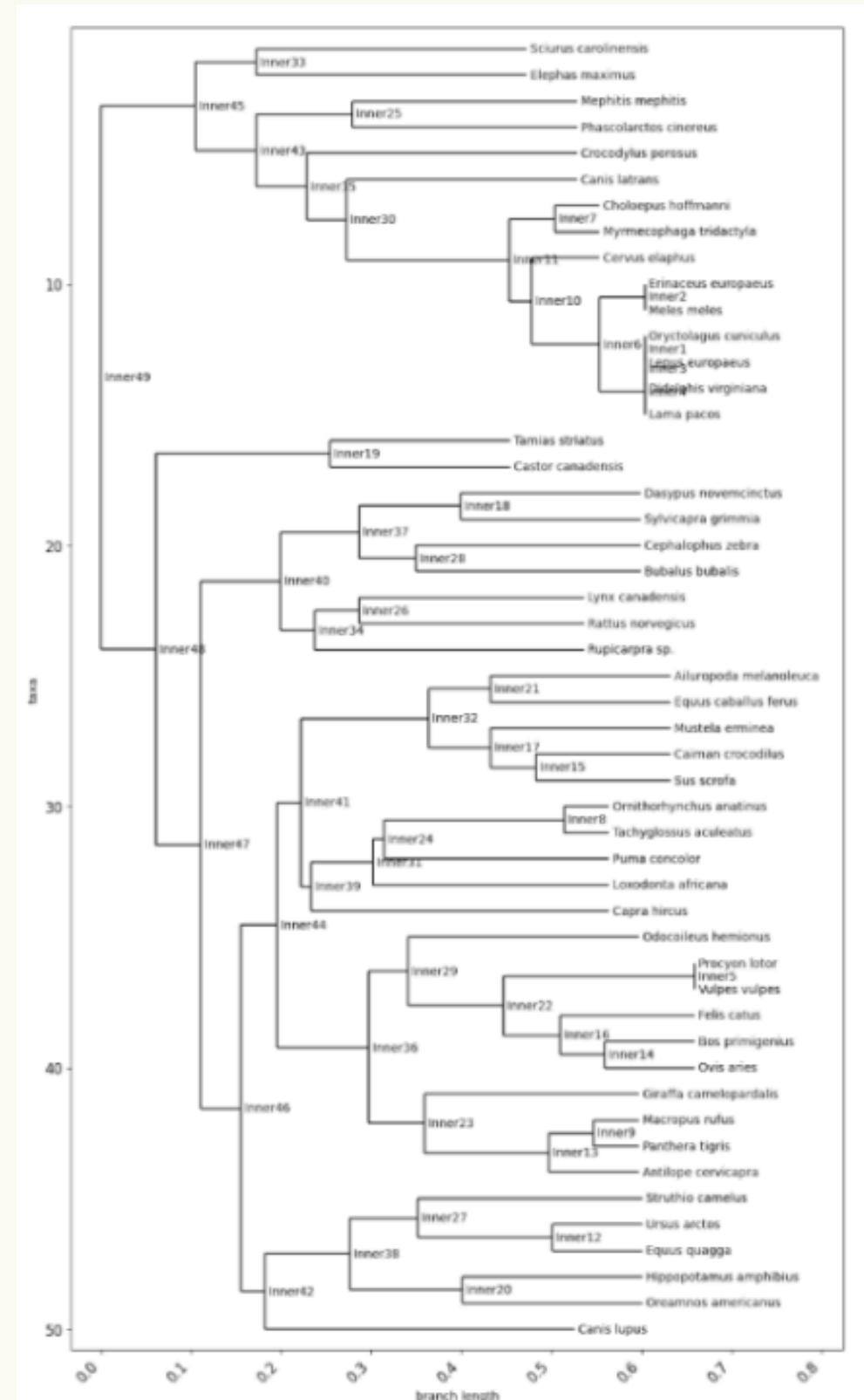
```
constructor = DistanceTreeConstructor()  
phylo_tree = constructor.upgma(distance_matrix)
```

< DistanceTreeConstructor > dengan metode UPGMA (Unweighted Pair Group Method with Arithmetic Mean) yang sederhana. Untuk lebih spesifik dengan metode Maximum Likelihood dapat digunakan PhyML.

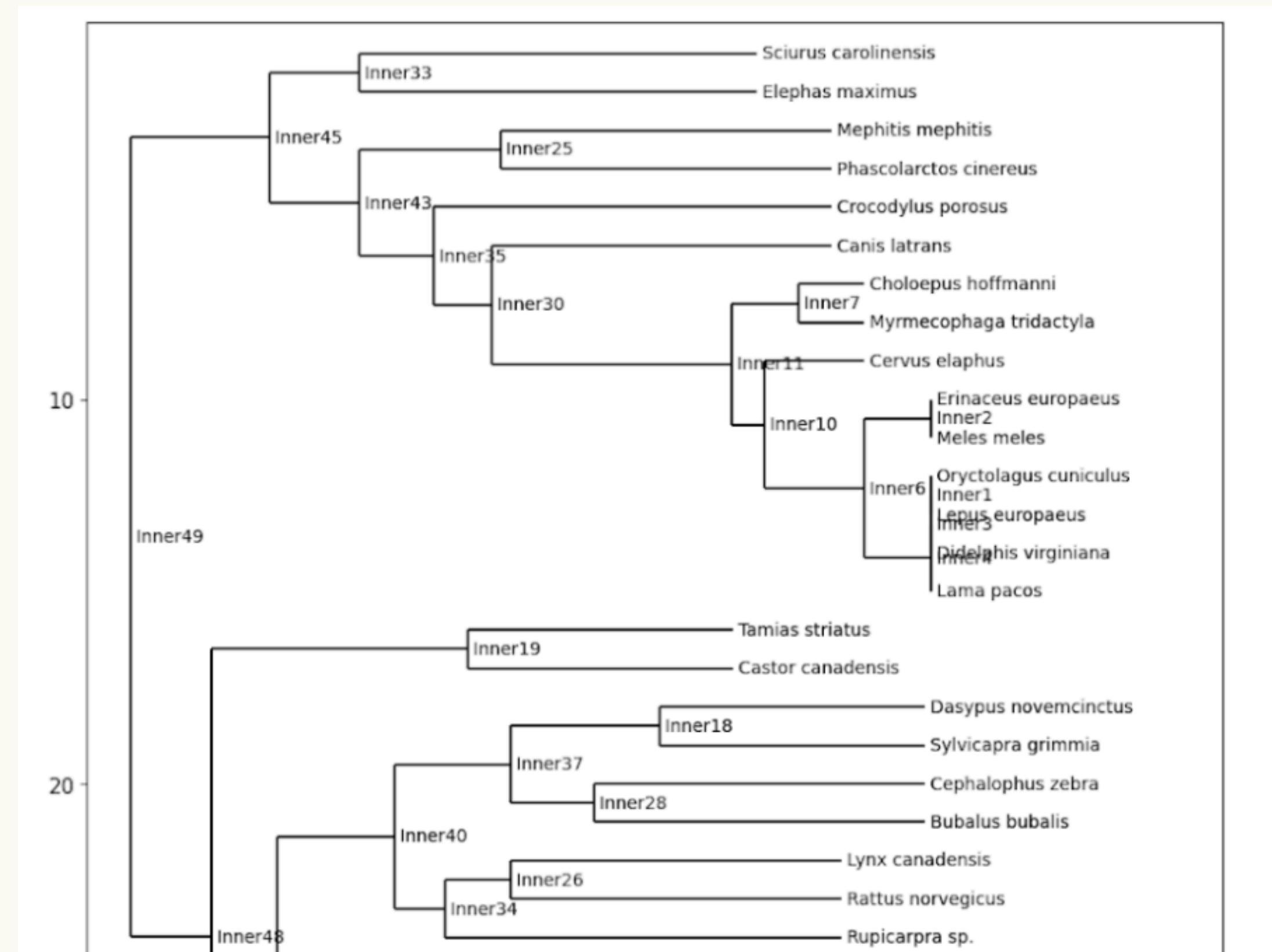
## Menampilkan Pohon dan Mengatur Ukuran

```
fig = plt.figure(figsize=(15, 20))  
ax = fig.add_subplot(1, 1, 1)  
Phylo.draw(phylo_tree, label_func=lambda x: x.name, axes=ax, do_show=False)  
plt.tick_params(axis='both', which='major', labelsize=12)  
plt.xticks(rotation=45, ha='right')  
plt.subplots_adjust(left=0.3)  
plt.show()
```

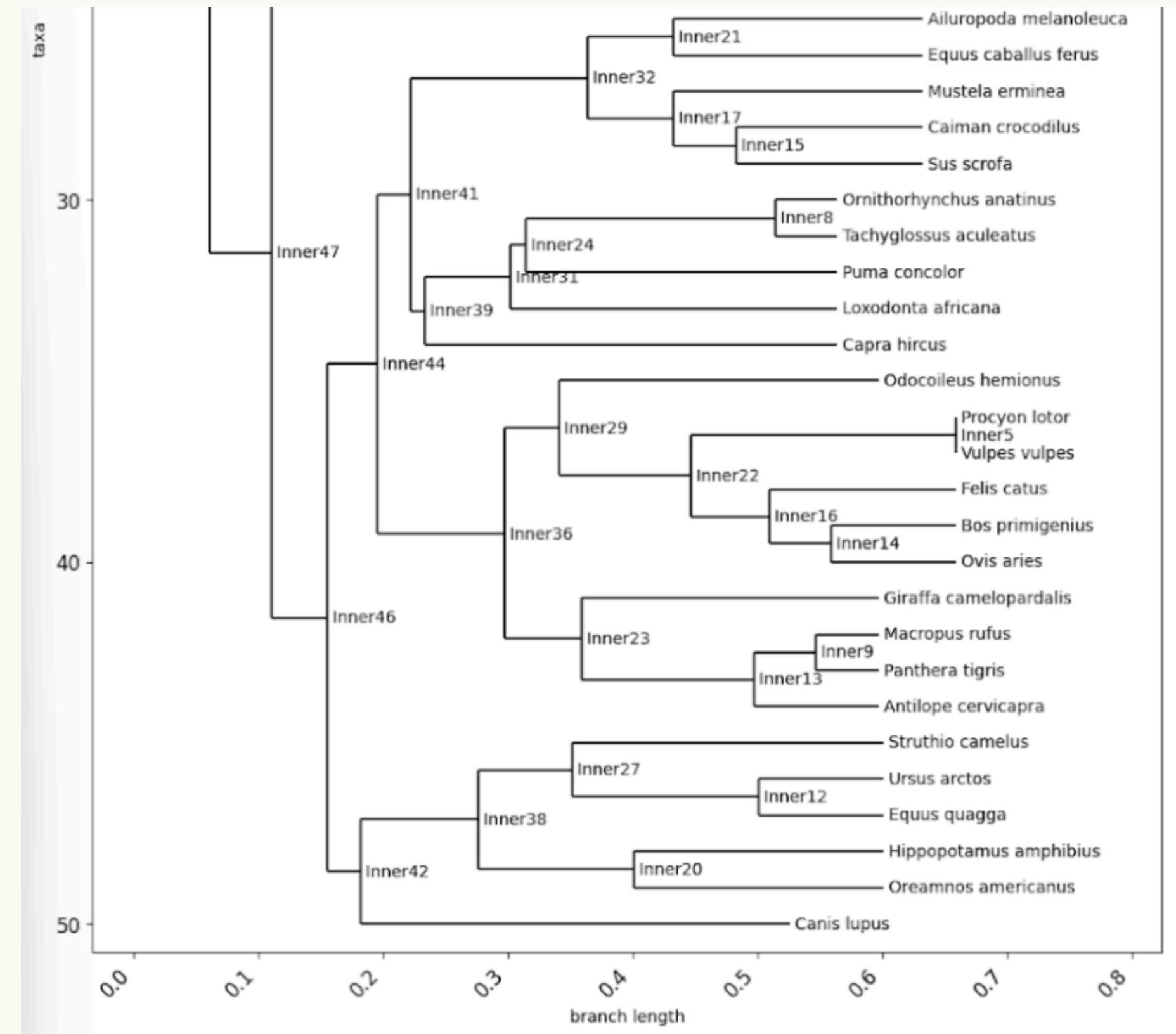
# Output (secara keseluruhan)



## Output



# Output



# MAXIMUM LIKELIHOOD METHOD PADA R

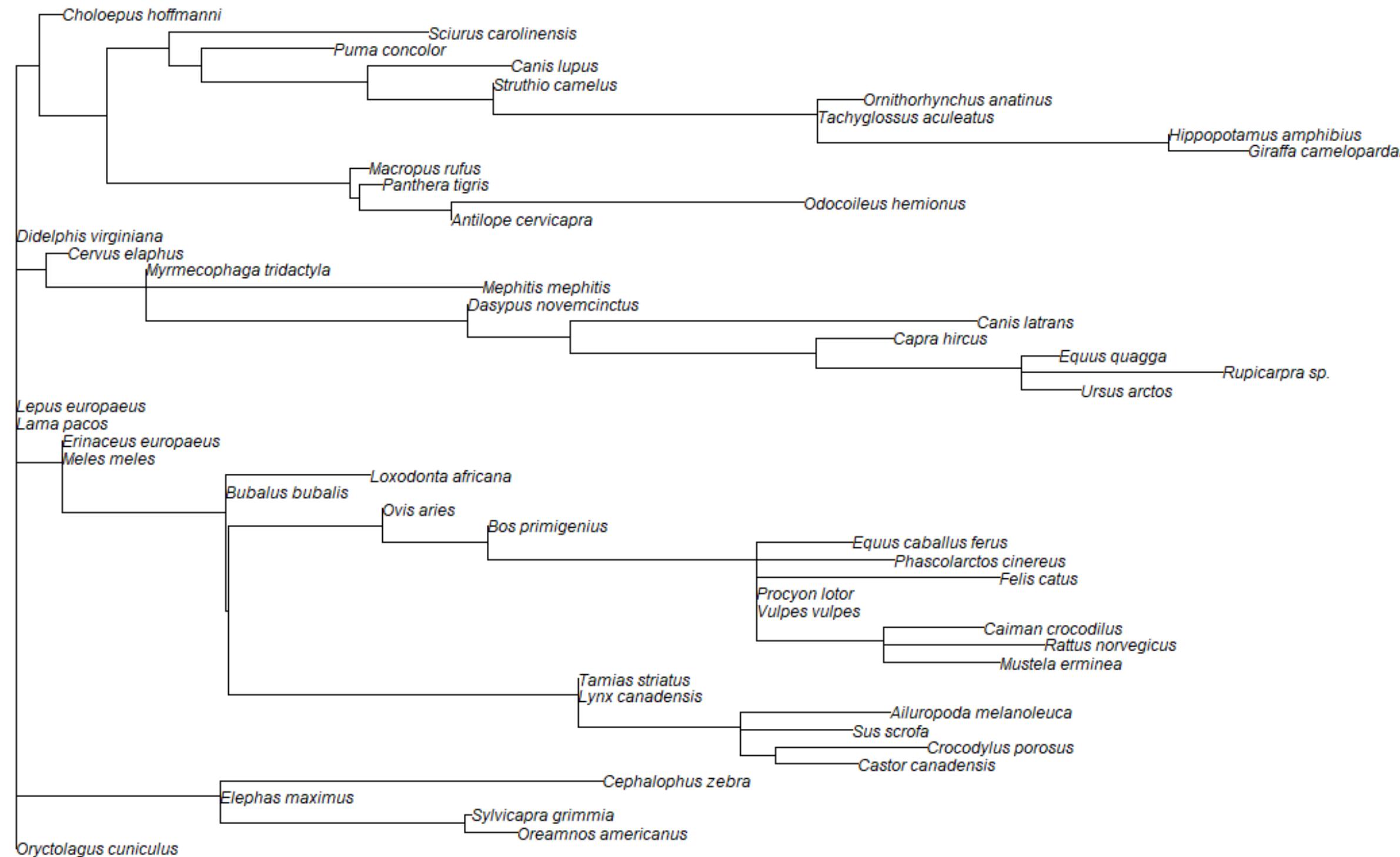
```
library(phangorn)
# transforming the DNAbin format into the phyDat format
phydata <- as.phyDat(dna_K00927)
# Comparing nucleotide substitution models
modeltest <- modelTest(phydata)
# inferring a tree using maximum Likelihood method
pml <- pml_bb(modeltest)
# bootstrap analysis
bstrees <- bootstrap.pml(pml)
# ploting the ML tree with the bootstrap values
plotBS(pml$tree, bstrees)
```

Sumber:

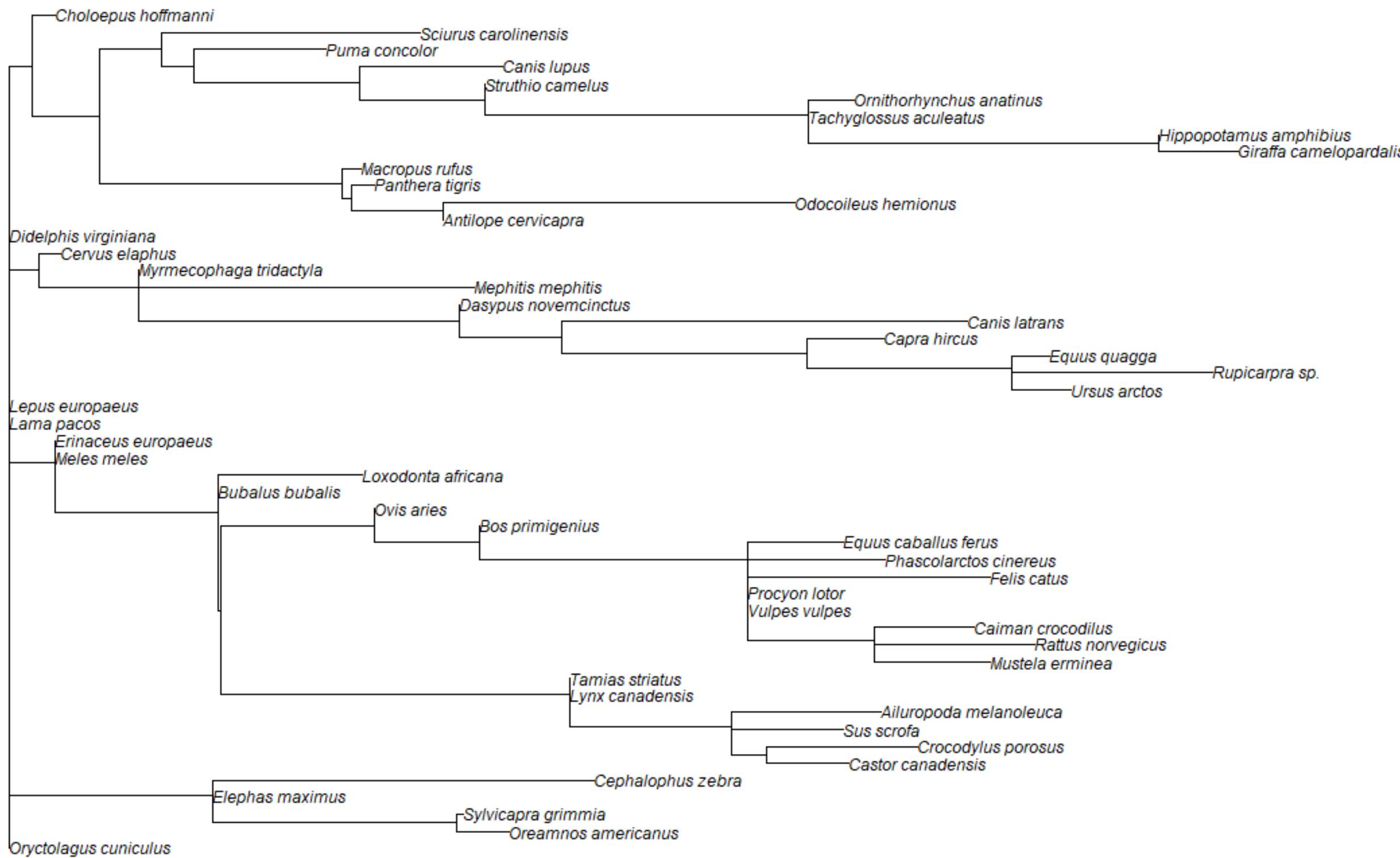
Munjal, G., Hanmandlu, M., & Srivastava, S. (2018). Phylogenetics Algorithms and Applications. Ambient Communications and Computer Systems

```
1 library(ape)
2 library(msa)
3 library(phangorn)
4 library(Biostrings)
5
6 # Membaca file FASTA
7 sequences <- readDNAStringSet("C:/Users/anyel/OneDrive/Pictures/Documents/sequence.txt")
8
9 # Penyelarasan urutan menggunakan msa
10 alignment <- msa(sequences)
11
12 # Mengkonversi ke format phyDat
13 phy_data <- phyDat(as.DNAbin(alignment), type = "DNA")
14
15 # Menghitung matriks jarak
16 dm <- dist.ml(phy_data)
17
18 # Membuat pohon filogenetik
19 tree <- NJ(dm)
20
21 # Memfitting model
22 fit <- pml(tree, phy_data)
23 fit <- optim.pml(fit, model = "K80", optGamma = TRUE, optInv = TRUE)
24
25 # Menggambar pohon
26 png("phylogenetic_tree_k80.png", width = 1200, height = 800) # Ukuran lebar dan tinggi dalam pixel
27 plot(fit$tree, main = "Maximum Likelihood Phylogenetic Tree, Model= Kimura 80",
28       cex.main = 1.5, cex.axis = 0.8) # Atur ukuran teks
29 dev.off()
```

### Maximum Likelihood Phylogenetic Tree, Model= Kimura 80



### Maximum Likelihood Phylogenetic Tree, Model= Kimura 80



## Interpretasi

- Pohon filogenetik ini disertai dengan cabang yang menunjukkan hubungan kekerabatan antar spesies.
- Panjang cabang menunjukkan seberapa besar perbedaan genetik antara spesies.
- Dari pohon ini, dapat diidentifikasi kelompok-kelompok spesies yang lebih erat. Misalnya, *Canis lupus* dan *Canis latrans* yang mungkin memiliki kekerabatan yang lebih dekat dibandingkan dengan spesies lain, seperti *Crocodylus porosus*.
- Titik-titik cabang internal (Inner1, Inner2, dll.) mewakili nenek moyang bersama dari sekelompok spesies. Mereka tidak mewakili spesies yang sebenarnya, tetapi membantu menggambarkan proses evolusi.



```
library(babette)
library(ape)

# running BEAST2 for Bayesian phylogenetic analysis
outputs <- bbt_run_from_model("./data/ntK00927.fas",
                                inference_model = create_inference_model(),
                                beast2_options = create_beast2_options())

# extracting the tree with the highest posterior
treeBI <- outputs$ntK00927_trees[[which.max(outputs$estimates$posterior)]]

# merging all Convergent samples into a List(multiphylo)
con.trees <- outputs$ntK00927_trees[3:10001]

# calculating the posterior of the bifurcations in the BI tree
posterior <- round(prop.clades(treeBI, con.trees)/9999*100)

# Visualizing the BI tree
plot(treeBI)

# adding posterior values of the bifurcations
drawSupportOnEdges(posterior)
```

# BAYESIAN METHOD

```
library(devtools)
#devtools::install_github("thibautjombart/apex")
library(apex)

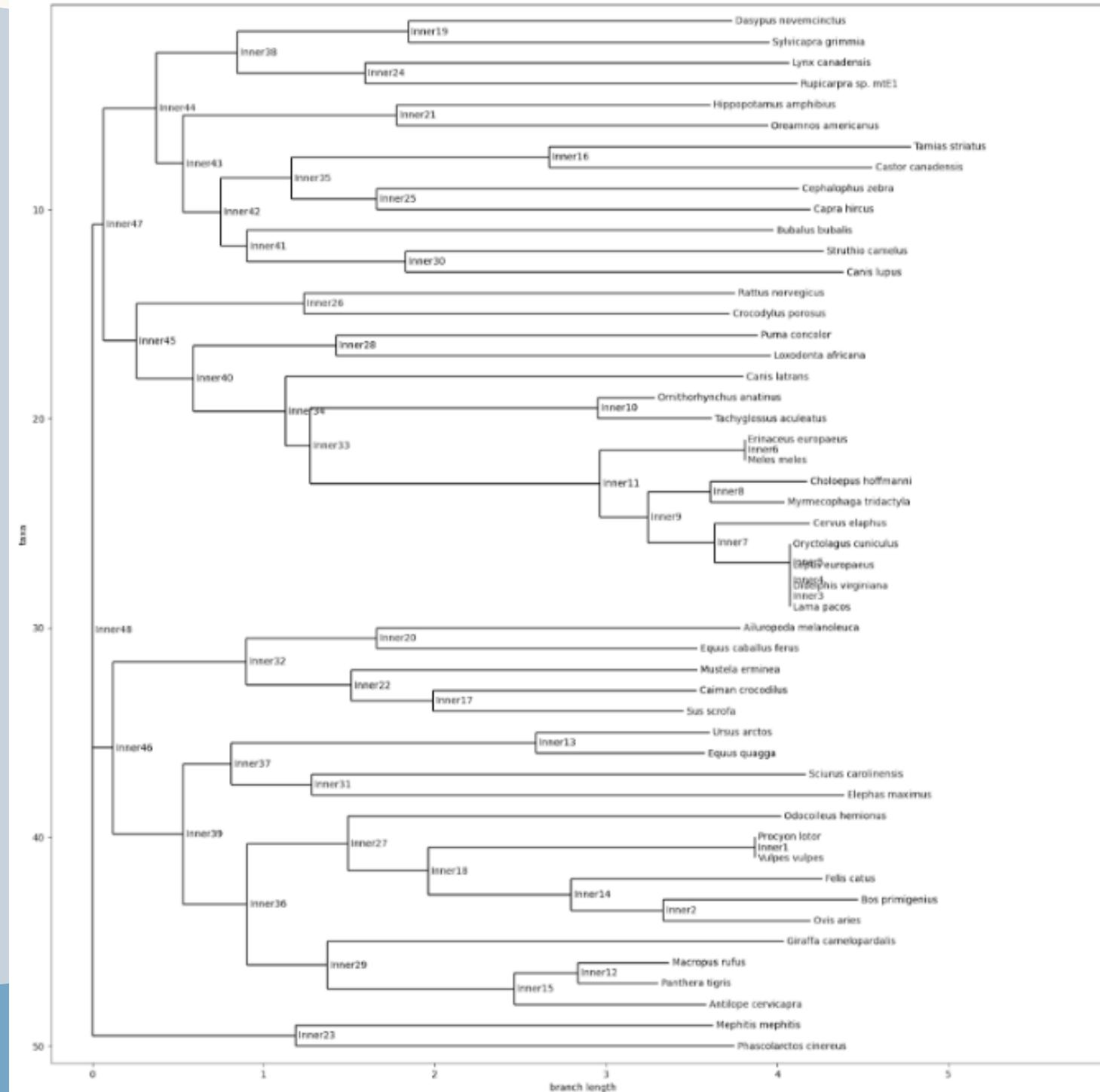
# storing these ten genes in the same list
dna <- c("ntK01939", "ntK03644", "ntK00797", "ntK00826", "ntK00088",
        "ntK02257", "ntK00164", "ntK00820", "ntK06158", "ntK00008")
i <- 0
DNAbin_list <- list()
for (k in dna) {
  i <- i + 1
  dna <- fasta2DNAbin(paste0("./data/", k, ".fas"))
  DNAbin_list[[i]] <- dna
}
# transforming the List of DNAbins into the multidna format
multidna <- new("multidna", DNAbin_list)
# concatenating the DNA alignments into a single alignment matrix
phydata <- as.phyDat(concatenate(multidna))
# Comparing nucleotide substitution models
modeltest <- modelTest(phydata)
# inferring a tree using maximum Likelihood method
pml <- pml_bb(modeltest)
# Visualizing the ML tree
plot(pml$tree)
```

# CONCATENATION METHOD

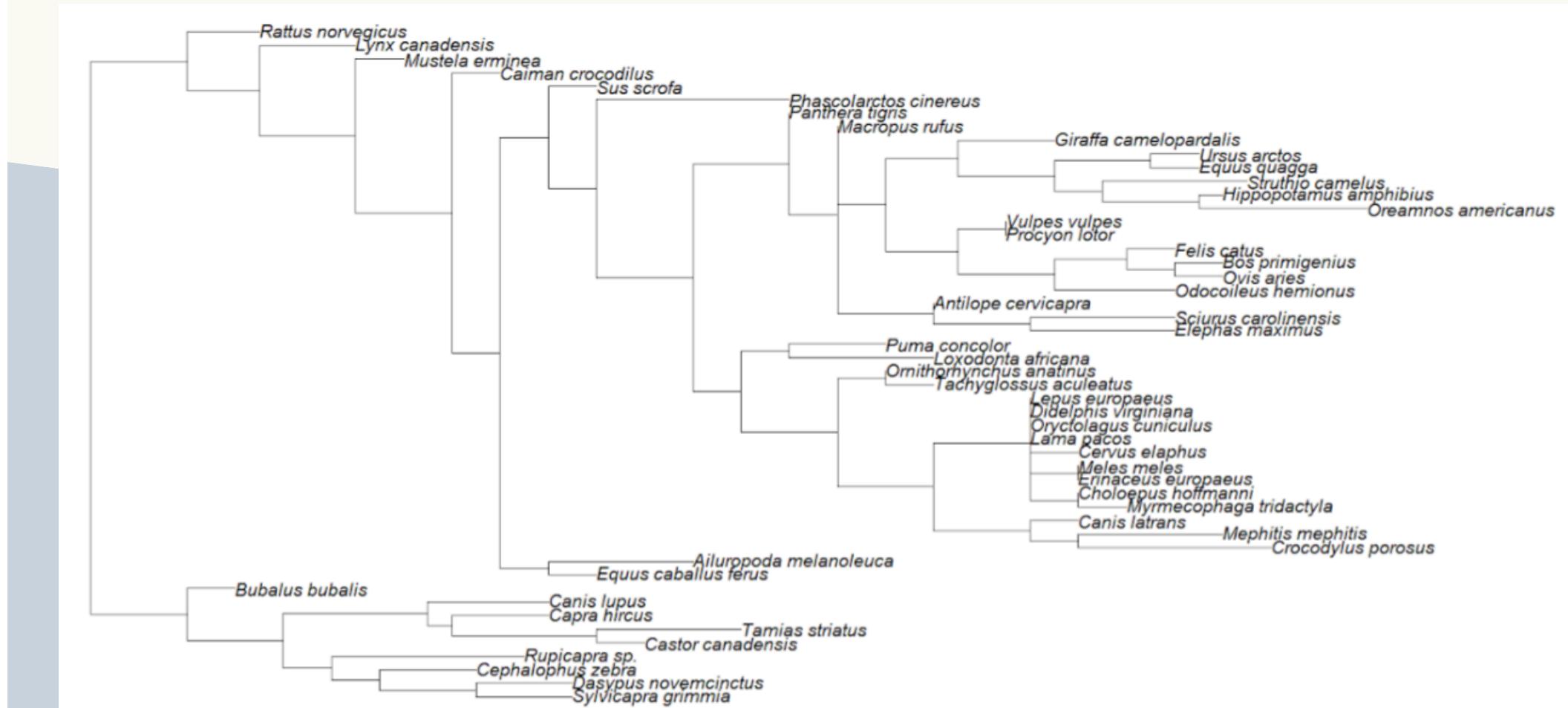
```
library(phangorn)
# constructing single gene trees, storing which in a List(multiphylo)
dna <- c("ntK01939", "ntK03644", "ntK00797", "ntK00826", "ntK00088",
        "ntK02257", "ntK00164", "ntK00820", "ntK06158", "ntK00008")
i <- 0
mtree <- c()
for (k in dna) {
  i <- i + 1
  dna <- fasta2DNAbin(paste0("./data/", k, ".fas"))
  # transforming the DNAbin format into the phyDat format
  phydata <- as.phyDat(dna)
  # Comparing nucleotide substitution models
  modeltest <- modelTest(phydata)
  # inferring a tree using maximum Likelihood method
  pml <- pml_bb(modeltest)
  # storing the ML tree into the List
  mtree[[i]] <- pml$tree
}
class(mtree) <- "multiPhylo"
# estimating a supertree from a set of trees
supertree <- superTree(mtree)
# Visualizing the supertree
plot(supertree)
```

# COALESCENCE METHOD

# PERBANDINGAN HASIL POHON ANTAR METODE

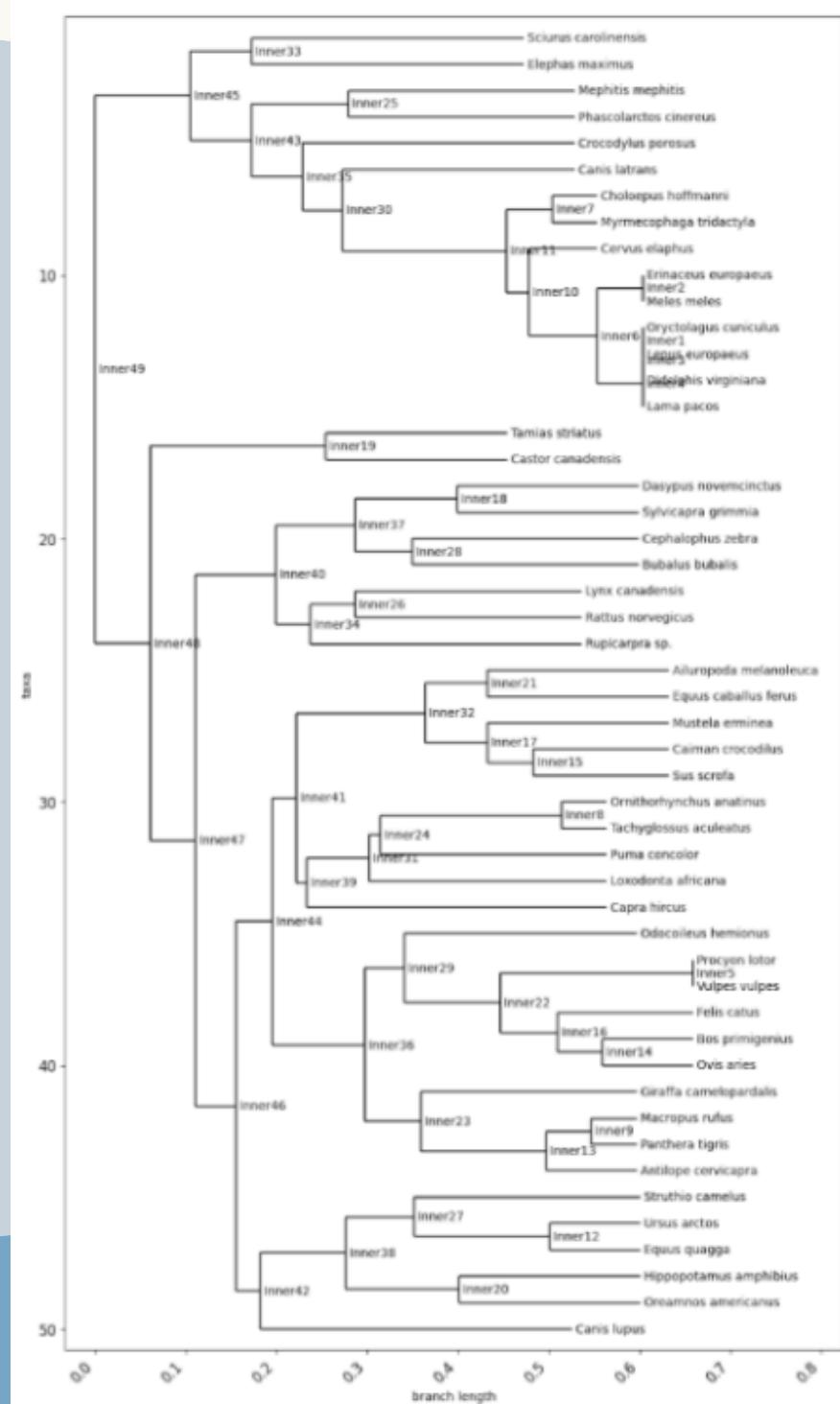


Distance Based



Parsimony Method

# PERBANDINGAN HASIL POHON ANTAR METODE



# PERBANDINGAN HASIL POHON ANTAR METODE

**BAYESIAN INFERENCE:** Seharusnya menjadi yang lebih akurat, namun pada gambar terlihat tidak terlalu memberikan pola cabang yang jelas. Hal ini mungkin disebabkan oleh data barisan DNA yang kurang panjang atau kurang lengkap.

**PARSIMONY:** Terlihat lebih sederhana dan memperlihatkan cabang-cabang kekerabatan, tetapi cenderung mengabaikan perbedaan jarak evolusi yang signifikan di antara spesies yang berbeda.

**MAXIMUM LIKELIHOOD:** Tampak lebih akurat dan bercabang secara sistematis, cabang yang dihasilkan cenderung mencerminkan model evolusi.

**DISTANCE- BASED:** Membuat jarak cabang yang konsisten, tetapi hanya memperhitungkan perbedaan keseluruhan tanpa mempertimbangkan perubahan karakter evolusi.

# KESIMPULAN & SARAN

## KESIMPULAN

Rekonstruksi filogenetik penting untuk memahami hubungan evolusi spesies melalui data genetik, morfologis, dan molekuler. Teknik seperti Maximum Parsimony (MP), Maximum Likelihood (ML), dan Bayesian Inference (BI) membantu menyusun pohon filogenetik akurat dengan kelebihan dan kekurangan masing-masing. MP sederhana, tetapi kurang menggambarkan kompleksitas evolusi, sedangkan ML dan BI lebih akurat namun memerlukan waktu komputasi tinggi. Pendekatan konkatenasi dan koalesensi menambah kedalaman analisis dengan menyelaraskan perbedaan antara pohon gen dan pohon spesies, memungkinkan penelitian lebih mendalam di biologi dan ekologi.

## SARAN

Untuk meningkatkan rekonstruksi filogenetik, gunakan data genomik lengkap dan pilih metode sesuai karakteristik dataset; koalesensi untuk data dengan heterogenitas tinggi dan konkatenasi untuk data homogen. Kombinasi metode seperti Maximum Likelihood dan Bayesian Inference dapat meningkatkan akurasi. Pengembangan perangkat lunak dan algoritma baru juga penting untuk efisiensi komputasi pada analisis dataset besar.

# REFERENSI

- Edwards, S. V., Liu, L., & Pearl, D. K. (2007). "High-resolution species trees without concatenation." *Proceedings of the National Academy of Sciences*, 104(14), 5936–5941. <https://doi.org/10.1073/pnas.0607004104>
- Felsenstein, J. (2004). *Inferring Phylogenies*. Sinauer Associates.
- Isaev, A. (2004). *Introduction to mathematical methods in bioinformatics*. Berlin German: Springer.
- Lemey, P., Salemi, M., & Vandamme, A. M. (Eds.). (2009). *The Phylogenetic Handbook: A Practical Approach to Phylogenetic Analysis and Hypothesis Testing*. Cambridge University Press.
- Mirarab, S., & Warnow, T. (2015). "ASTRAL-II: Coalescent-based species tree estimation with many hundreds of taxa and thousands of genes." *Bioinformatics*, 31(12), i44–i52. <https://doi.org/10.1093/bioinformatics/btv234>
- Munjal, G., Hanmandlu, M., & Srivastava, S. (2018). *Phylogenetics Algorithms and Applications*. Ambient Communications and Computer Systems: RACCCS-2018, 904, 187–194. [https://doi.org/10.1007/978-981-13-5934-7\\_17](https://doi.org/10.1007/978-981-13-5934-7_17)
- Zou, Y., Zhang, Z., Zeng, Y., Hu, H., Hao, Y., Huang, S., & Li, B. (2024). Common Methods for Phylogenetic Tree Construction and Their Implementation in R. *Bioengineering* (Basel, Switzerland), 11(5), 480. <https://doi.org/10.3390/bioengineering11050480>