

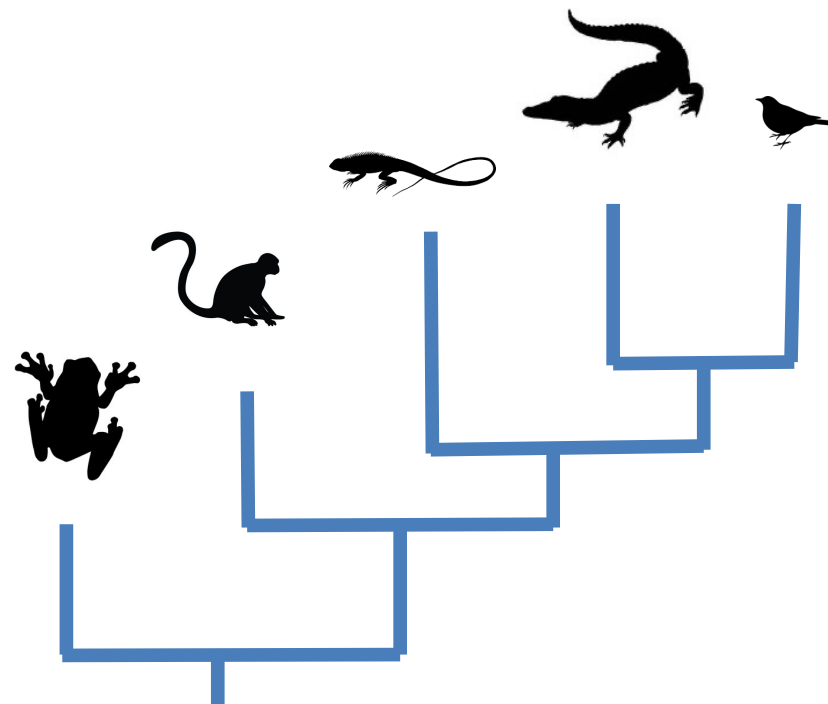
Para que servem as árvores filogenéticas?

As árvores representam a diversificação das espécies ao longo do tempo

Eles desempenham um papel predominante na biologia moderna e são conhecimentos básicos não apenas em Sistemática, mas também para abordar questões em diversos campos.

- Biologia molecular
- Genética e epigenética
- Biologia do desenvolvimento e Evo-Devo
 - Ecologia
- Epidemiologia e Medicina

Árvores de espécies



Árvores genéticas

```
ATCAAAATTTGGCGGC  
ATCAAAATTTGGCGGC  
ATCAAAATTTGGCGGC  
ATCAAAATTTGGCGGC  
ATCAAAATTTGGCGGC
```

Locus 1

```
TTGCAAAGGGGCGCGGGCC  
TTGCAAAGGGGCGCGGGCC  
TTGCAAAGGGGCGCGGGCC  
TTGCAAAGGGGCGCGGGCC  
TTGCAAAGGGGCGCGGGCC
```

Locus 2

...

```
CCTGCTCCCGGTCGTA  
CCTGCTCCCGGTCGTA  
CCTGCTCCCGGTCGTA  
CCTGCTCCCGGTCGTA  
CCTGCTCCCGGTCGTA
```

Locus *n*



Sequências de DNA

Locus 1

```
ATCAAAATTTGGCGGC
ATCAAAATTTGGCGGC
ATCAAAATTTGGCGGC
ATCAAAATTTGGCGGC
ATCAAAATTTGGCGGC
```

Locus 2

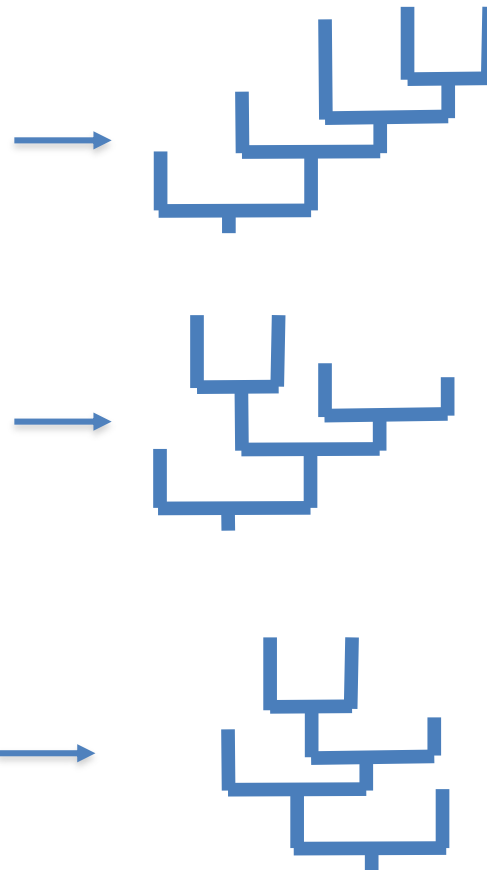
```
TTGCAAAGGGGCGC
TTGCAAAGGGGCGC
TTGCAAAGGGGCGC
TTGCAAAGGGGCGC
TTGCAAAGGGGCGC
```

⋮

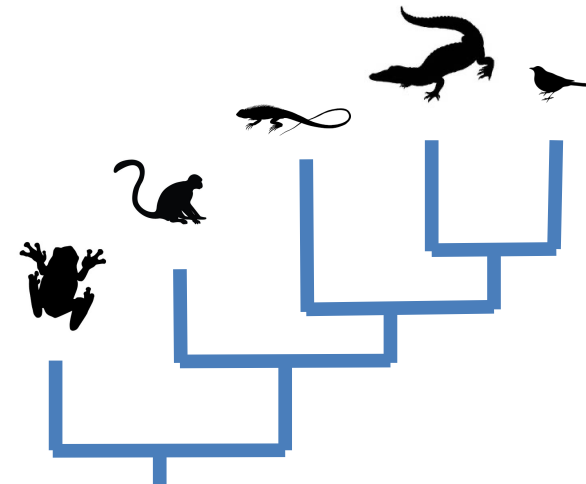
Locus n

```
CCTGCTCCCGGTCGTA
CCTGCTCCCGGTCGTA
CCTGCTCCCGGTCGTA
CCTGCTCCCGGTCGTA
CCTGCTCCCGGTCGTA
```

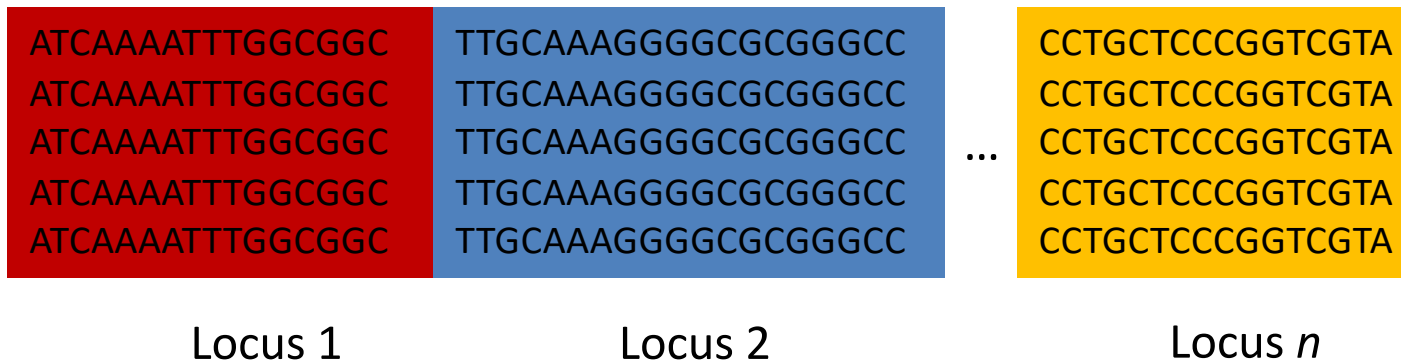
Árvores genéticas independentes



Árvores de espécies



Concatenação de loci independentes

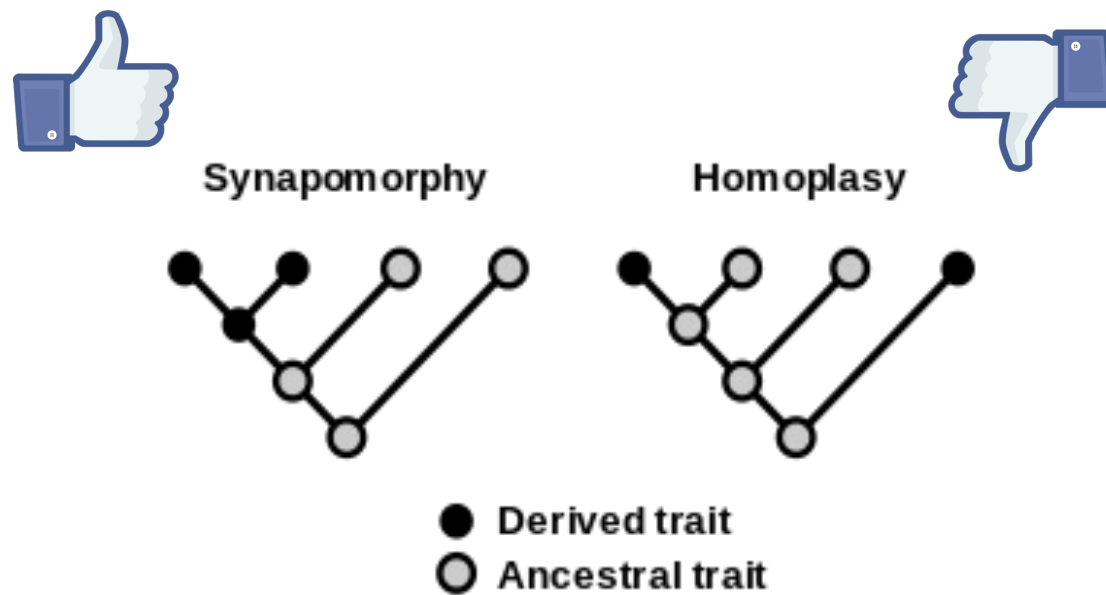


Não faça isso!!!

Métodos para reconstruções filogenéticas

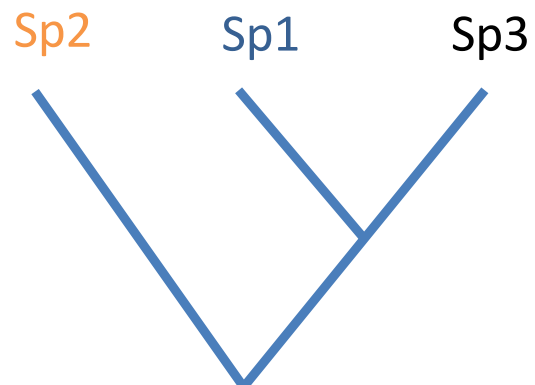
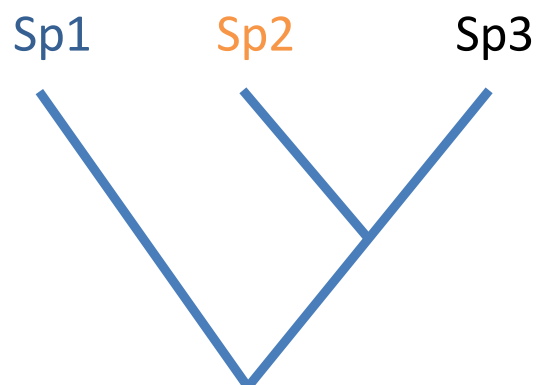
1. Árvores genéticas: locus de comprimento necessário
 - Parcimônia
 - Maximum likelihood
 - Bayesiano
2. Árvores de espécies: baseadas em árvores genéticas (locus de comprimento) ou polimorfismos de nucleotídeo único (SNPs)

Maximum Parsimony

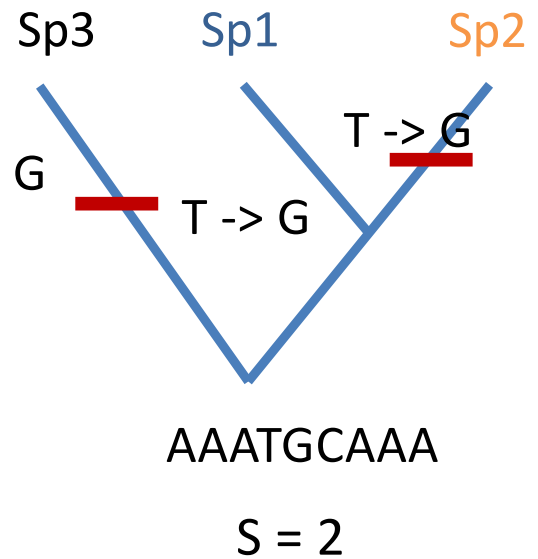
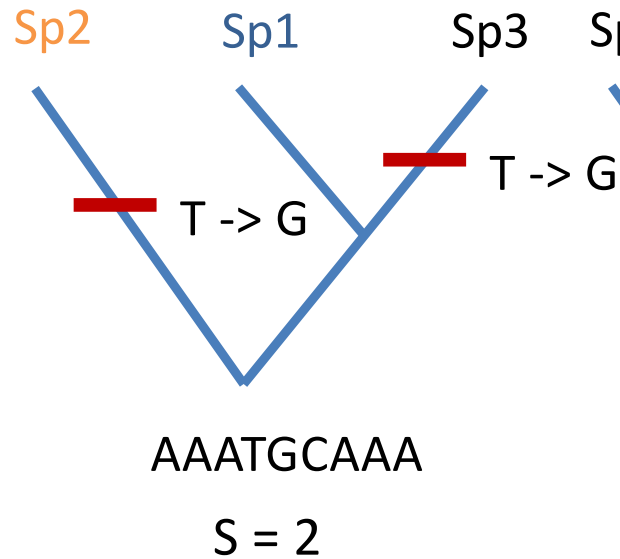
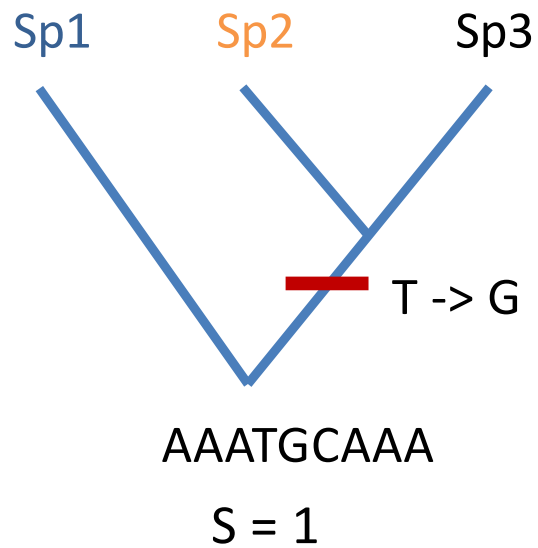


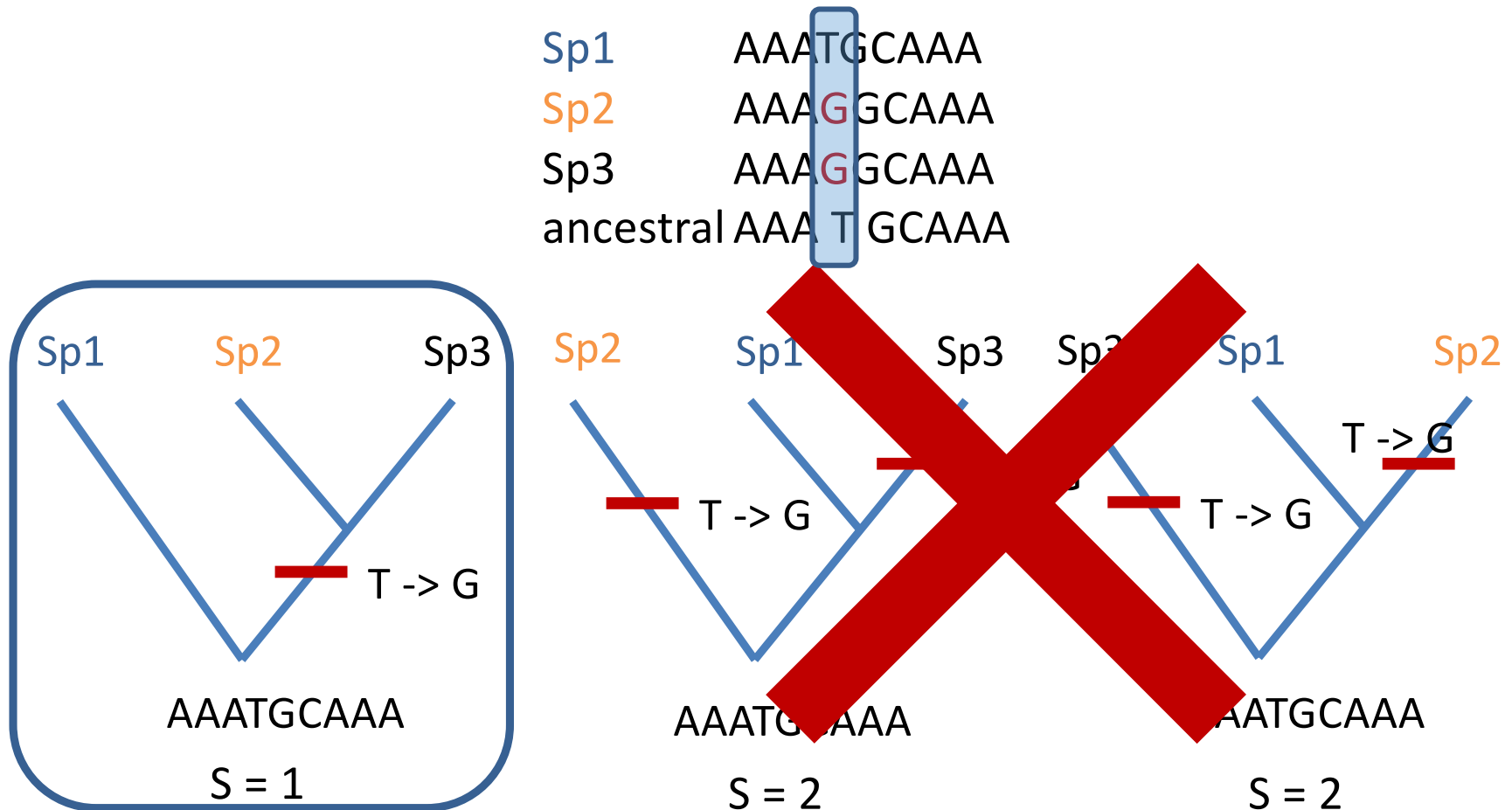
Assume que a homoplasia é estranha

Sp1	AAATGCAAA
Sp2	AAA G GCAAA
Sp3	AAA G GCAAA
ancestral	AAA T GCAAA

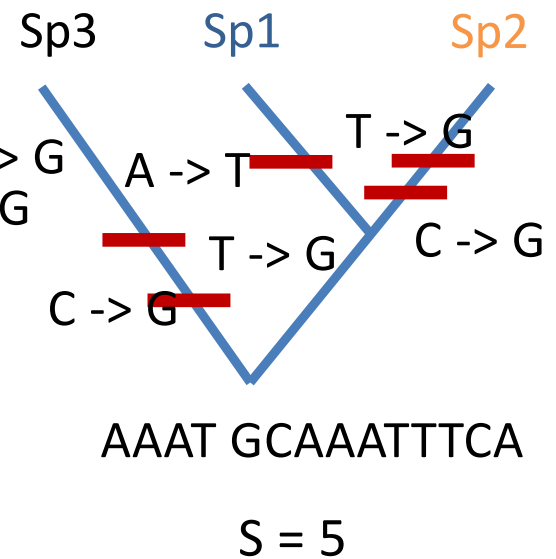
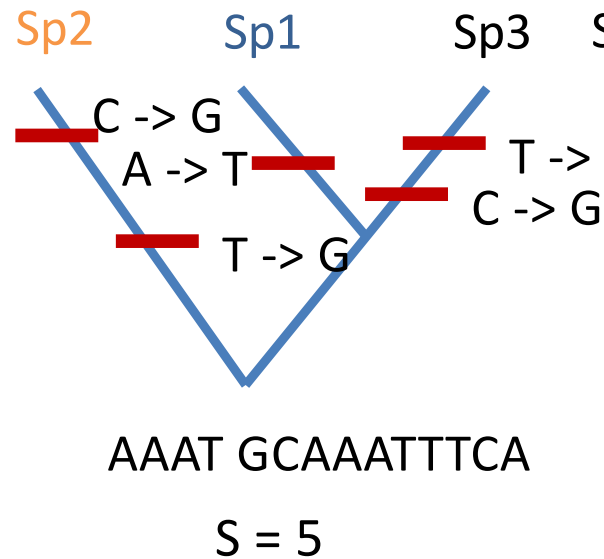
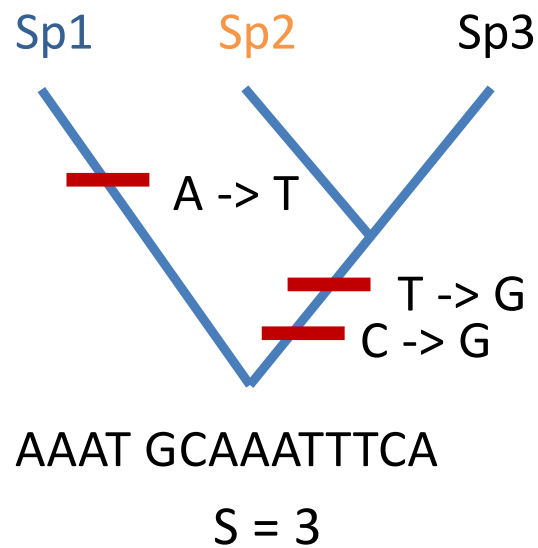


Sp1	AAATGCAAA
Sp2	AAAGGCAAA
Sp3	AAAGGCAAA
ancestral	AAATGCAAA

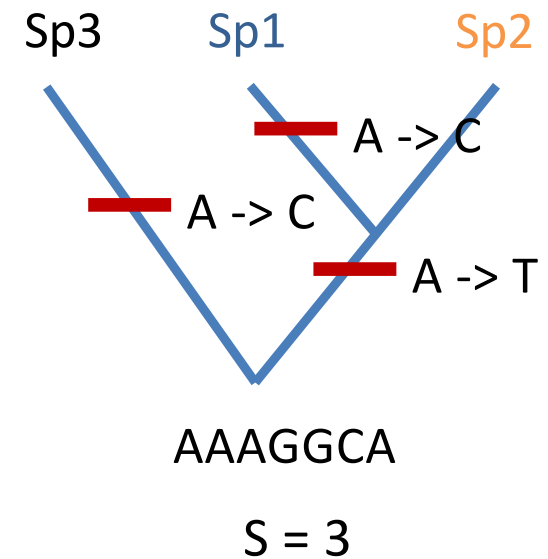
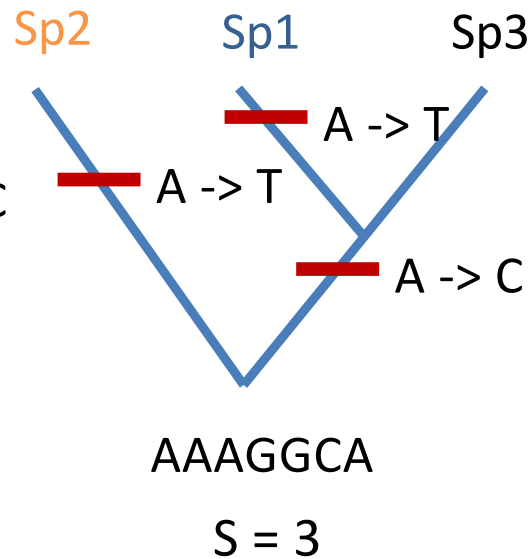
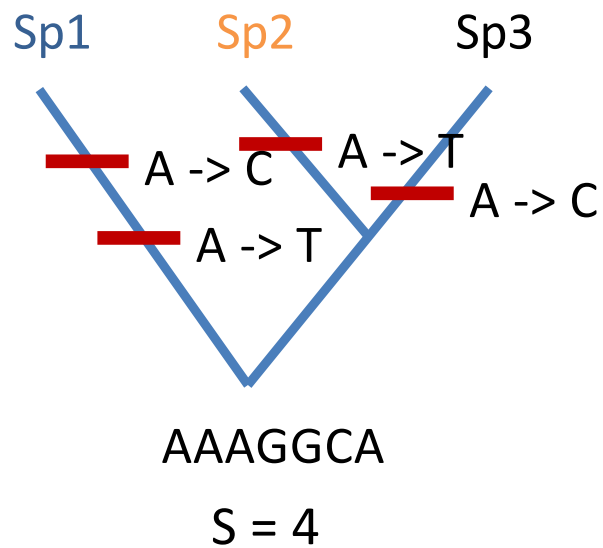




Sp1	AAATGCAAA TTTCT
Sp2	AAA G GCAAATT G A
Sp3	AAA G GCAAATT G A
ancestral	AAAT GCAAATTCA



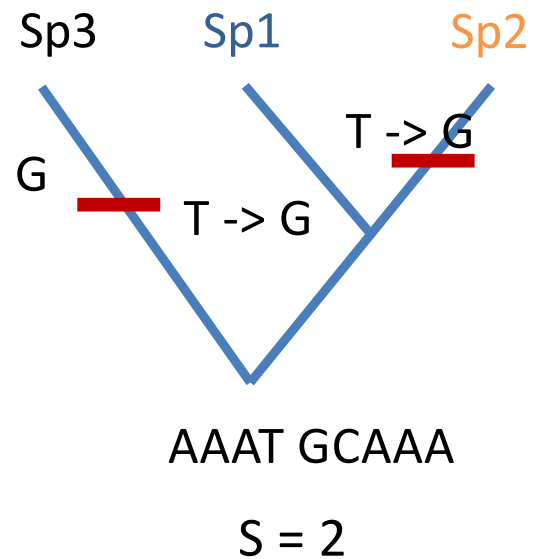
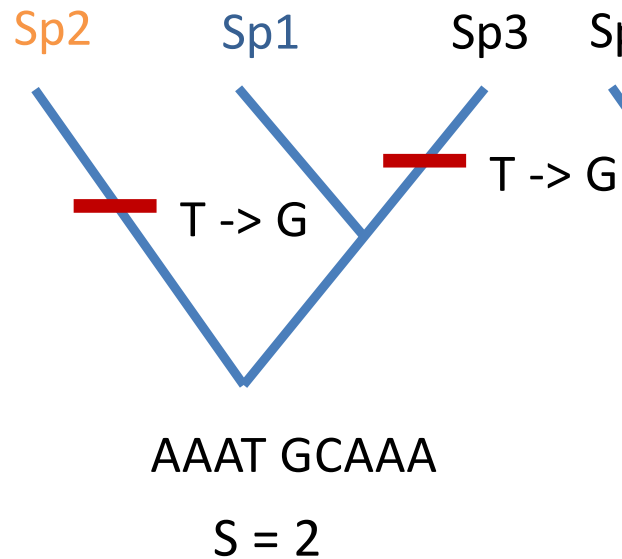
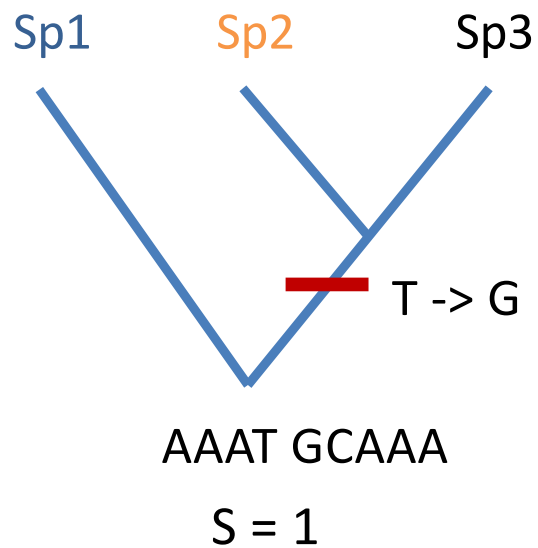
Sp1	AT CGGCA
Sp2	AT AGGCA
Sp3	AACGGCA
ancestral	AAAGGCA





Sua vez!!!

Sp1 AAATGCAAA
 Sp2 AAAGCAAA
 Sp3 AAAGCAAA
 ancestral AAAT GCAAA

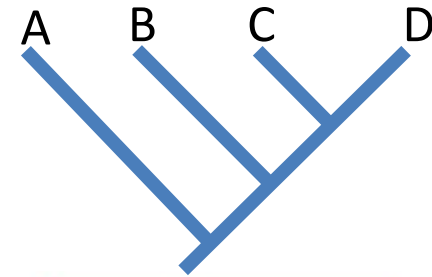


Cálculo Bootstrap

Cálculo Bootstrap

Original data

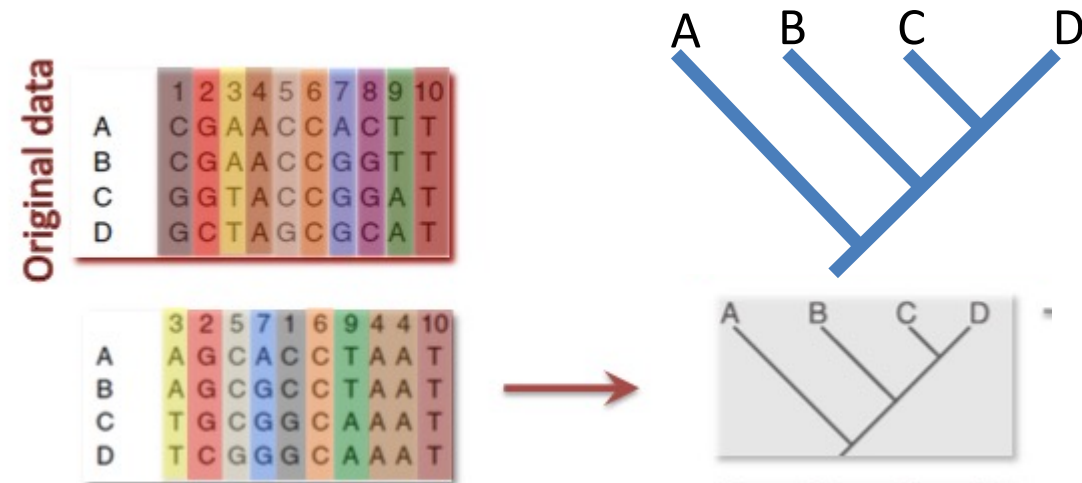
	1	2	3	4	5	6	7	8	9	10
A	C	G	A	A	C	C	A	C	T	T
B	C	G	A	A	C	C	G	G	T	T
C	G	G	T	A	C	C	G	G	A	T
D	G	C	T	A	G	C	G	C	A	T



Bootstrap replicates

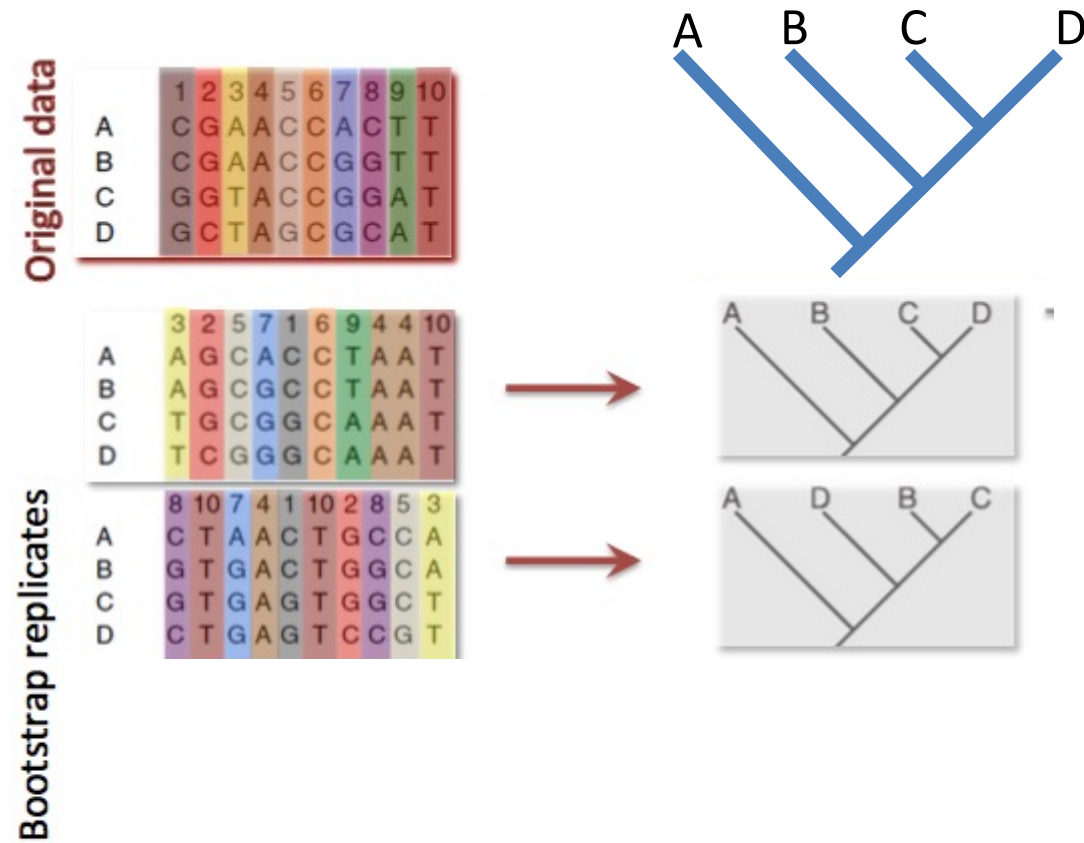
Felsenstein (1985)

Cálculo Bootstrap



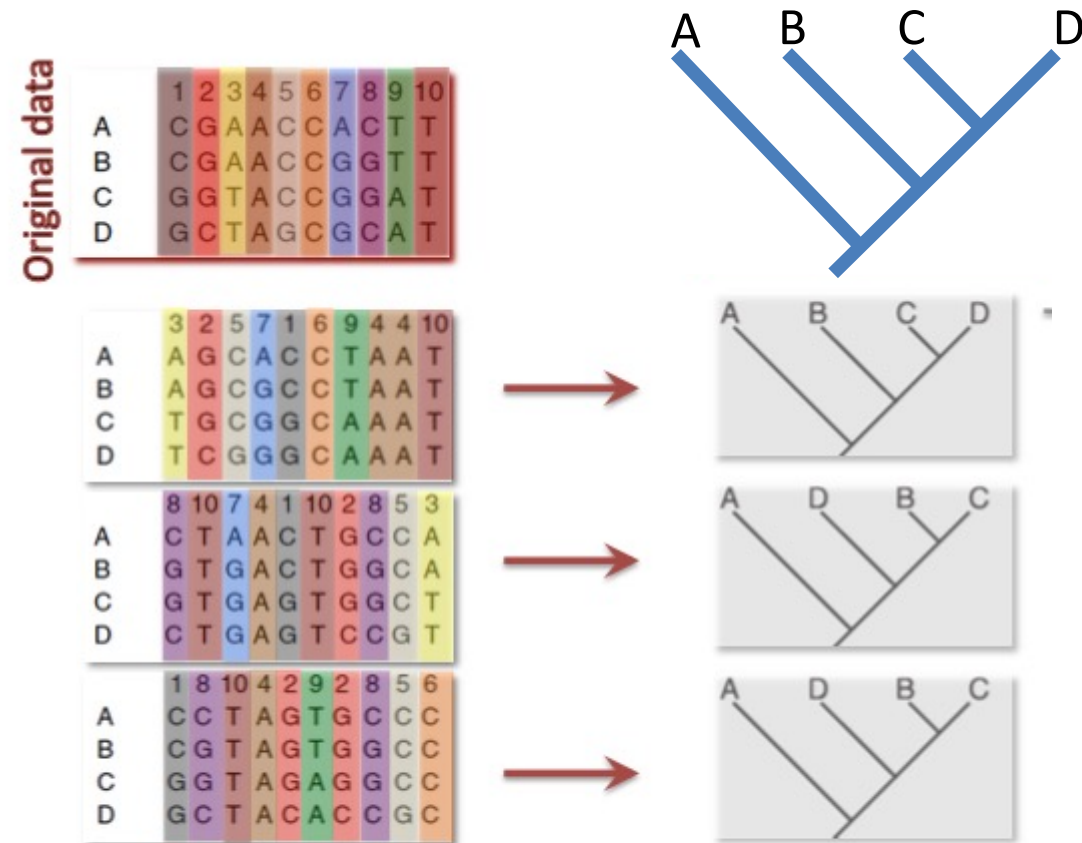
Felsenstein (1985)

Cálculo Bootstrap



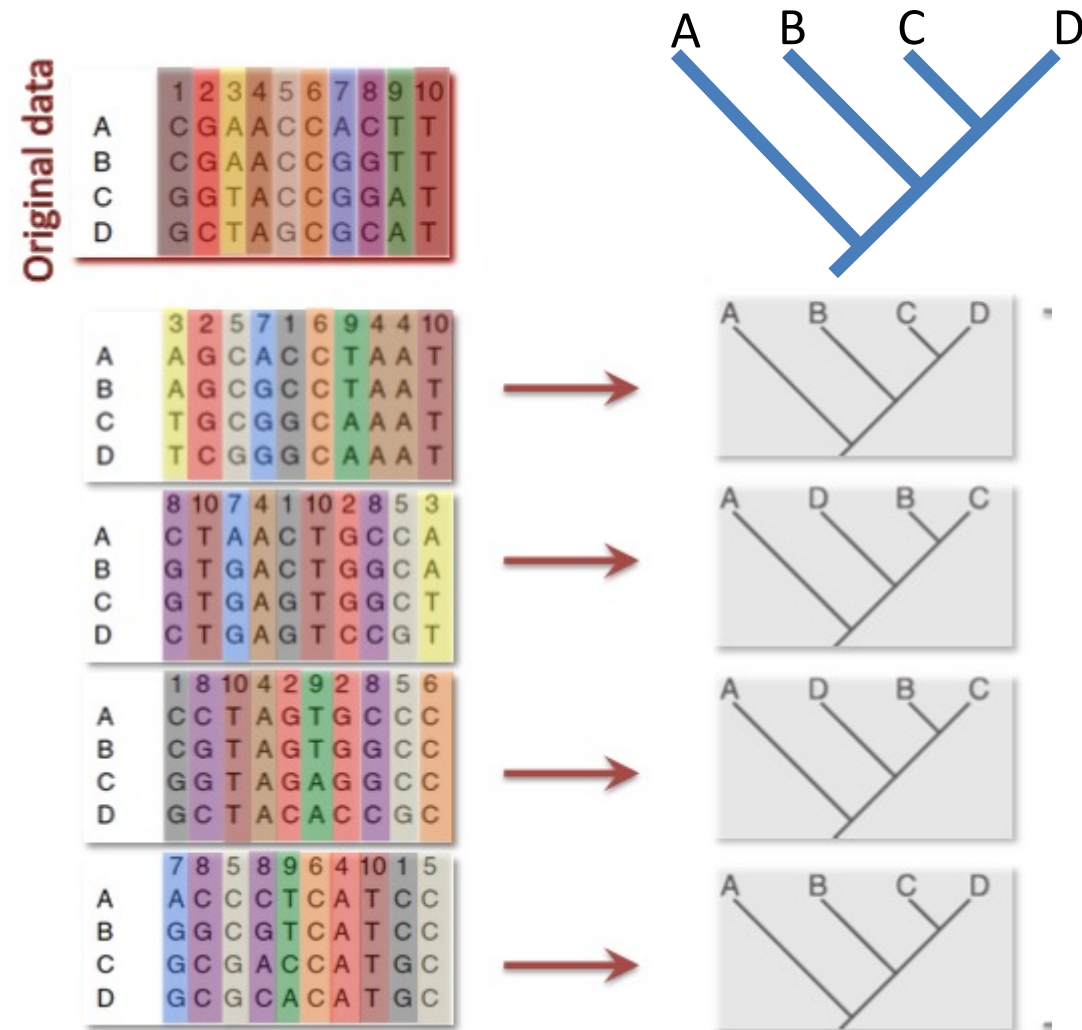
Felsenstein (1985)

Cálculo Bootstrap

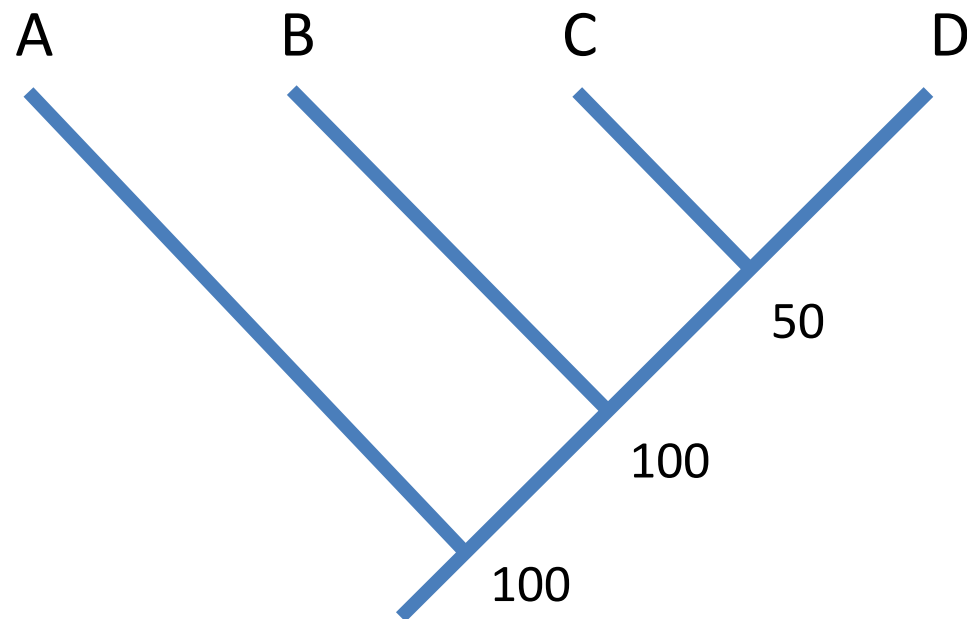


Felsenstein (1985)

Cálculo Bootstrap



Felsenstein (1985)



Programas para calcular árvores de parcimônia:

- TNT (Goloboff 2008)
- PAUP* (Swofford 2003)

Métodos para reconstruções filogenéticas

1. Árvores genéticas: locus de comprimento necessário
 - Parcimônia
 - Maximum likelihood
 - Bayesiano
2. Árvores de espécies: baseadas em árvores genéticas (locus de comprimento) ou polimorfismos de nucleotídeo único (SNPs)

Modelo probabilístico

$$\text{Likelihood} = P(D | M)$$

D = Data → Sequências de DNA

M = Modelo → topologia, comprimento do ramo,
modelo de substituição, frequências das base

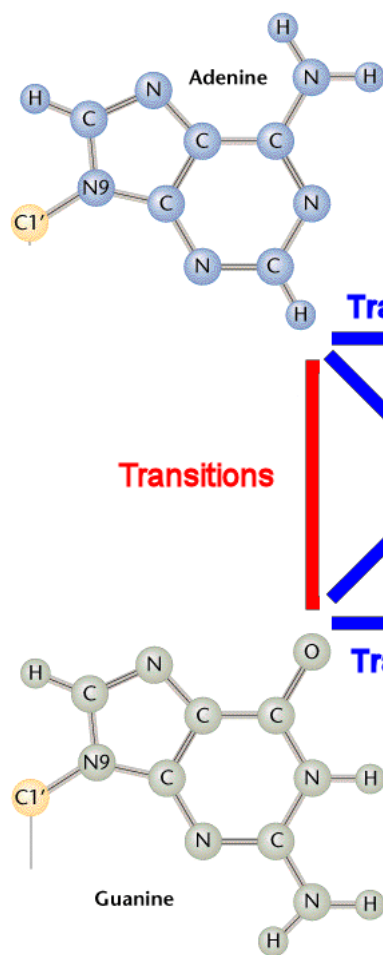
Diferentes modelos de substituições moleculares

Jukes and Cantor (1969)

JC

	T	C	A	G
T	-	$\mu/4$	$\mu/4$	$\mu/4$
C	$\mu/4$	-	$\mu/4$	$\mu/4$
A	$\mu/4$	$\mu/4$	-	$\mu/4$
G	$\mu/4$	$\mu/4$	$\mu/4$	-

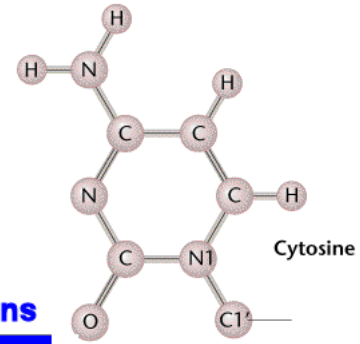
Purinas



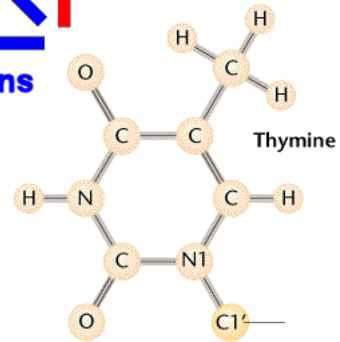
Transversions

Transitions

Transversions



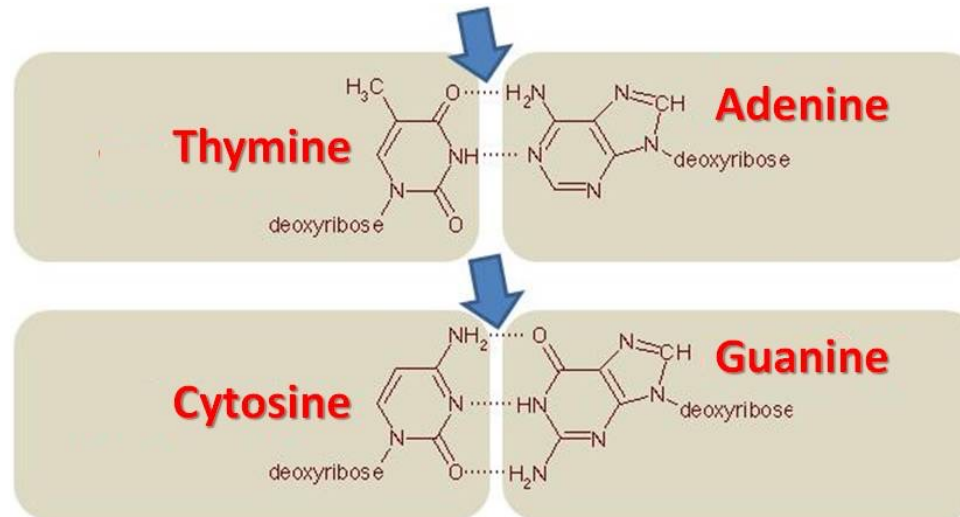
Transitions



Pirimidinas

As frequências básicas não são idênticas ao longo dos cromossomos

Ligações de hidrogênio



Frequências

GC > TA

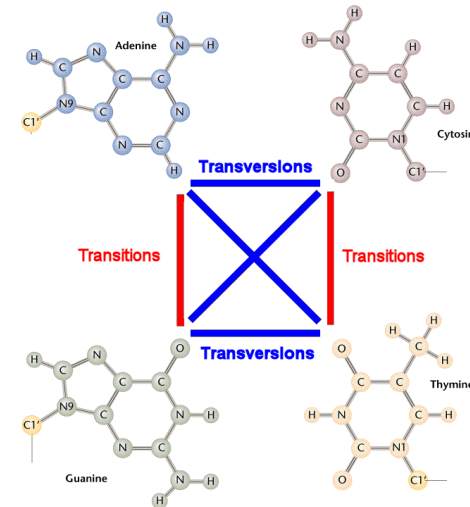
Different models of molecular substitutions

Hasegawa, Kishino and Yano (1985)

HKY

	T	C	A	G
T	-	$k\mu_c$	μ_a	$k\mu_g$
C	$k\mu_t$	-	μ_a	μ_g
A	μ_t	μ_c	-	μ_g
G	μ_t	μ_c	$k\mu_a$	-

k = parâmetro de transição



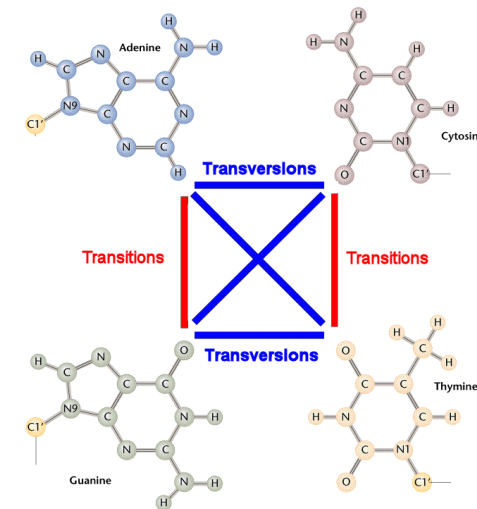
HKY modela diferenças em frequências e transversões/transições

Different models of molecular substitutions

Generalized time reversible (1986)

GTR

$$\begin{pmatrix} \cdot & \alpha\pi_C & \beta\pi_G & \gamma\pi_T \\ \alpha\pi_A & \cdot & \delta\pi_G & \epsilon\pi_T \\ \beta\pi_A & \delta\pi_C & \cdot & \phi\pi_T \\ \gamma\pi_A & \epsilon\pi_C & \phi\pi_G & \cdot \end{pmatrix}$$



GTR é o modelo mais complexo disponível
Cada mutação tem uma probabilidade particular

Modelo probabilístico

$$\text{Likelihood} = P(D | M)$$

Very complex
model !!!!

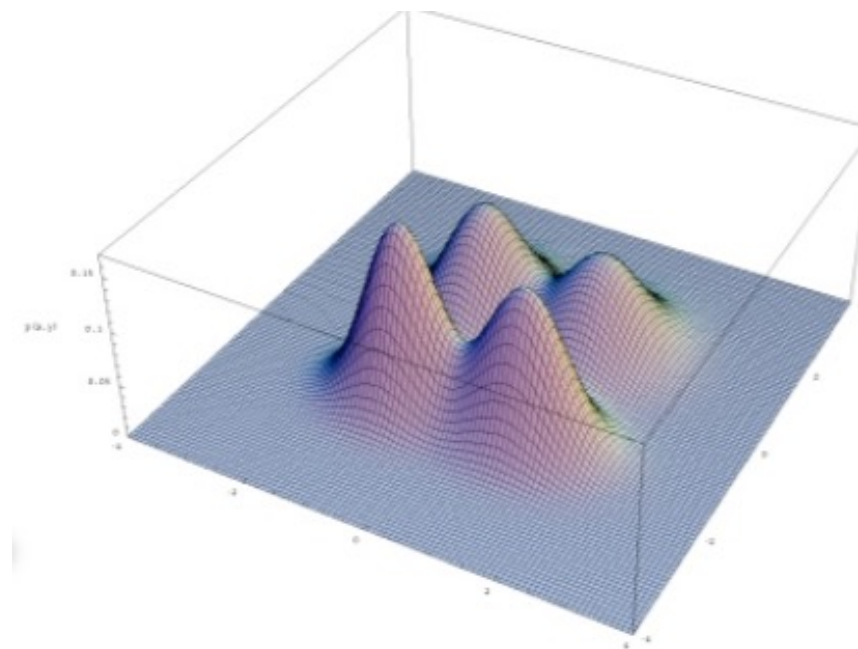


D = Data → Sequências de DNA

M = Modelo → topologia, comprimento do ramo,
modelo de substituição, frequências das base

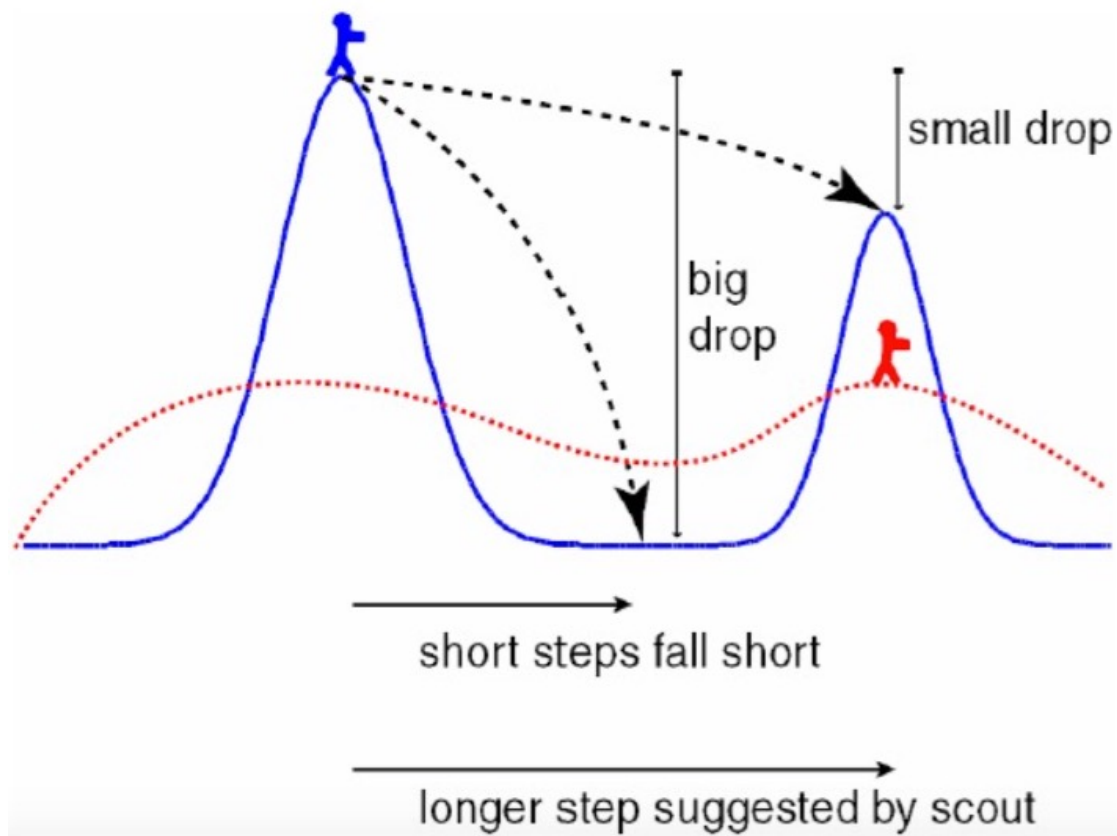
# Taxa (N)	# Unrooted trees
3	1
4	3
5	15
6	105
7	945
8	10,935
9	135,135
10	2,027,025
.	.
.	.
.	.
.	.
30	$\approx 3.58 \times 10^{36}$

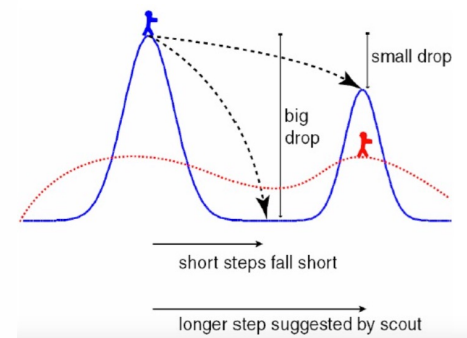
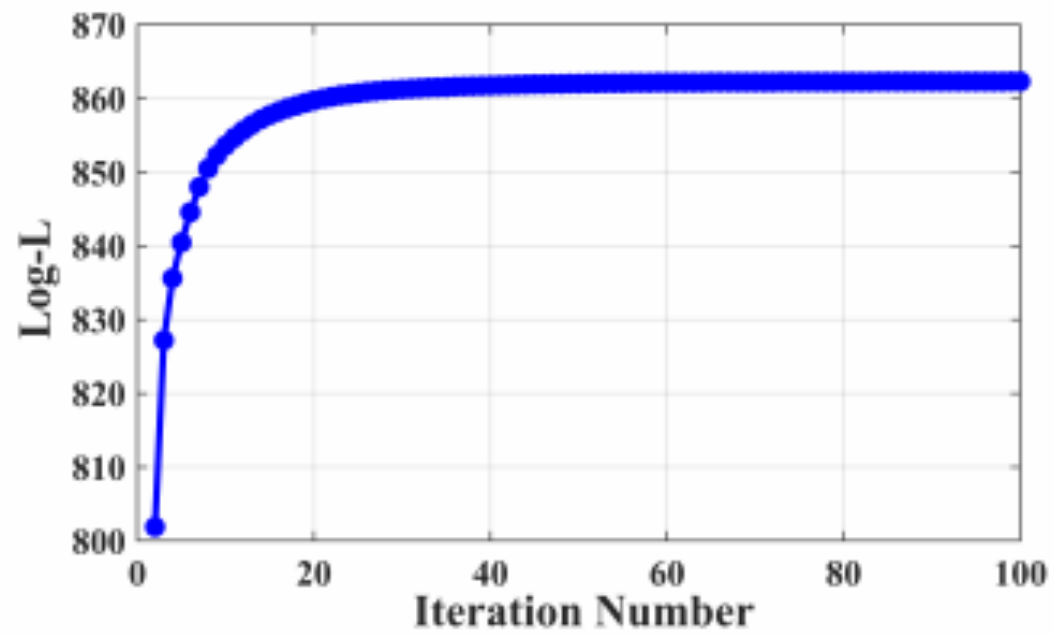
$(2N - 5)!! = \# \text{ unrooted trees for } N \text{ taxa}$



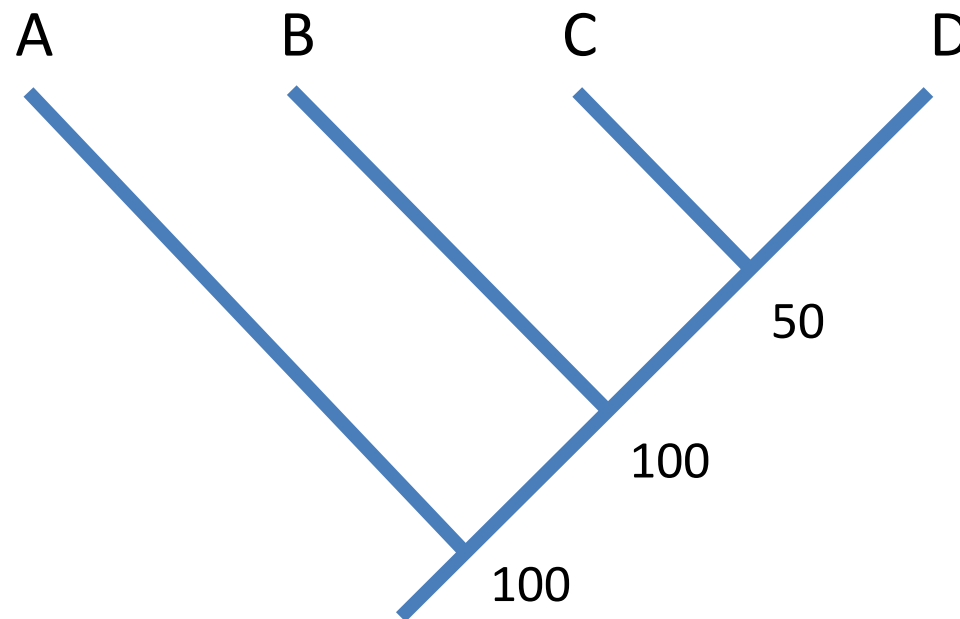
Pesquisa heurística

Markov Chain Monte Carlo (MCMC)



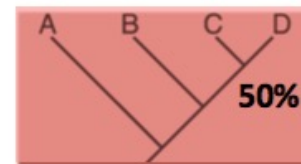


Same bootstrap calculations!!!



Original data

	1	2	3	4	5	6	7	8	9	10
A	C	G	A	A	C	C	A	C	T	T
B	C	G	A	A	C	C	G	G	T	T
C	G	G	T	A	C	C	G	G	A	T
D	G	C	T	A	G	C	G	C	A	T



Programas para calcular árvores de ML

- PAUP* (Swofford 2003)
- Garli (Zwickl 2006)
- PhyML (Guindon 2010)
- RAxML (Stamatakis 2014)

Métodos para reconstruções filogenéticas

1. Árvores genéticas: locus de comprimento necessário
 - Parcimônia
 - Maximum likelihood
 - Bayesiano
2. Árvores de espécies: baseadas em árvores genéticas (locus de comprimento) ou polimorfismos de nucleotídeo único (SNPs)

$$P(M | D) = \frac{P(D | M) \times P(M)}{P(D)}$$

likelihood

Probabilidade *prior*

D = Data → Sequências de DNA

M = Modelo → topologia, comprimento do ramo,
modelo de substituição, frequências das base

Hipótese:



$$P(M | D) = \frac{P(D | M) \times P(M)}{P(D)}$$

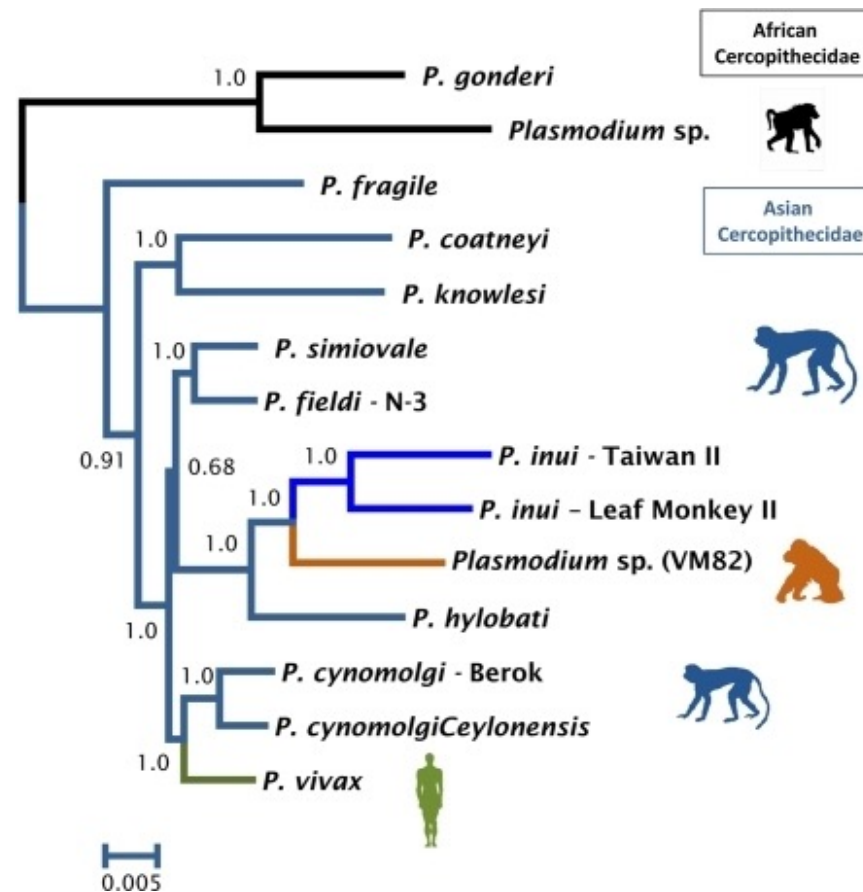
likelihood

Probabilidade *prior*

D = Data → Sequências de DNA

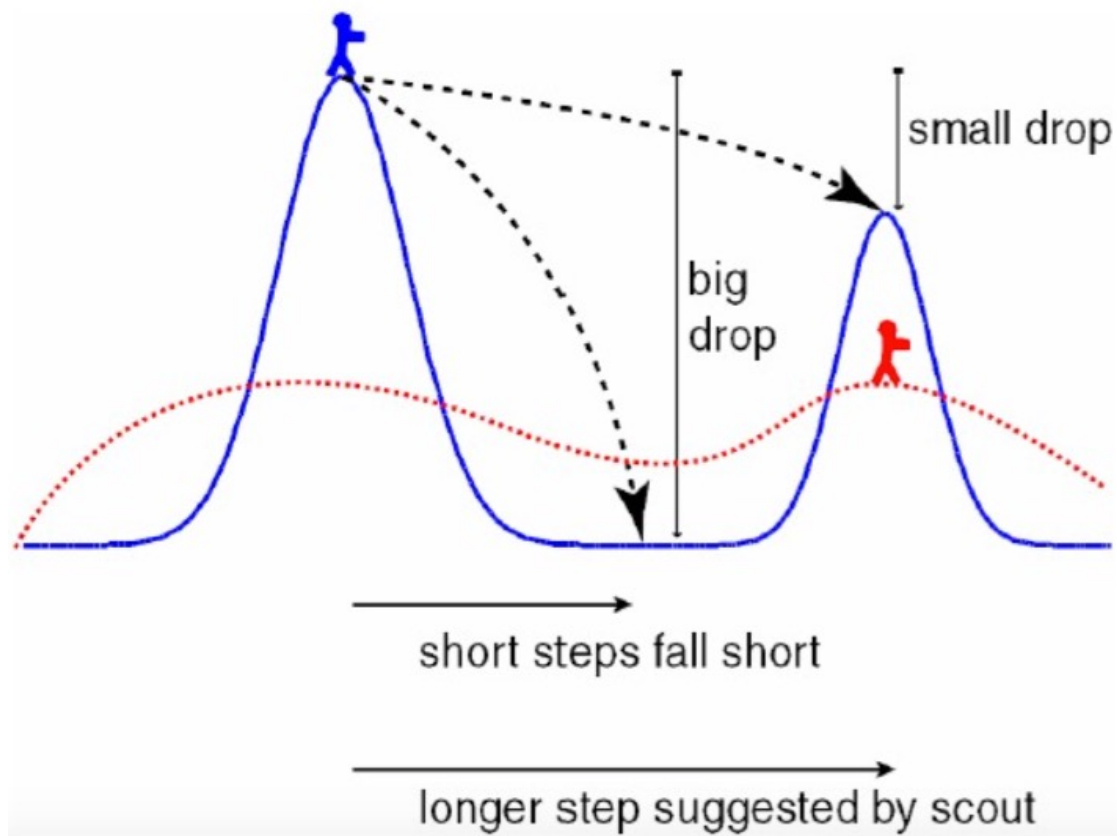
M = Modelo → topologia, comprimento do ramo,
modelo de substituição, frequências das base

A probabilidade posterior é calculada durante a estimativa da topologia em árvore, portanto, análises extras com bootstrap não são necessárias



Pacheco et al. (2012)

Markov Chain Monte Carlo (MCMC)



Programas para computação para árvores bayesianas

- MrBayes (Huelsenbeck and Ronquist 2001)
- BEAST (Drummond et al. 2012)

Métodos para reconstruções filogenéticas

1. Árvores genéticas: locus de comprimento necessário
 - Parcimônia
 - Maximum likelihood
 - Bayesiano
2. Árvores de espécies: baseadas em árvores genéticas (locus de comprimento) ou polimorfismos de nucleotídeo único (SNPs)

Sequências de DNA

Locus 1

```
ATCAAAATTTGGCGGC
ATCAAAATTTGGCGGC
ATCAAAATTTGGCGGC
ATCAAAATTTGGCGGC
ATCAAAATTTGGCGGC
```

Locus 2

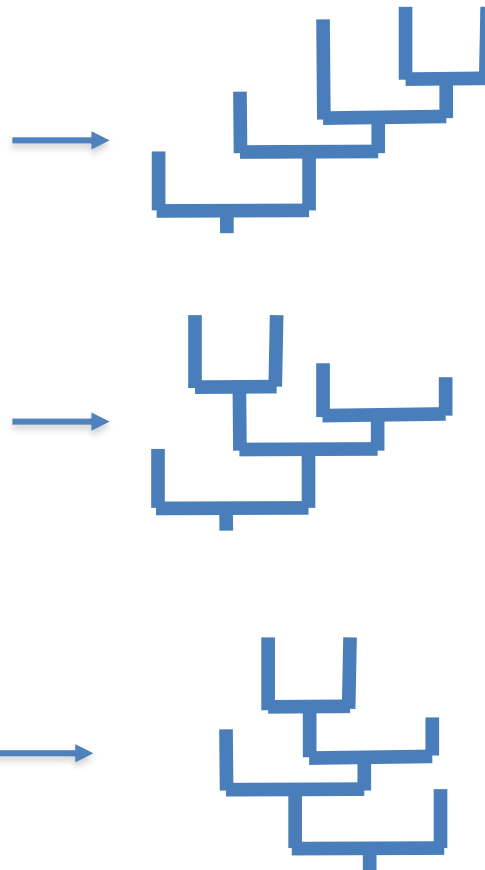
```
TTGCAAAGGGGCGC
TTGCAAAGGGGCGC
TTGCAAAGGGGCGC
TTGCAAAGGGGCGC
TTGCAAAGGGGCGC
```

⋮

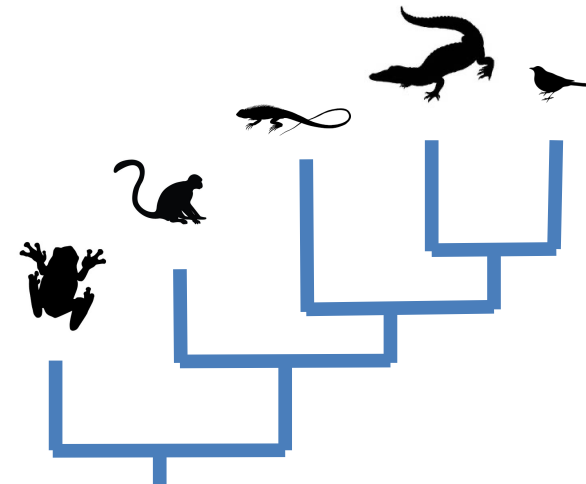
Locus n

```
CCTGCTCCCGGTCGTA
CCTGCTCCCGGTCGTA
CCTGCTCCCGGTCGTA
CCTGCTCCCGGTCGTA
CCTGCTCCCGGTCGTA
```

Árvores genéticas independentes



Árvores de espécies



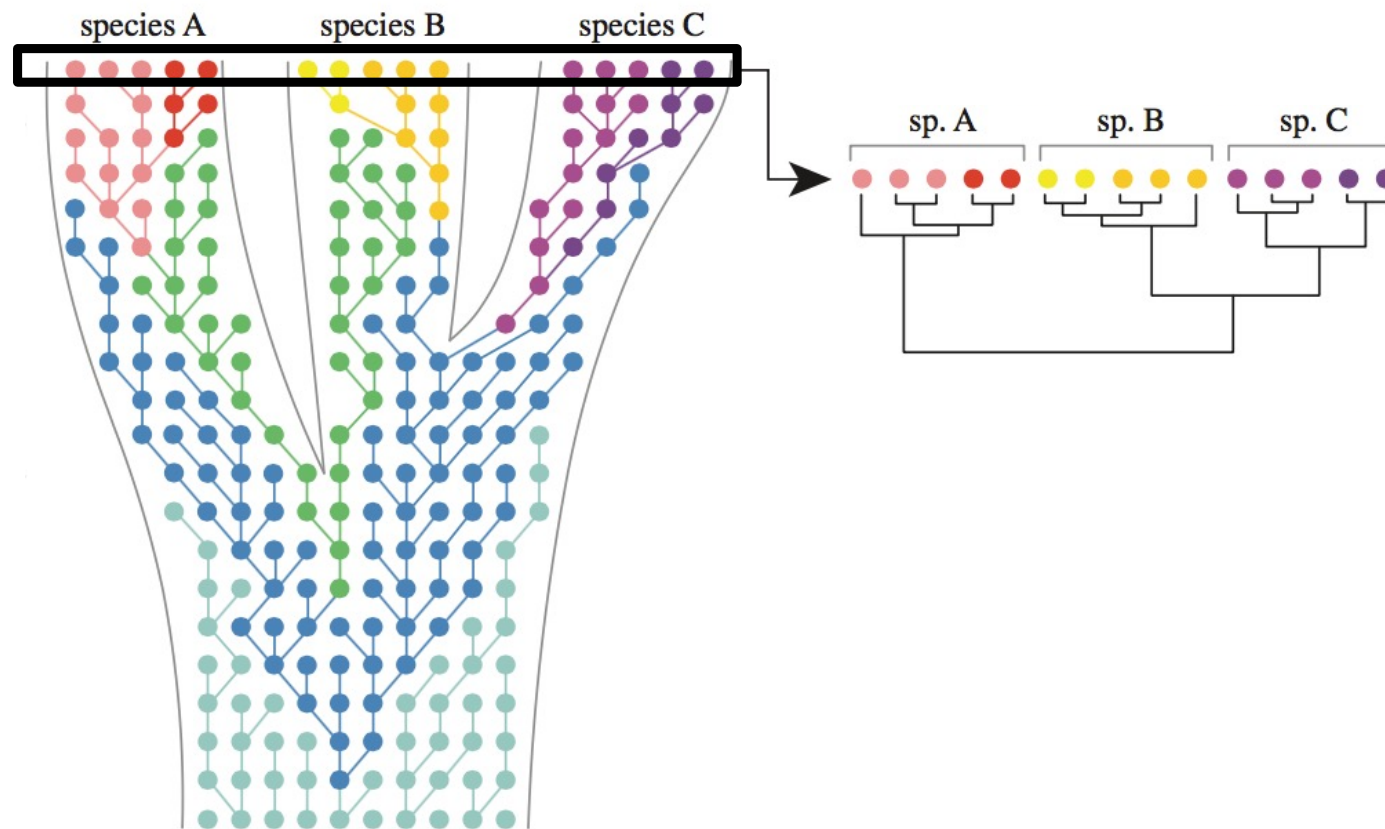


Fig. de Leliaert et al. 2014

species A

species B

species C

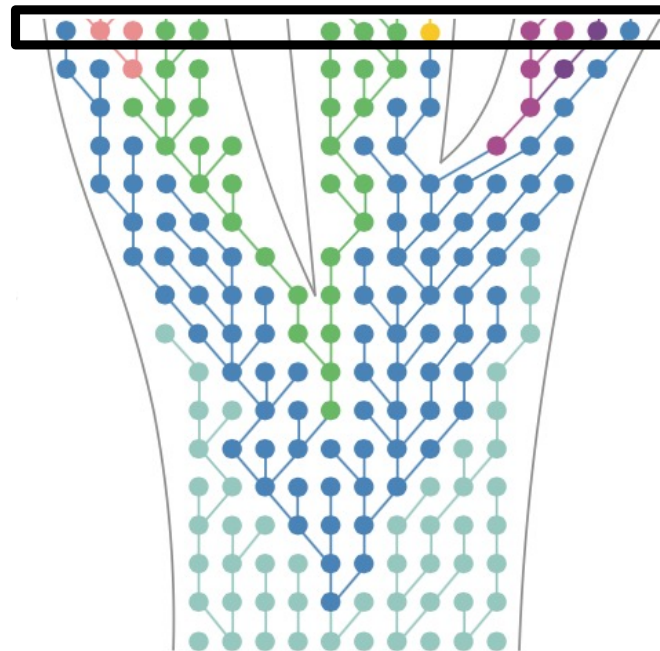


Fig. de Leliaert et al. 2014

species A

species B

species C

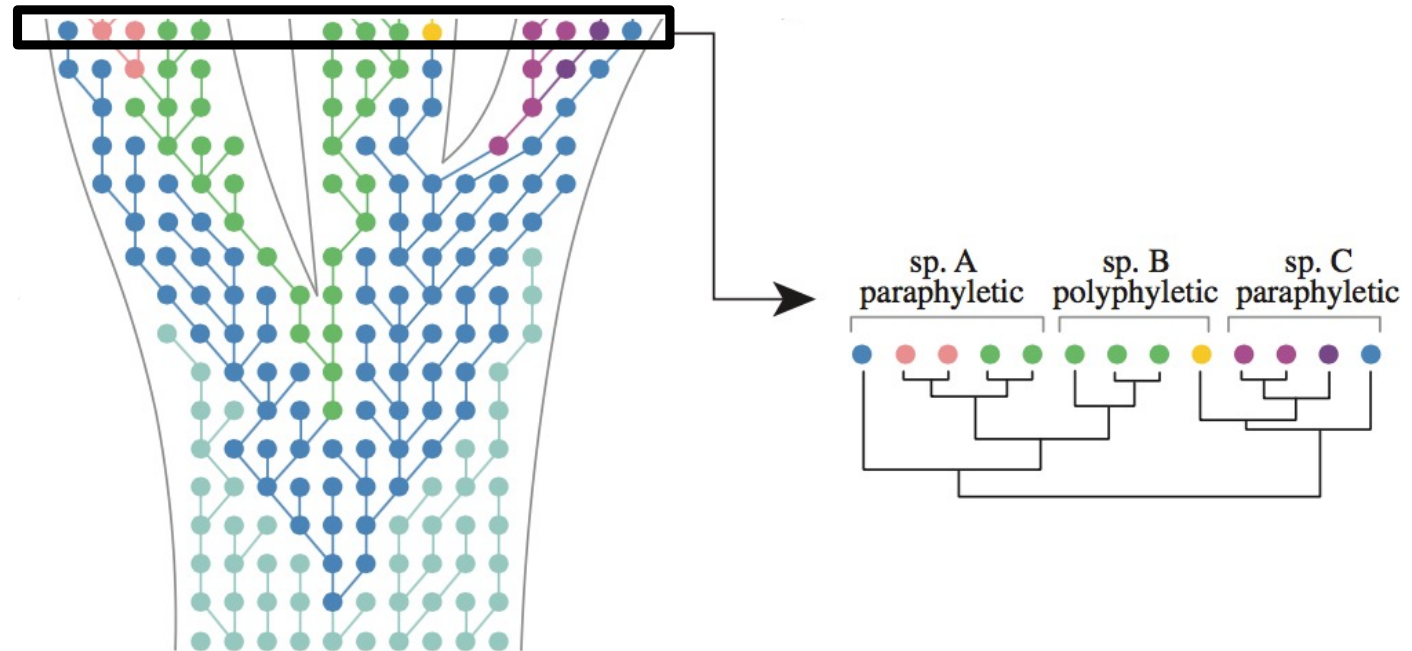
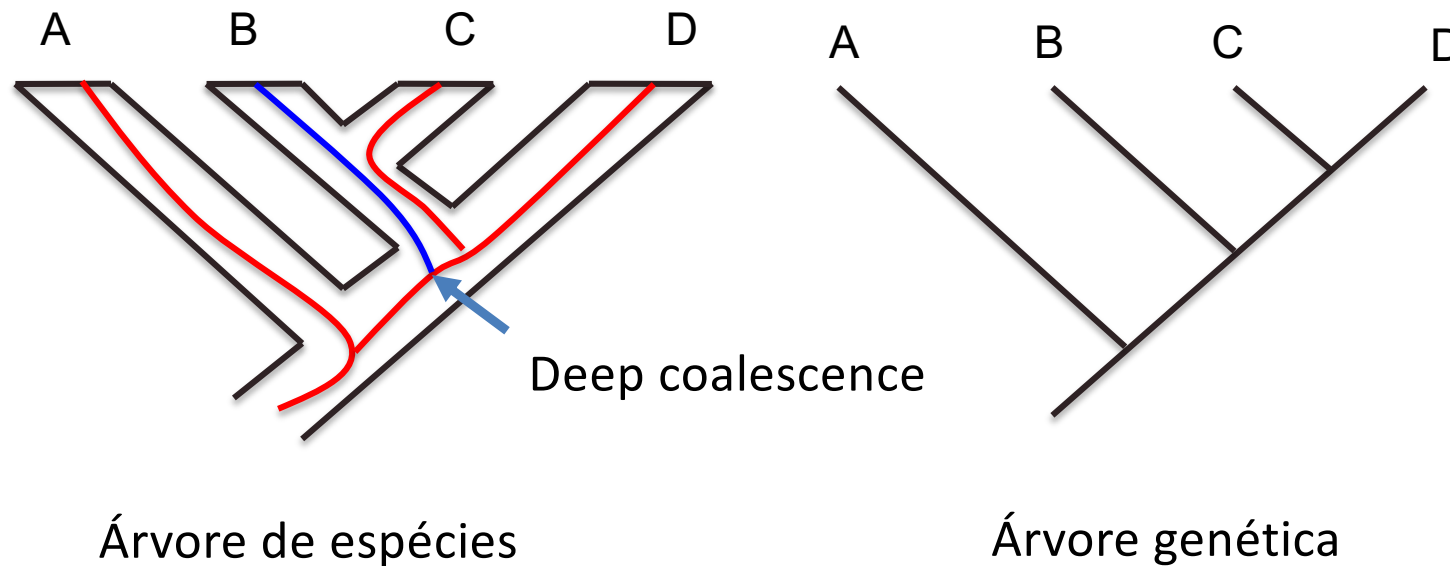


Fig. de Leliaert et al. 2014

Incomplete Lineage Sorting (ILS)



As árvores genéticas que discordam da tendência central não estão erradas; pelo contrário, fazem parte do padrão difuso que constitui a história genética.

Maddison 1997

Sanger Sequences

- BEST (Liu 2008)
- *BEAST (Drummond & Rambaut 2007)
- BUCKy (Larget et al. 2010)
- MDC (Than & Nakhleh 2009)
- STEM (Kubatko et al. 2009)

Bayes

Parsimony

Likelihood

Escala genômica

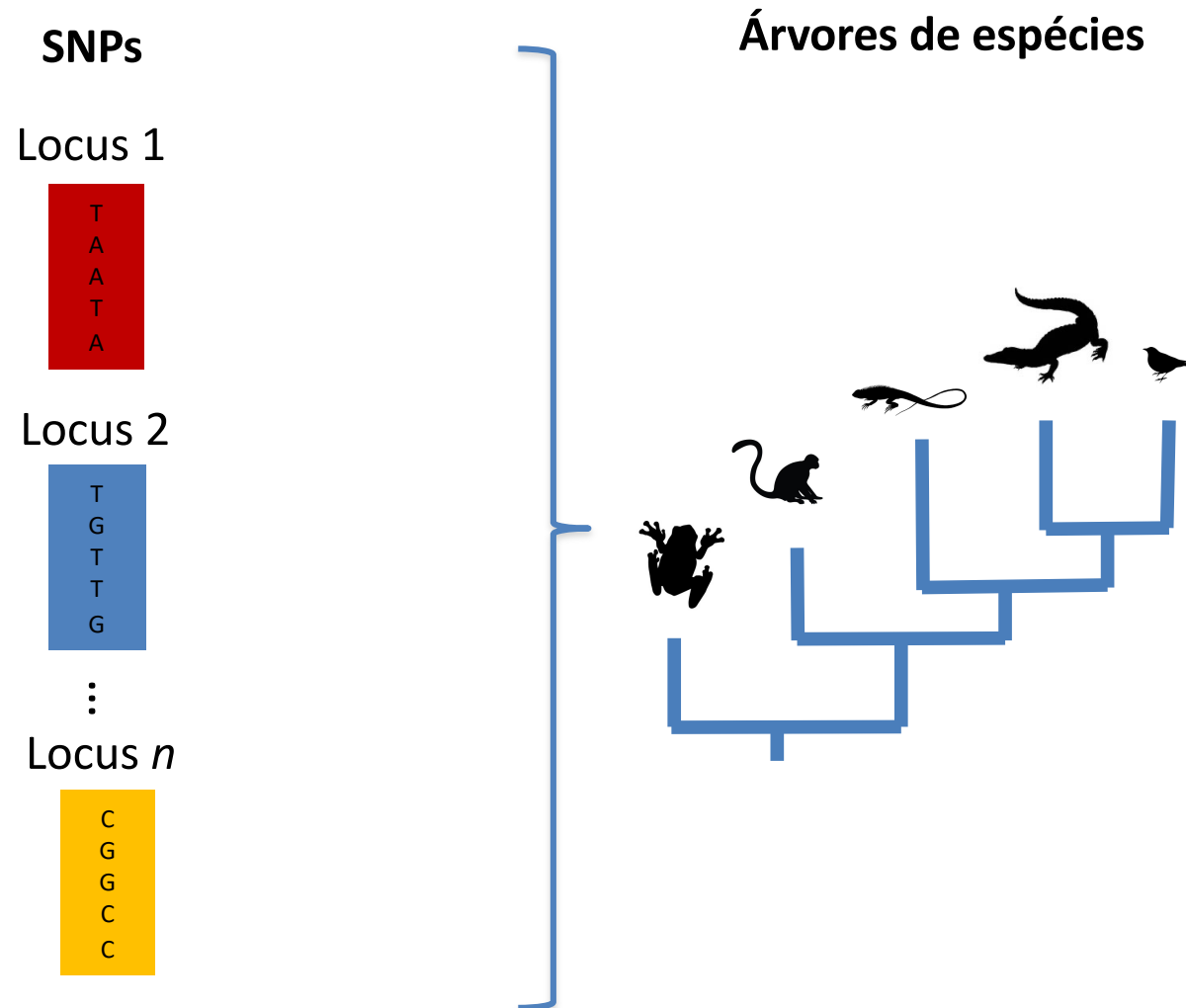
- starBEAST2 (Ogilvie et al. 2017)
- SNAPP (Bryant et al. 2012) [SNPs]
- MP-EST (Liu et al. 2010)
- ASTRAL (Mirab et al. 2015)
- SVDquartets (Chifman and Kubatko 2015) [SNPs]

Bayes

Likelihood

Programs for estimating species trees

Program	Type of data	loci	Species / individuals	time consumption
starBEAST2	long sequences	hundreds	few	slow
SNAPP	SNPs	hundreds	few	slow
MP-EST	gene trees	thousands	many	medium
ASTRAL	gene trees	thousands	many	fast
SVDquartets	SNPs	thousands	many	fast



Boa revisão de SNPs em filogenética por Leaché e Oak 2017 TREE

Muito obrigado!!!



**SOMETIMES,
EVOLUTION SUCKS.**



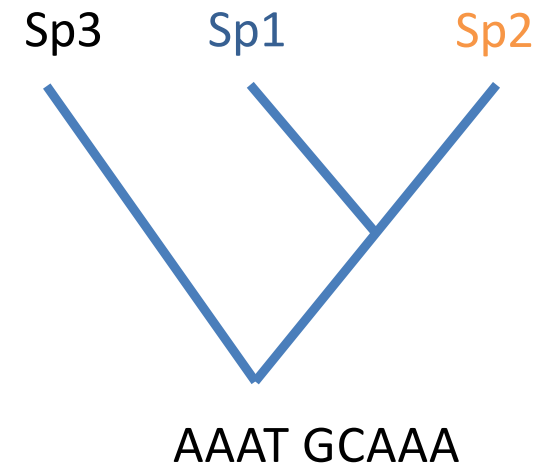
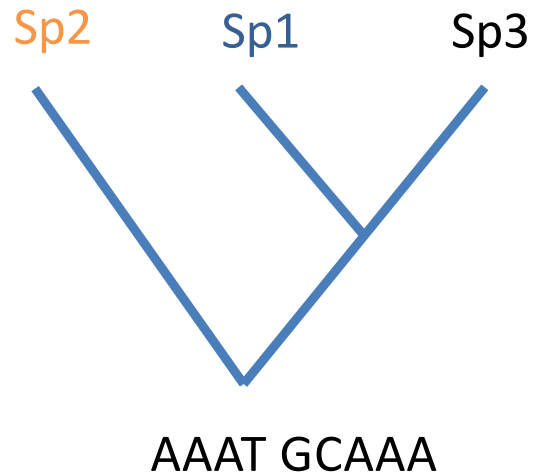
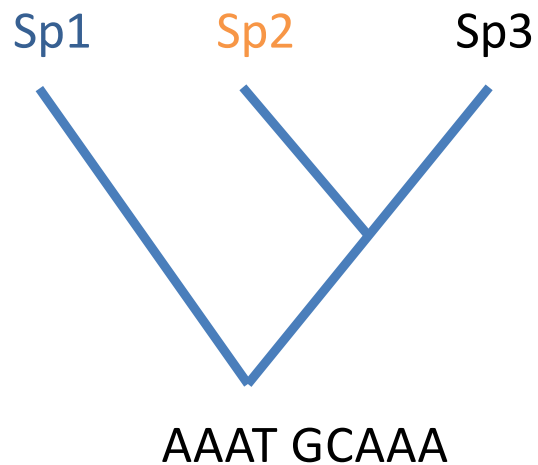
Melisa_Olave



molave@mendoza-conicet.gob.ar

Maximum Parsimony

Sp1 AAATGCAAA
Sp2 AAA**G**GCAAA
Sp3 AAA**G**GCAAA
ancestral AAAT GCAAA



Maximum Parsimony

Sp1 AAATGCAAA
Sp2 AAA**G**GCAAA
Sp3 AAA**G**GCAAA
ancestral AAAT GCAAA

