

TEORIA DA COALESCÊNCIA

Estudo das propriedades genealógicas
de linhagens

APLICABILIDADE

- É o instrumento interpretativo mais usado na análise dos processos populacionais que estão na base da variação de sequências de DNA.

CARACTERÍSTICAS

- Abordagem genealógica
- Modelação dos processos mutacionais no sentido presente→passado e não presente→futuro
- Limitação da análise à amostra disponível sem que seja necessário ter em conta toda a população
- Dissociação conceptual entre processos mutacionais e genealógicos

TIPO DE DADOS

SNP

AAGCTTGTTTCAGAGGAGTGA**C**AATCTGCTTACATCTTGACCTT

AAGCTTGTTTCAGAGGAGTGA**T**AATCTGCTTACATCTTGACCTT



TIPO DE DADOS

a

| | | Chimpanzee haplotype | Human haplotypes | | | | | | | |
|---------------------|-------|-------------------------|------------------|---|---|---|---|---|---|---|
| | | | A | B | C | D | E | F | G | H |
| Nucleotide position | 12 | T | . | . | . | C | C | C | C | C |
| | 397 | G | . | . | . | . | C | . | . | . |
| | 847 | C | . | . | . | . | . | . | T | T |
| | 1,261 | A | . | . | . | . | . | . | . | G |
| | 1,343 | T | A | A | . | . | . | . | . | . |
| | 1,344 | G | . | . | T | . | . | . | . | . |
| | 1,634 | T | . | . | . | . | . | C | C | C |
| | 1,639 | A | . | . | G | G | G | G | G | G |
| | 1,733 | A | . | . | . | G | G | . | . | . |
| | 1,951 | T | G | G | . | . | . | . | . | . |
| | 2,020 | G | A | A | . | . | . | . | . | . |
| | 2,263 | C | A | . | . | . | . | . | . | . |
| | 2,304 | T | . | . | C | C | C | C | C | C |

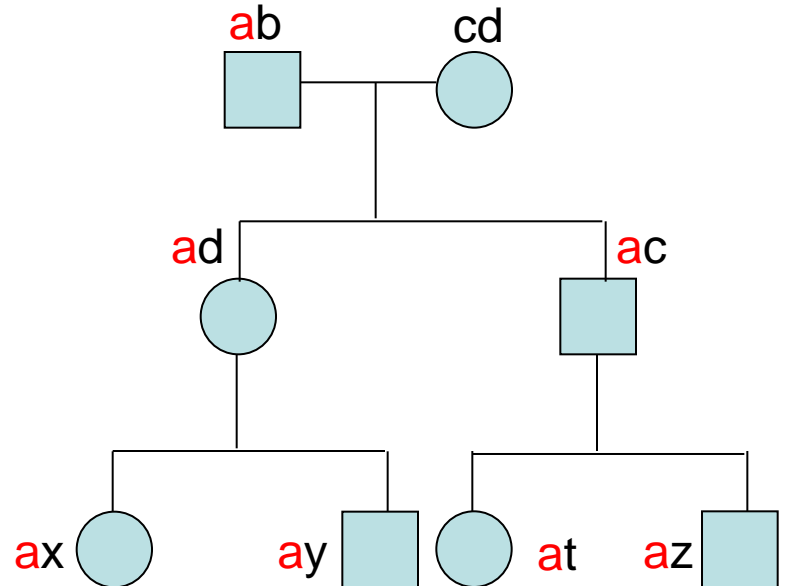
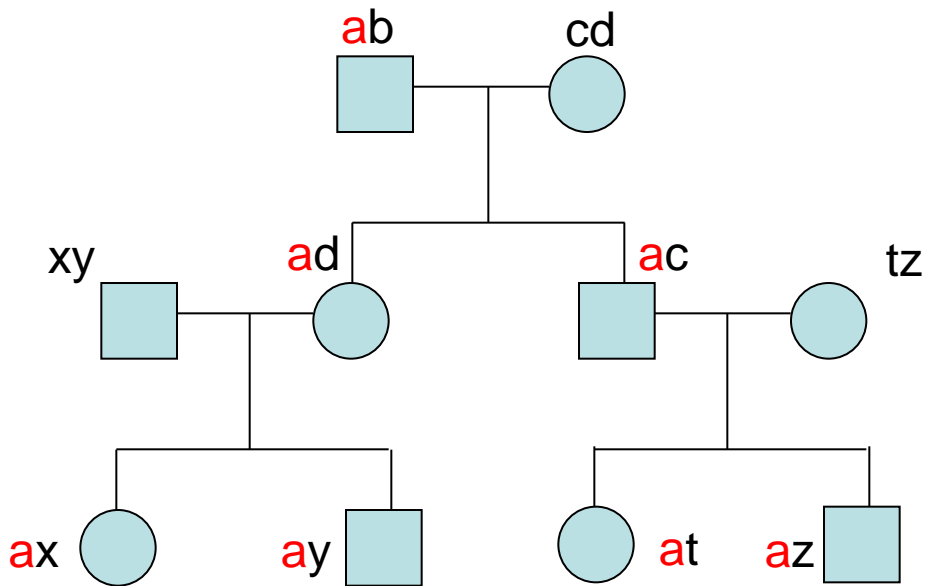
TIPO DE DADOS

site of variation

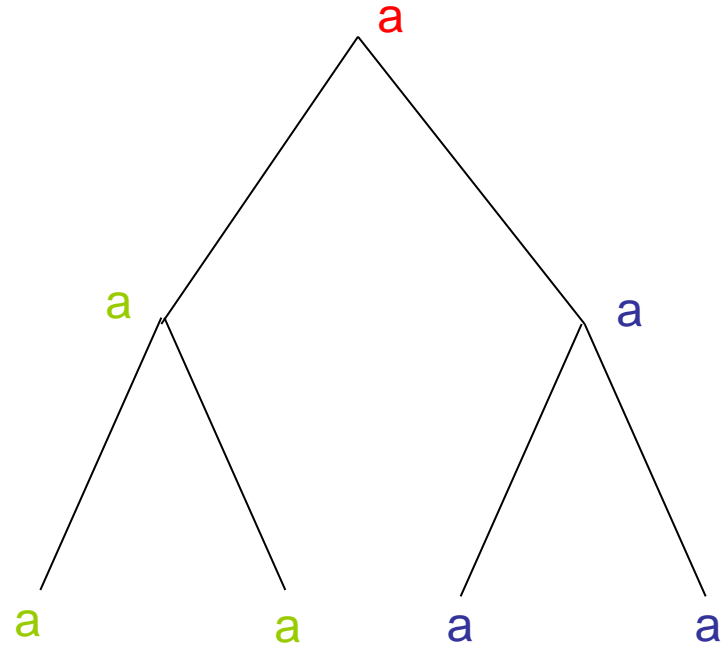
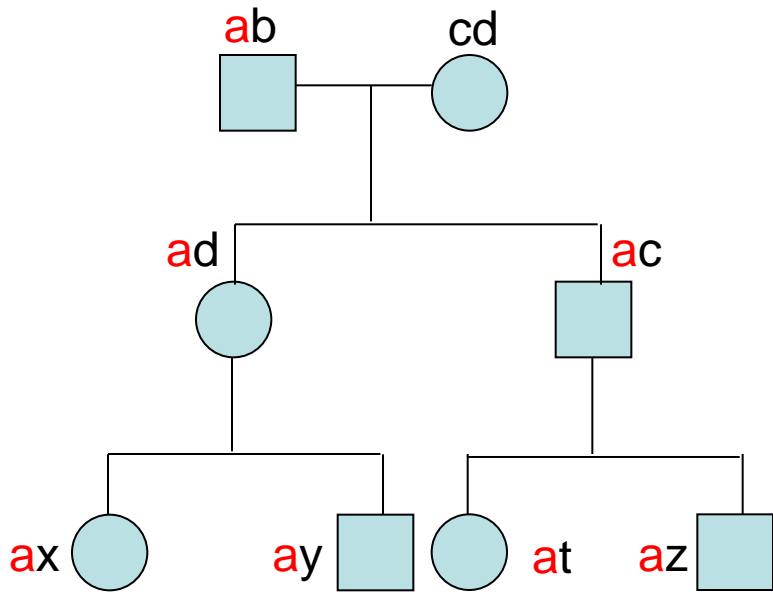
individuals

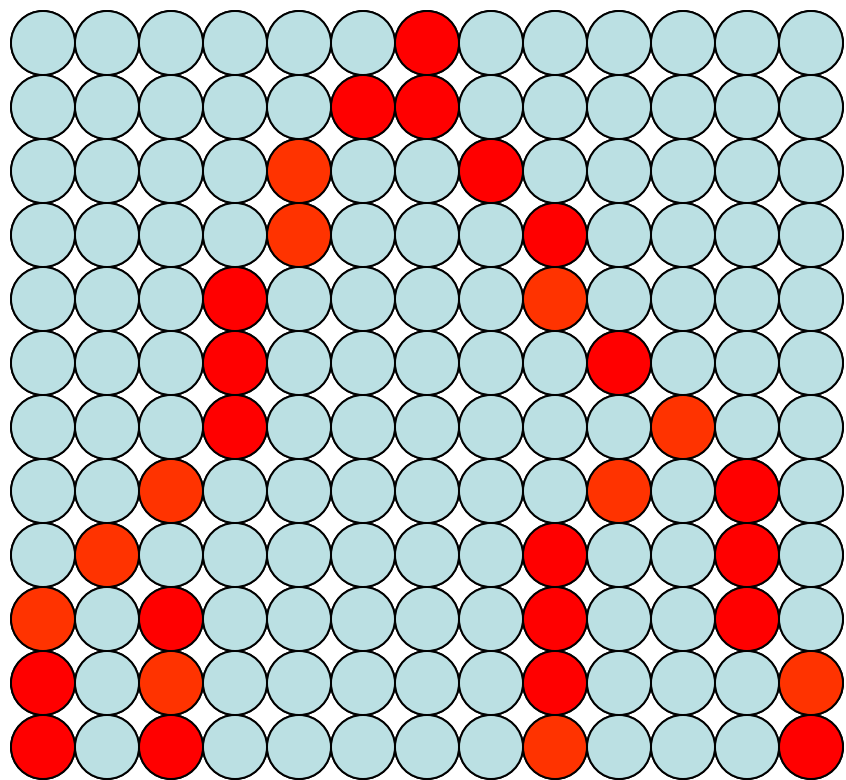


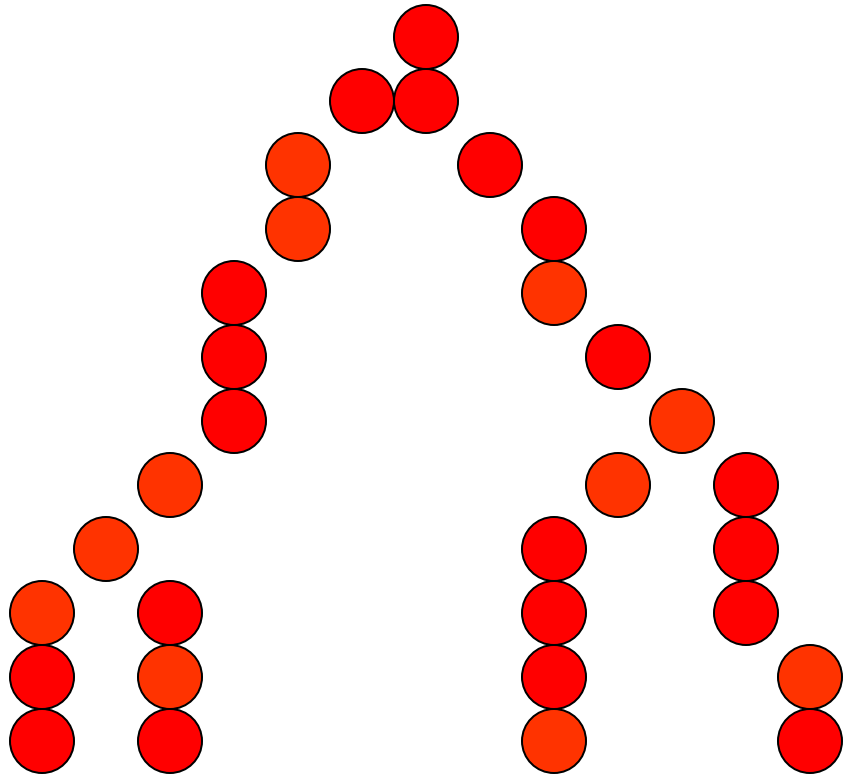
A ABORDAGEM GENEALÓGICA

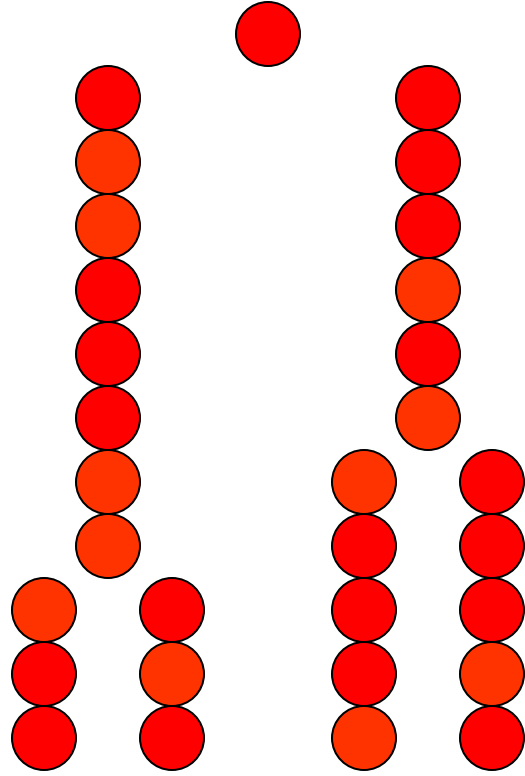


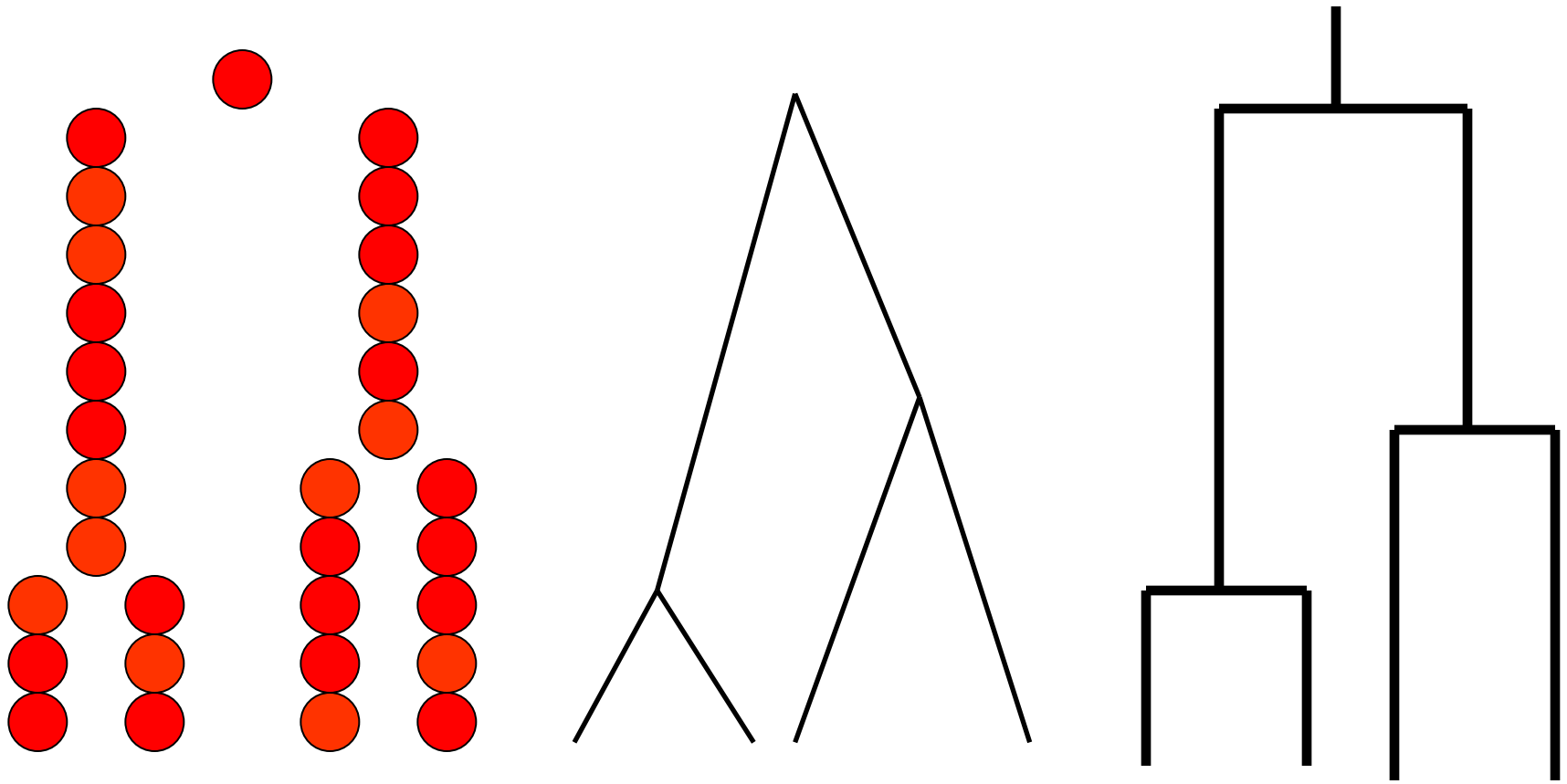
A ABORDAGEM GENEALÓGICA









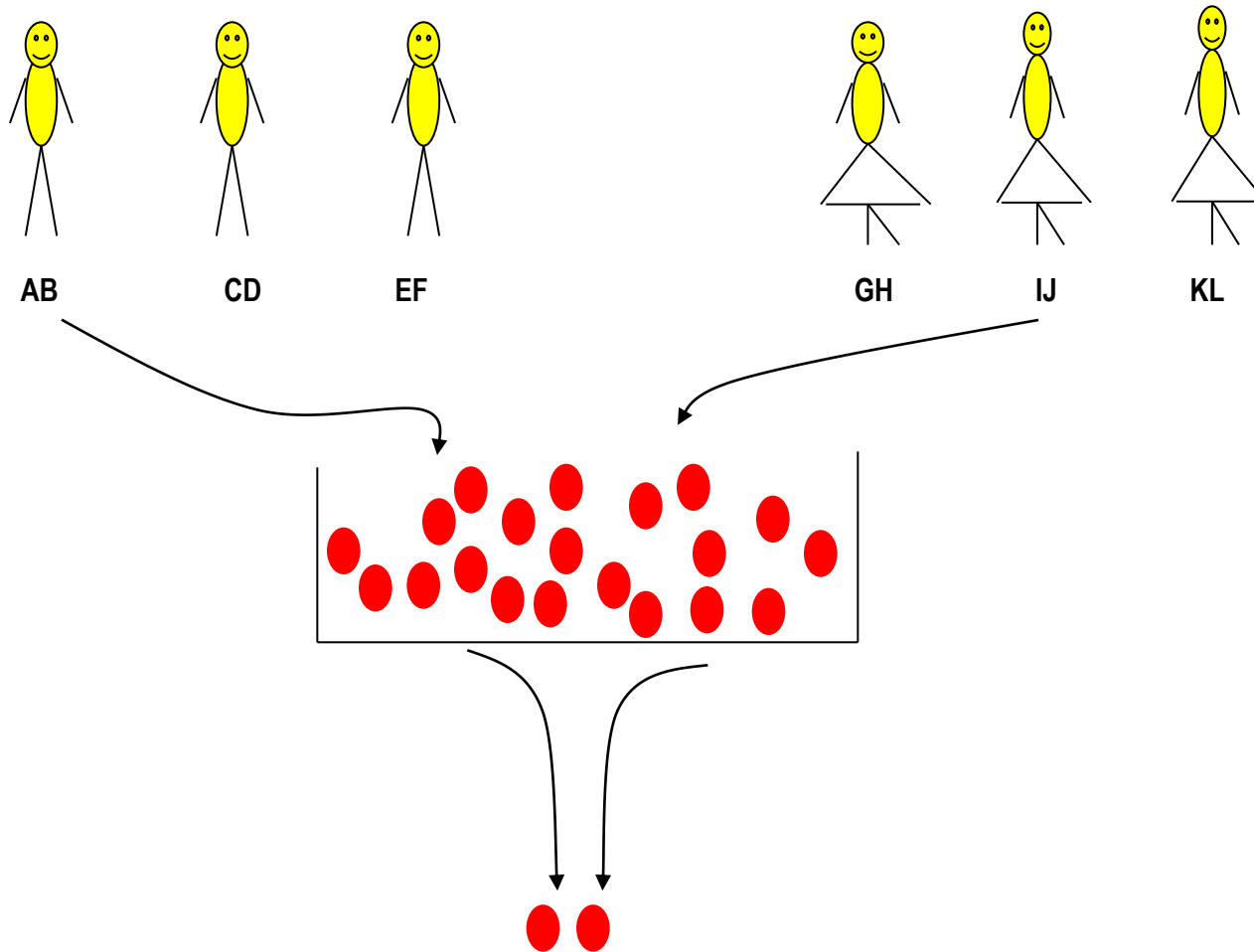


A teoria da coalescência centra-se na genealogia da amostra

MODELO DE WRIGHT-FISHER

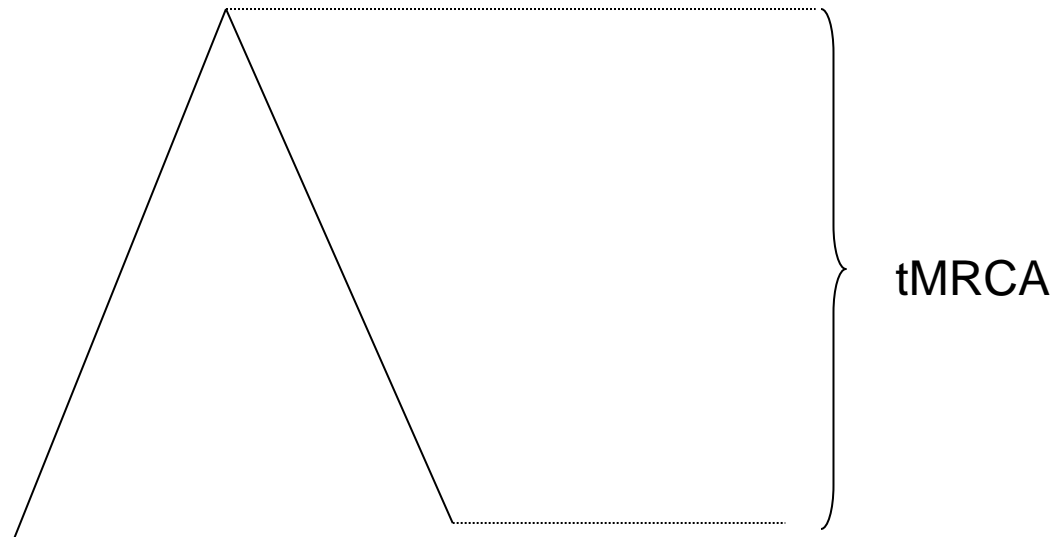
- N constante
- Acasalamento ao acaso
- Ausência subdivisões geográficas (ou outras)
- Ausência de selecção
- Gerações discretas
- ⇒ **O insucesso reprodutivo está distribuído ao acaso**
 - *Para valores de N grandes, o número de descendentes por indivíduo tem uma distribuição Poisson.*

MODELO DE WRIGHT-FISHER

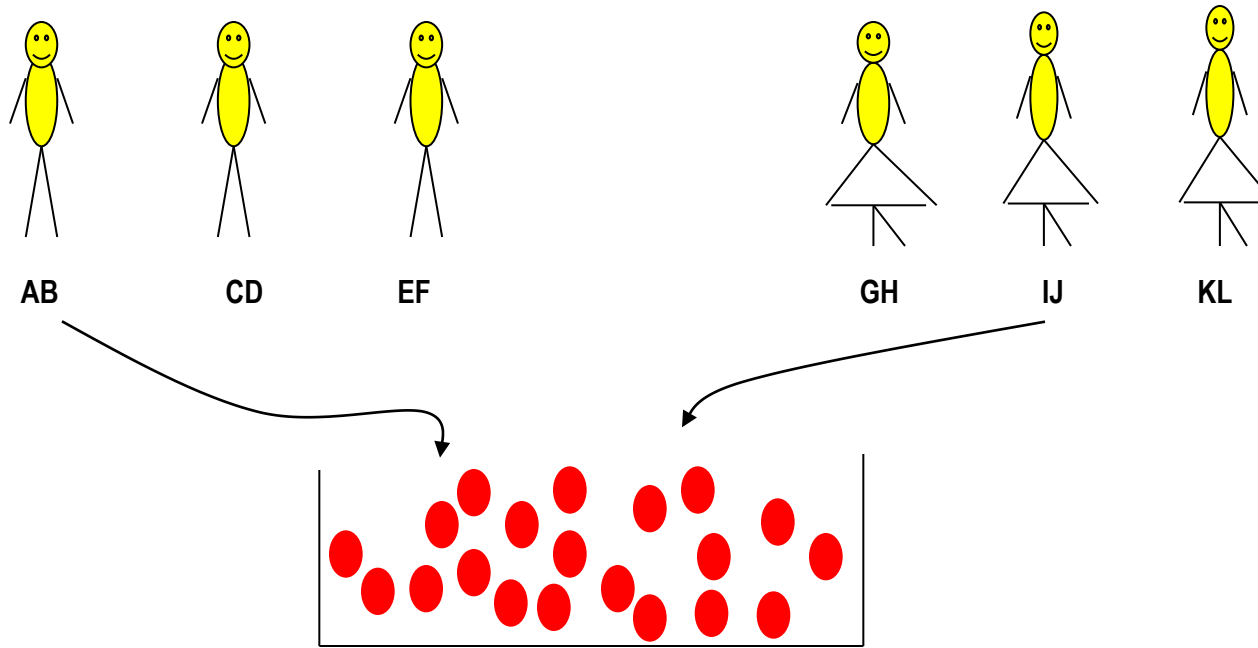


PROPRIEDADES DAS GENEALOGIAS NO MODELO DE WRIGHT-FISHER

TEMPO DE COALESCÊNCIA $n=2$



TEMPO DE COALESCÊNCIA $n=2$



Prob. de partilha de ancestral na geração anterior ●●

$$AA = [1/12]^2$$

$$BB = [1/12]^2$$

$$CC = [1/12]^2$$

$$DD = [1/12]^2$$

$$EE = [1/12]^2$$

$$FF = [1/12]^2$$

$$GG = [1/12]^2$$

$$HH = [1/12]^2$$

$$II = [1/12]^2$$

$$JJ = [1/12]^2$$

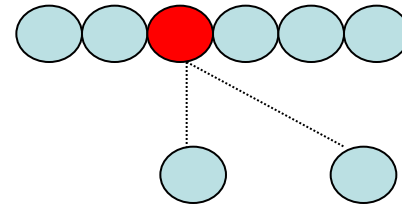
$$KK = [1/12]^2$$

$$LL = [1/12]^2$$

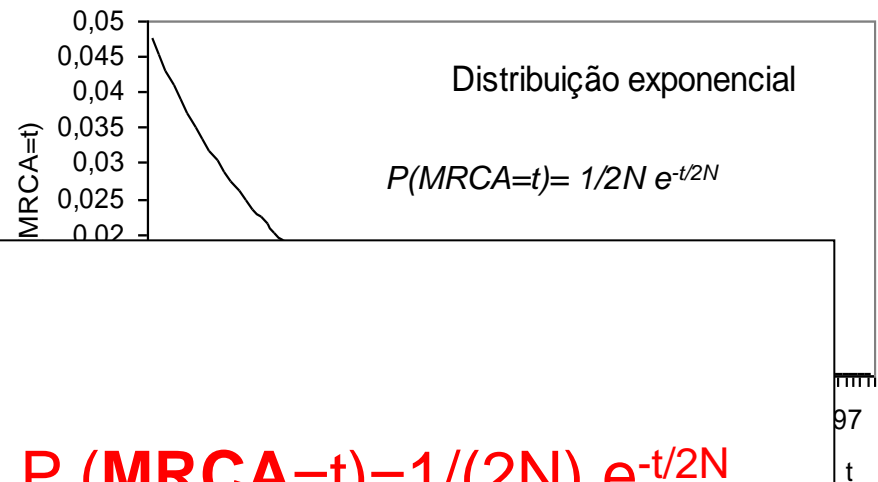
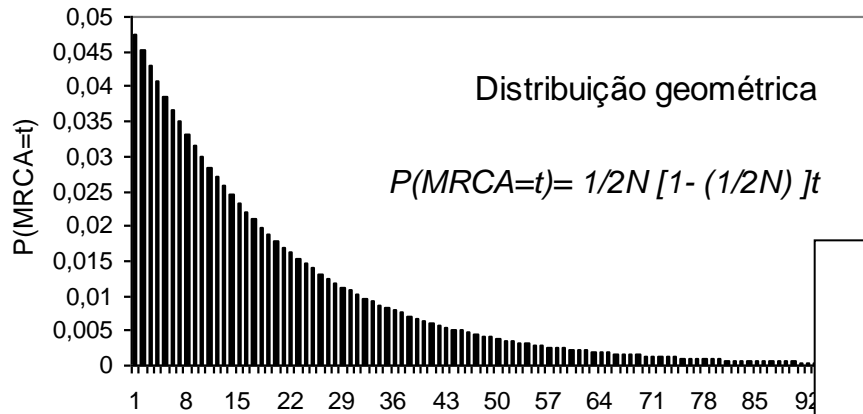
$$= [1/12]^2 \cdot 12 = 1/12 = 1/2N$$

TEMPO DE COALESCÊNCIA $n=2$

- $n=2$
- Prob. de partilha de ancestral na geração anterior:
- $1/2N$
- Prob. de não partilha de ancestral na geração anterior:
- $1 - (1/2N)$
- Probabilidade de coalescência na geração $t+1$
- $1/2N [1 - (1/2N)]^t$

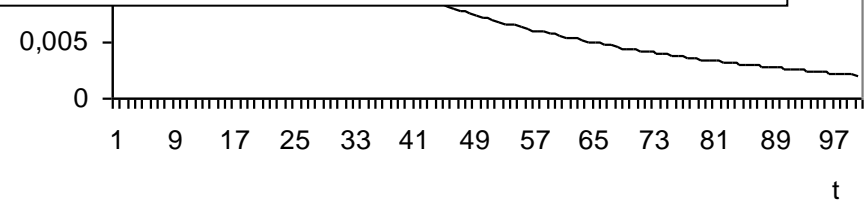
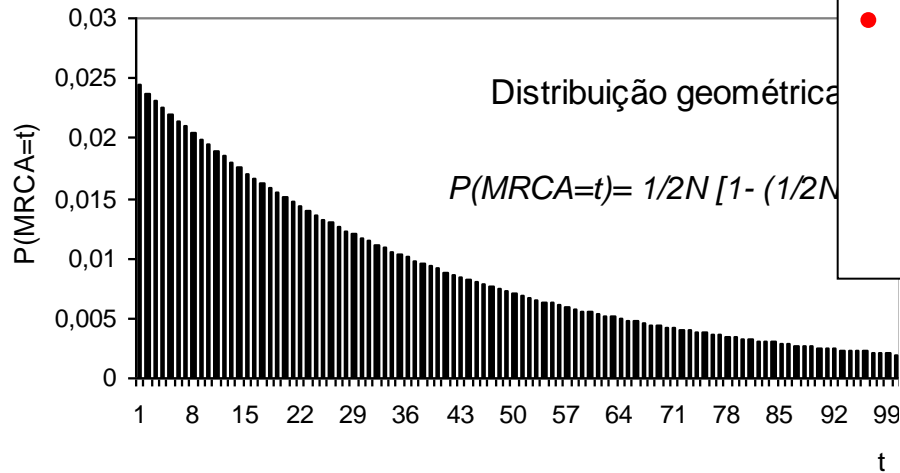


n=2; N=10

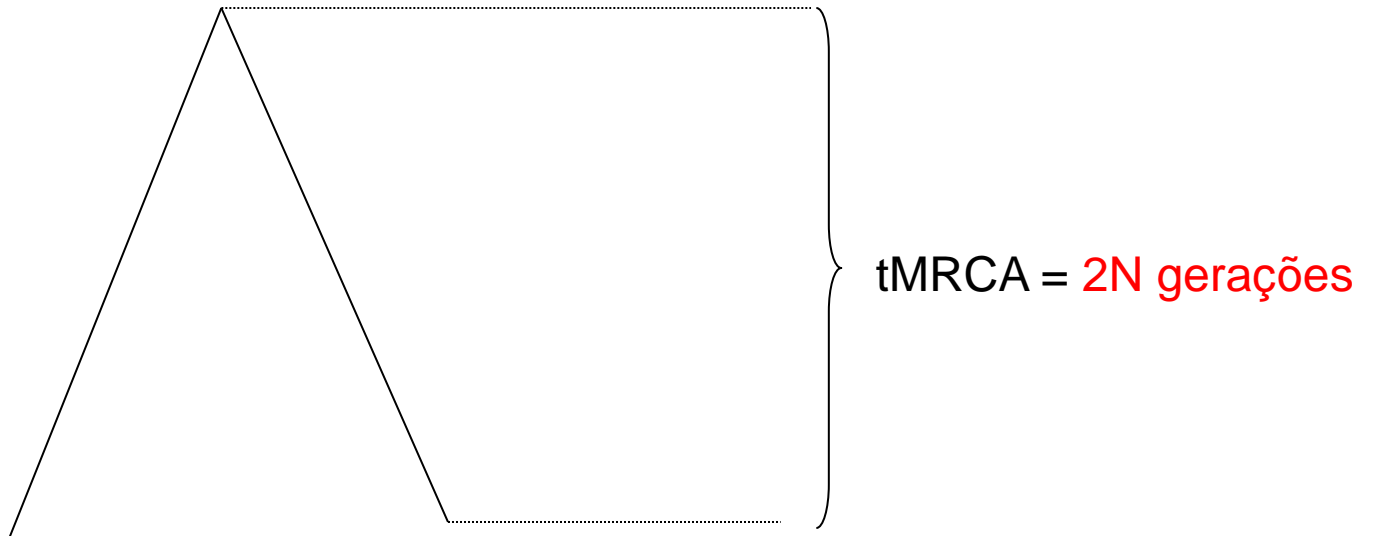


- $P(\underline{MRCA}=t) = \frac{1}{(2N)} e^{-t/2N}$
- $E(t) = 2N$ gerações

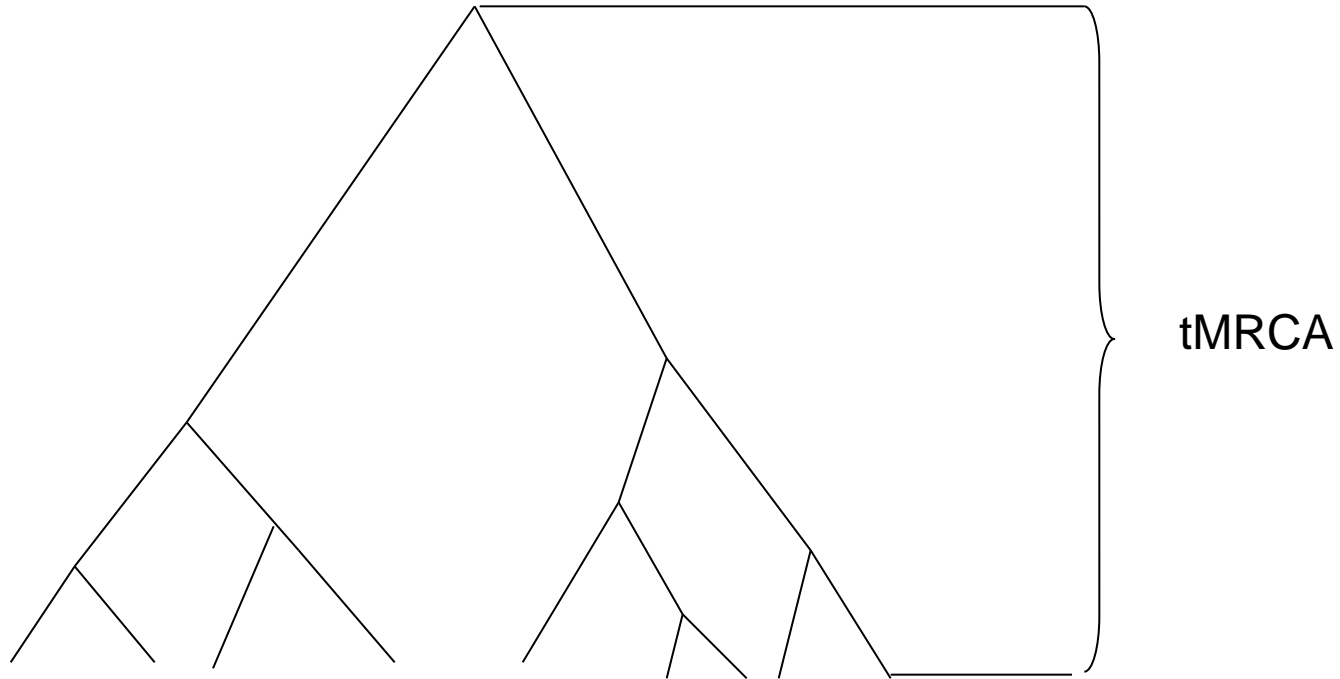
n=2; N=20



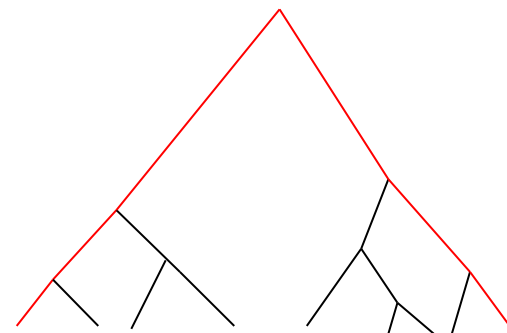
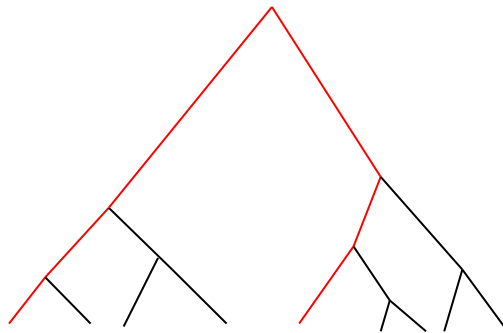
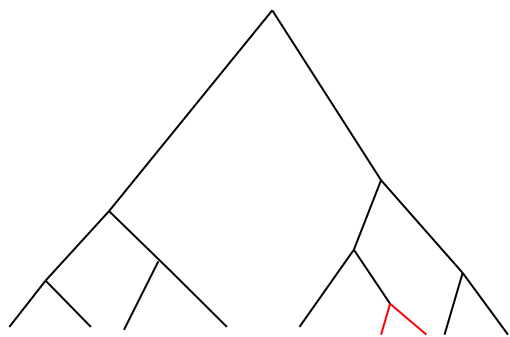
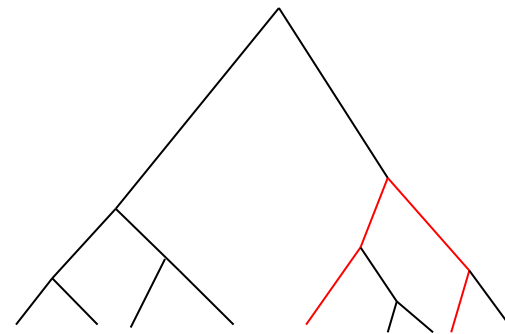
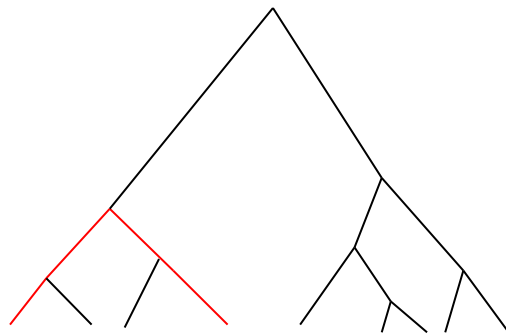
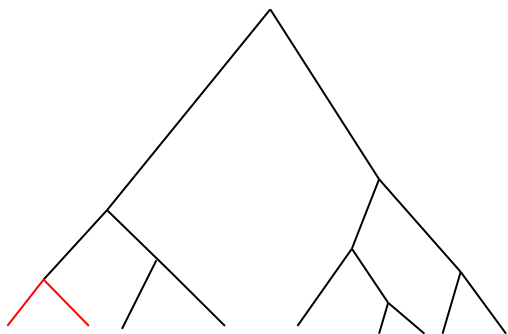
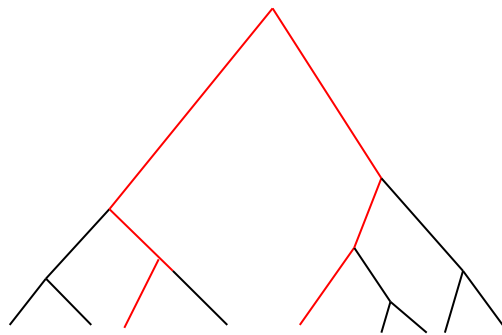
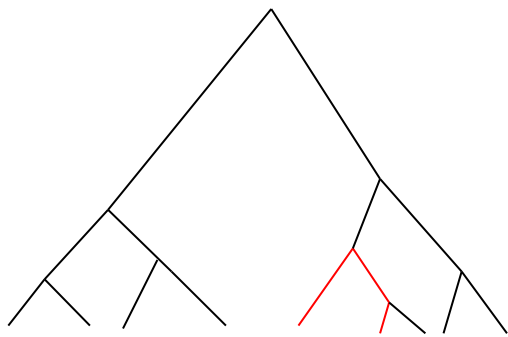
TEMPO DE COALESCÊNCIA $n=2$



TEMPO DE COALESCÊNCIA $n=n$



TEMPO DE COALESCÊNCIA $n=n$



TEMPO DE COALESCÊNCIA $n=n$

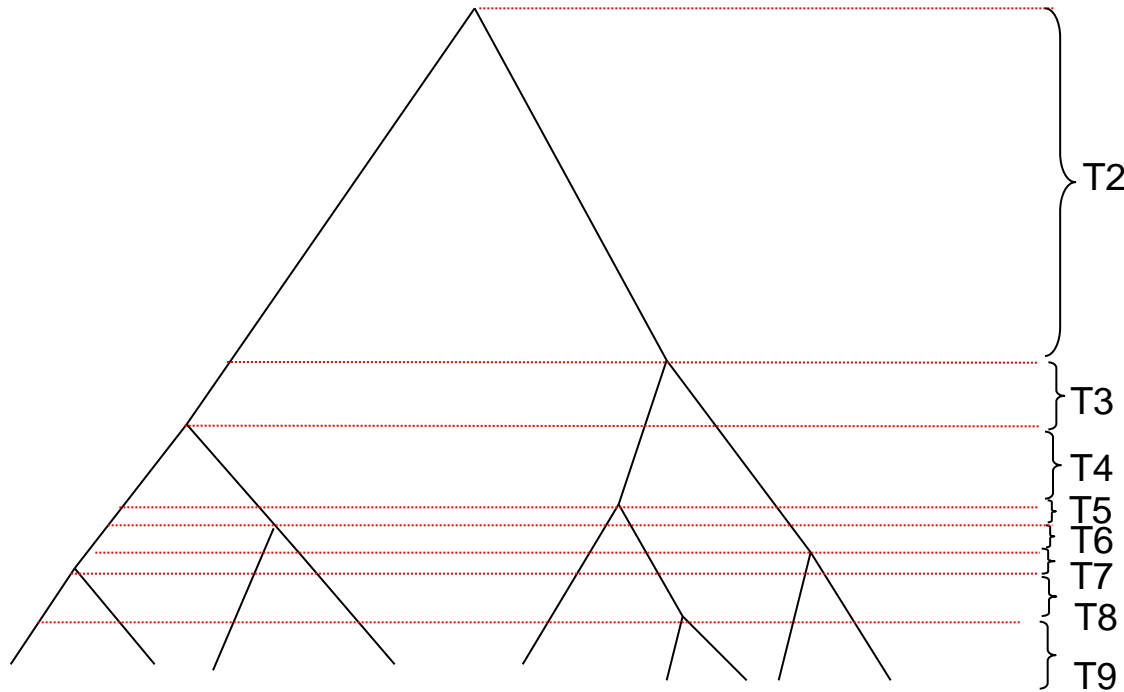
E (tMRCa com $n=2$) = $2N$

Número de pares possíveis = $n(n-1) / 2$

Tempo médio para a primeira coalescência numa amostra de n linhagens:

$$1 / [(n(n-1) / 2)] \cdot 2N = 4N / n(n-1)$$

TEMPO DE COALESCÊNCIA $n=n$



$$\text{TOTAL} = E(t_{\text{MRCA}}) = 3,56 N = 4N [1 - (1/n)]$$

$$4N / 2 (2-1) = 1 \cdot 2N$$

$$\begin{aligned} 4N / 3 (3-1) &= 1/3 \cdot 2N \\ 4N / 4 (4-1) &= 1/6 \cdot 2N \\ 4N / 5 (5-1) &= 1/10 \cdot 2N \\ 4N / 6 (6-1) &= 1/15 \cdot 2N \\ 4N / 7 (7-1) &= 1/21 \cdot 2N \\ 4N / 8 (8-1) &= 1/28 \cdot 2N \\ 4N / 9 (9-1) &= 1/36 \cdot 2N \end{aligned}$$

↓

$$0,78 * 2N$$

TEMPO DE COALESCÊNCIA $n=n$

- $E(t_{MRCA}) = 4N [1 - (1/n)]$

- $n \longrightarrow \infty$

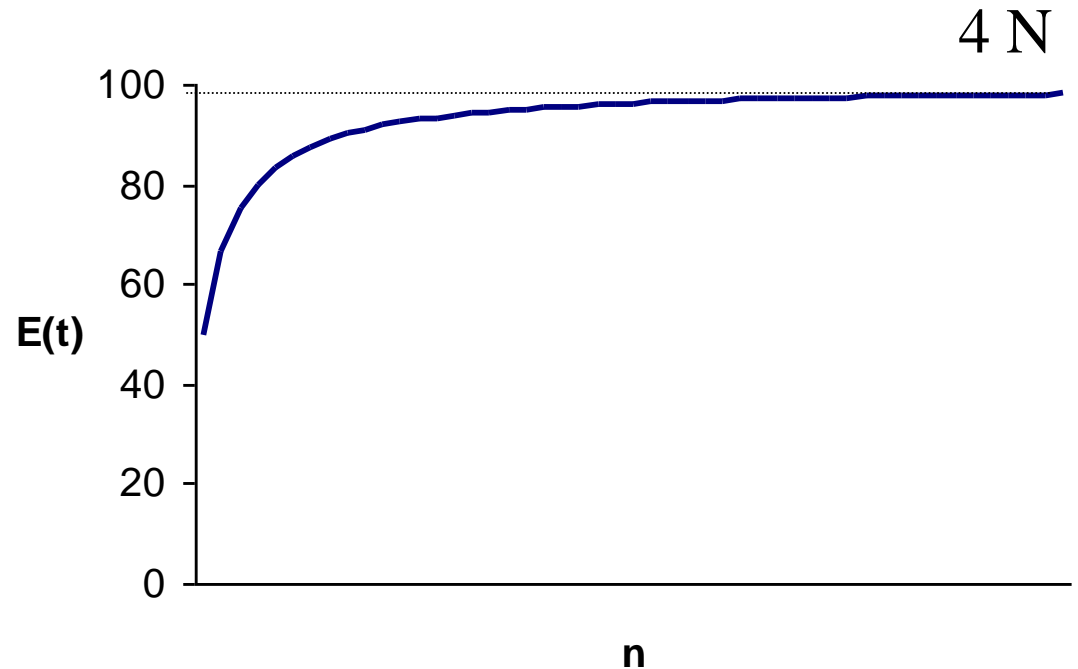
- $E(t_{MRCA}) = 4N$

- Ex:

- $n=2; E(t) = 2N$

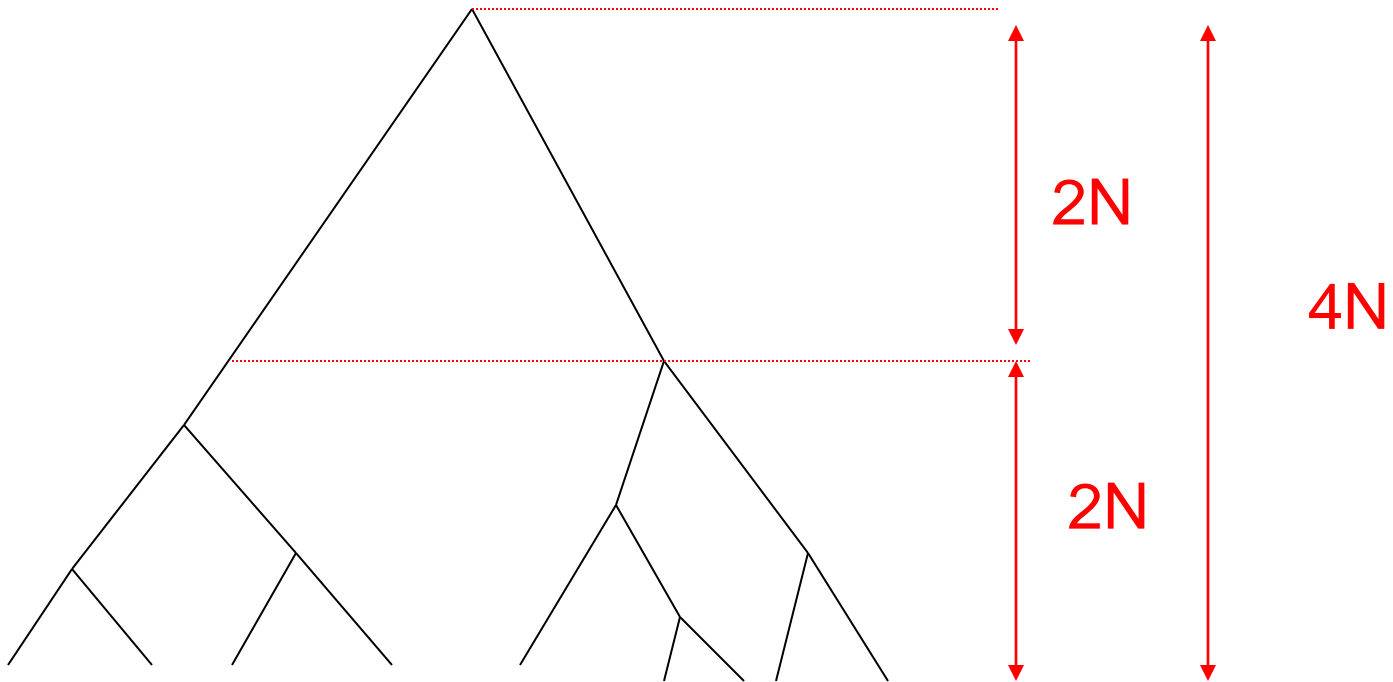
- $n=10; E(t) = 3,6N$

- .
- .
- .
- .



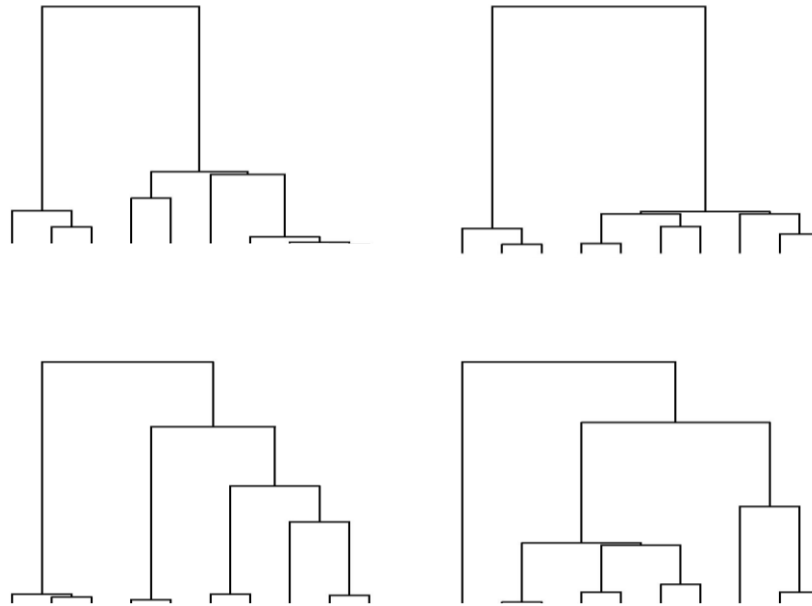
TEMPO DE COALESCÊNCIA

- Propriedades médias



TEMPO DE COALESCÊNCIA

- Aleatoriedade

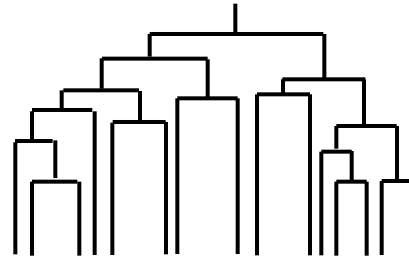


TEMPO DE COALESCÊNCIA

- Consequências
- Inutilidade de aumento da amostra: a estrutura da genealogia é basicamente determinada a partir da altura em que já há poucos ancestrais
- A inclusão da bifurcação mais antiga na amostra é obtida com um número surpreendentemente baixo de indivíduos:
- $(n-1)/(n+1)$ é a probabilidade de incluir o MRCA numa amostra de tamanho n
- A informação obtida a partir de um único gene é muito limitada.
- É melhor aumentar o número de loci analisados do que o número de amostras em cada locus.

ADIÇÃO DE MUTAÇÕES

ADIÇÃO DE MUTAÇÕES



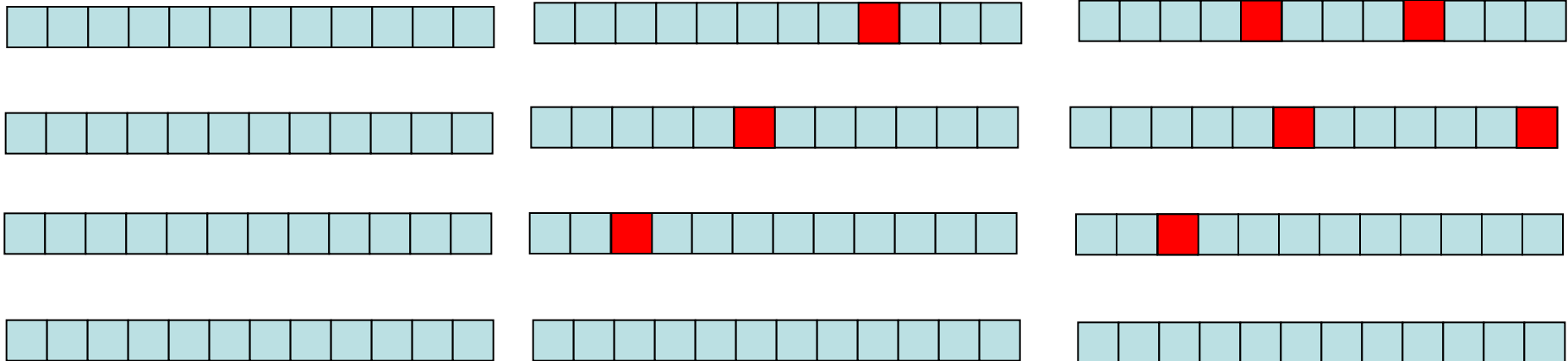
- Os padrões de diversidade genética observados reflectem a sobreposição de mutações numa genealogia

ADIÇÃO DE MUTAÇÕES-ISM

t

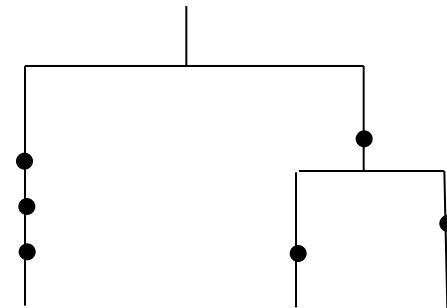
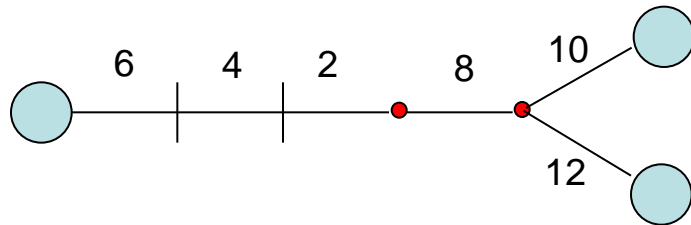
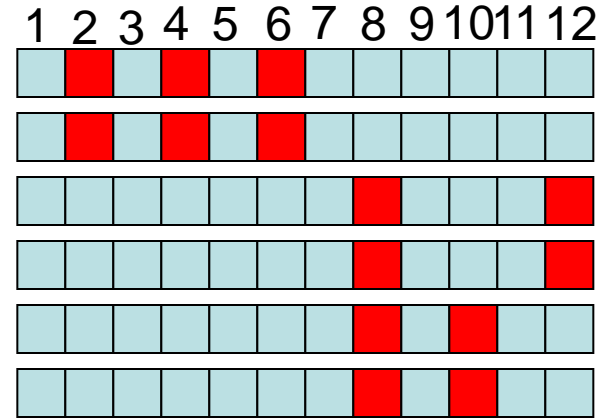
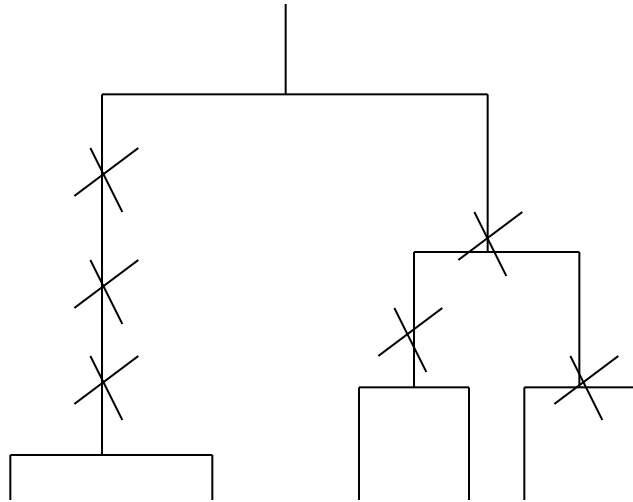
t + x

t + x + y



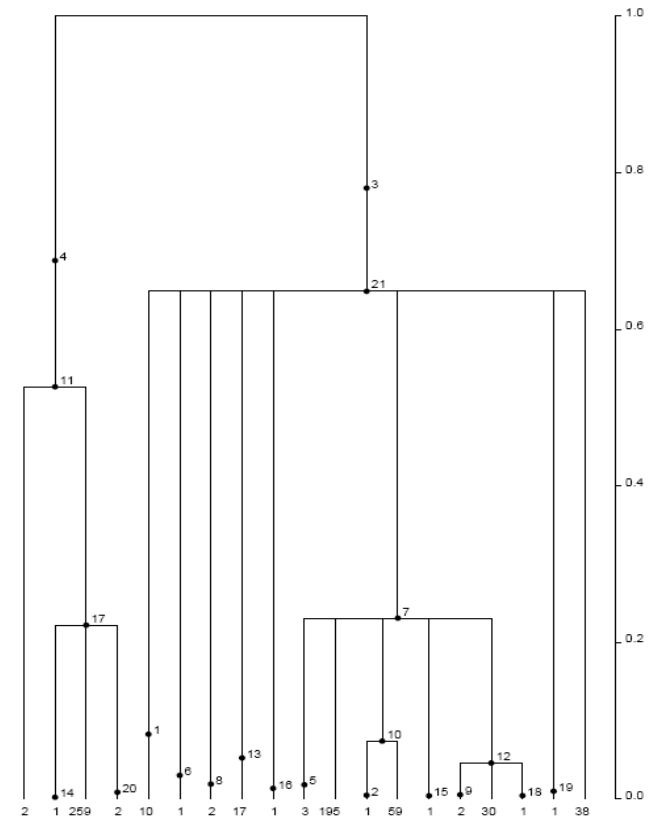
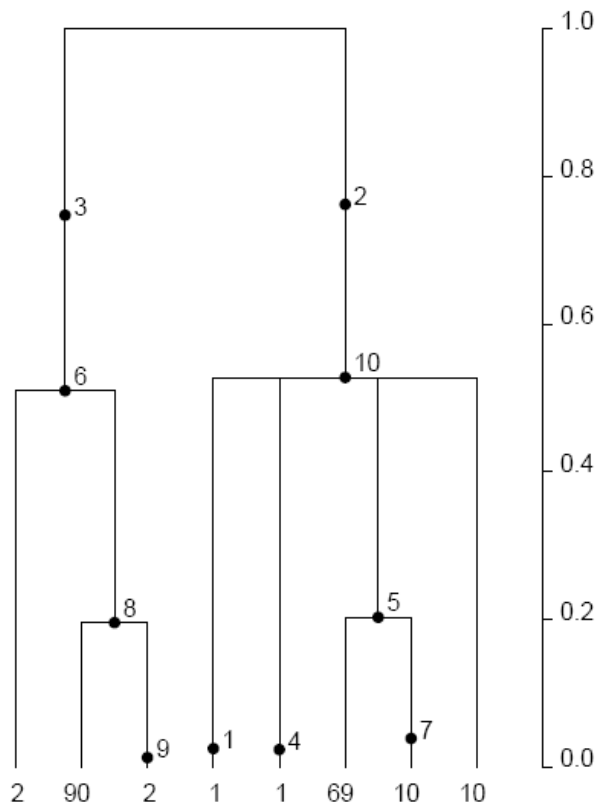
- Cada mutação ocorre numa posição diferente de uma sequência de DNA: ISM

ADIÇÃO DE MUTAÇÕES

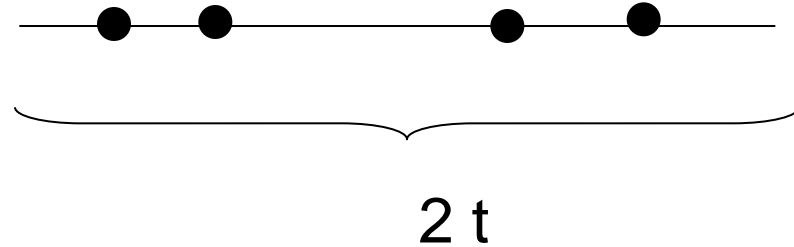
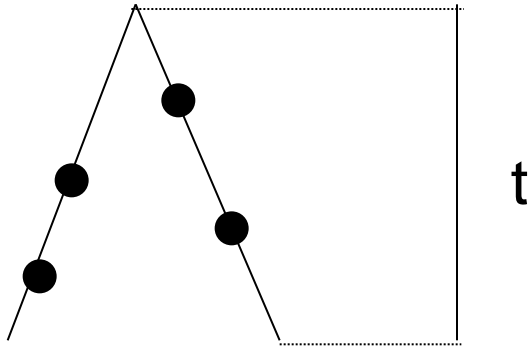


ADIÇÃO DE MUTAÇÕES

Exemplo de aplicação do programa *GENETREE*



MEDIDAS SUMÁRIAS DE DIVERSIDADE GENÉTICA



$$E(\pi) = 2 \mu t$$

$$E(t) = 2 N$$

Nº Médio de diferenças
entre um par de
sequências = π

$$E(\pi) = 2 \mu E(t)$$

μ = taxa de
mutação por
locus por
geração

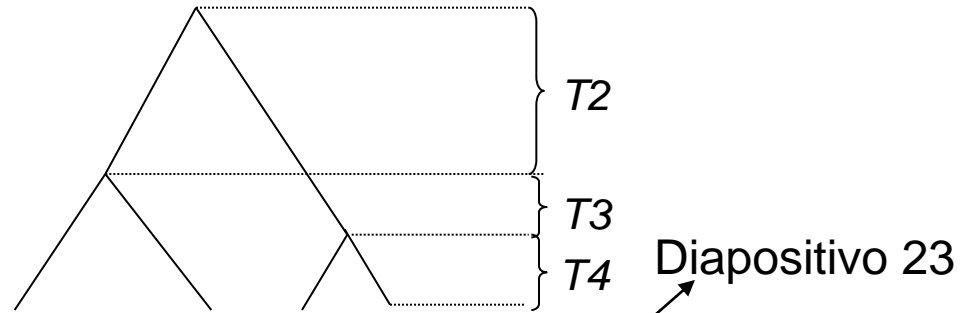
$$\pi = 4N\mu$$

ou

$$\theta = 4N\mu$$

MEDIDAS SUMÁRIAS DE DIVERSIDADE GENÉTICA

Nº de mutações (posições com segregação) numa amostra de tamanho $n = S$



Comprimento total da árvore

$$L = 4T_4 + 3T_3 + 2T_2$$

$$L = \sum_{i=2}^n iT_i \quad E(L) = \sum_{i=2}^n iE(T_i) = \sum_{i=2}^n i \left(\frac{4N}{i(i-1)} \right) = \sum_{i=2}^n \frac{4N}{i-1} = 4N \sum_{i=1}^{n-1} \frac{1}{i}$$

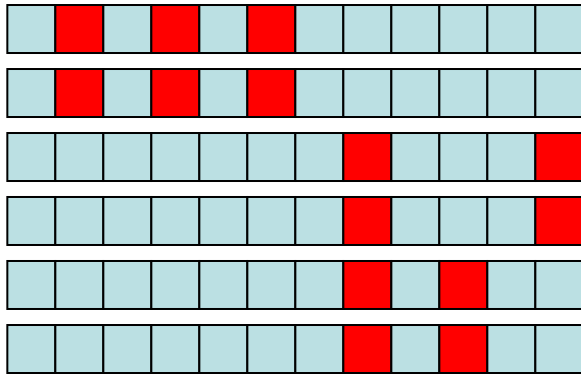
Nº esperado de mutações

$$E(S) = \mu 4N \sum_{i=1}^{n-1} \frac{1}{i} = \theta \sum_{i=1}^{n-1} \frac{1}{i}$$

$$\theta_w = S / \sum_{i=1}^{n-1} \frac{1}{i}$$

MEDIDAS SUMÁRIAS DE DIVERSIDADE GENÉTICA

Exemplo numérico



$$S=6$$

$$\theta = 6 / (1/1 + 1/2 + 1/3 + 1/4 + 1/5)$$

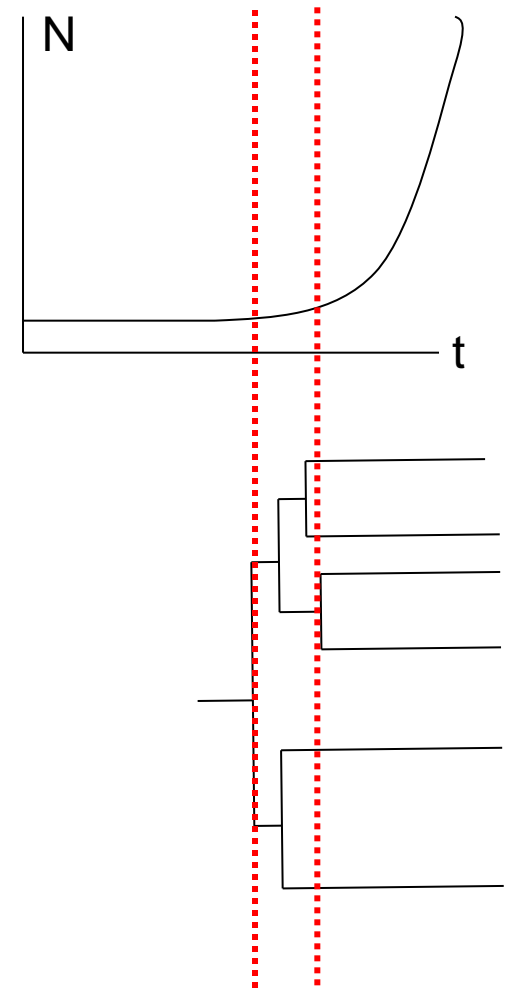
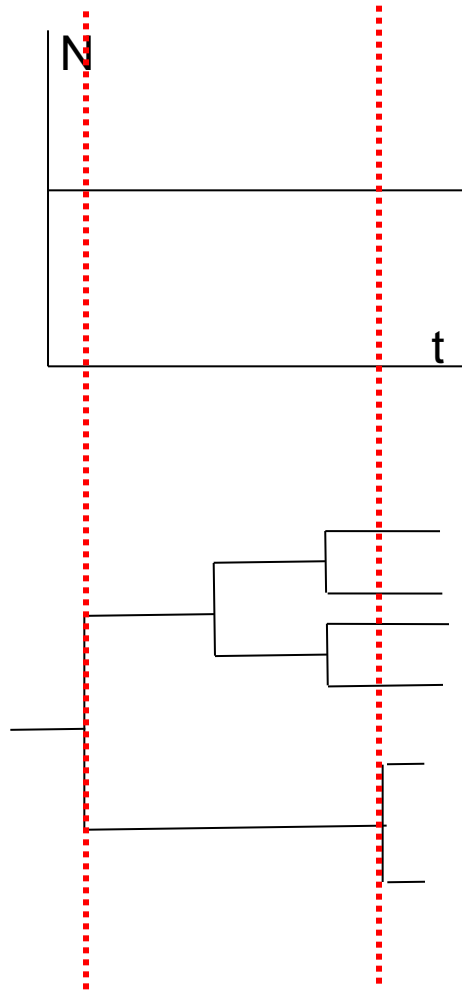
$$\theta = 2,63$$

$$\theta_w = S / \sum_{i=1}^{n-1} \frac{1}{i}$$

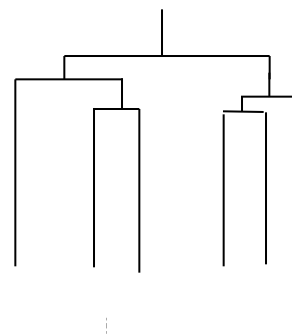
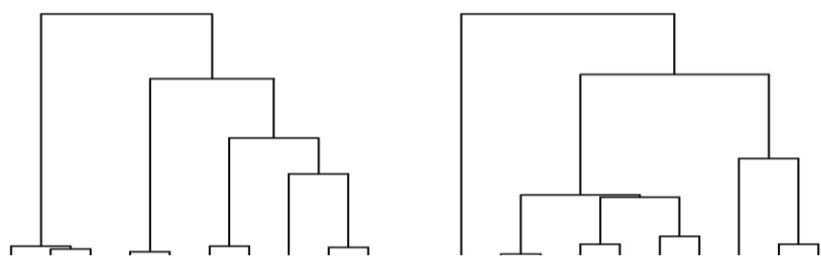
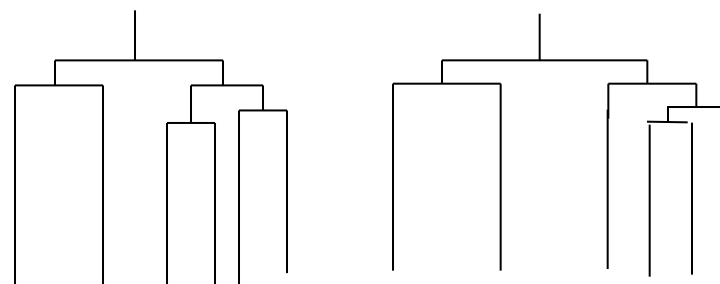
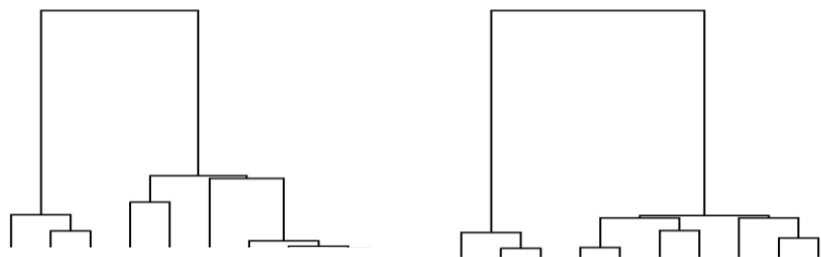
VARIAÇÃO NO TAMANHO POPULACIONAL

Crescimento da população

Topologia das genealogias



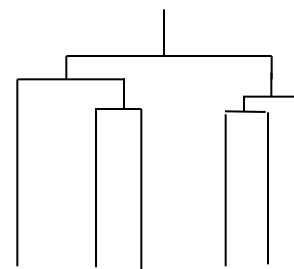
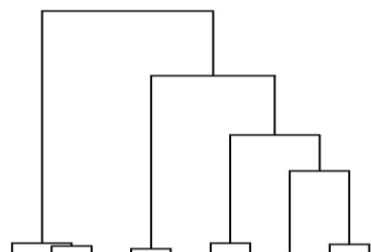
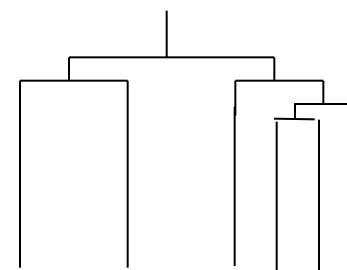
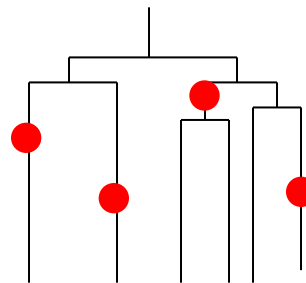
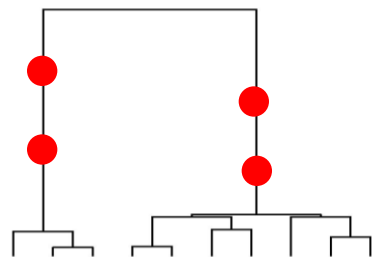
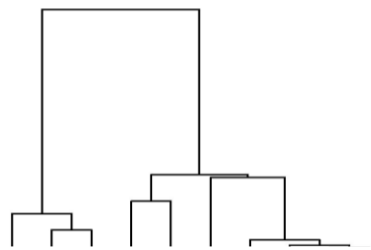
- Aleatoriedade das genealogias



Estável
(maior aleatoriedade)

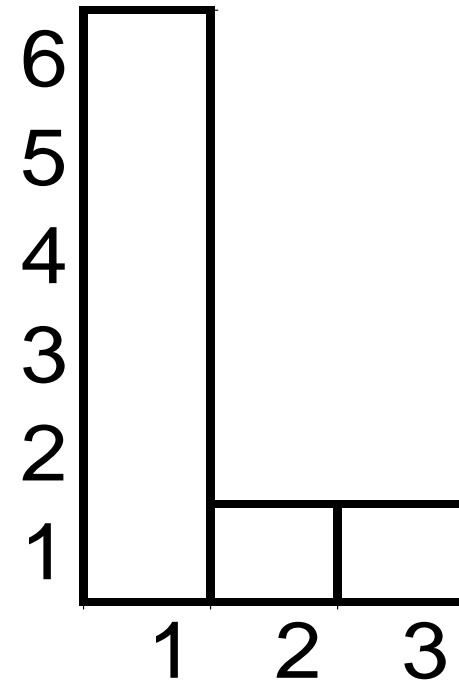
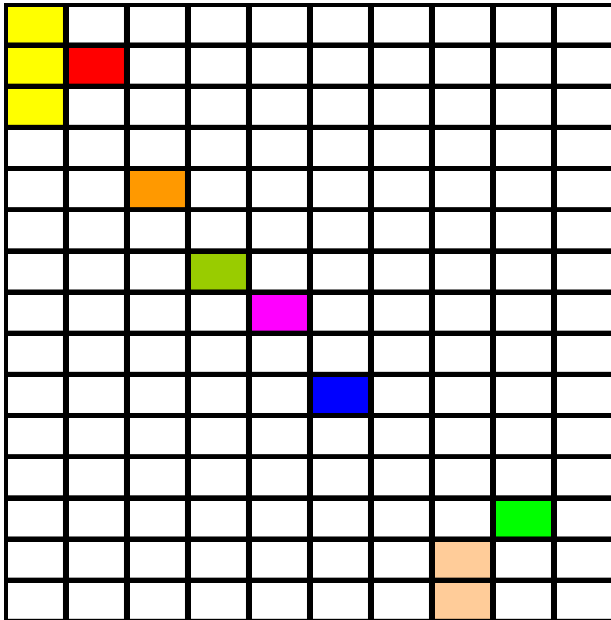
Crescimento
(menor aleatoriedade)

Distribuição de mutações



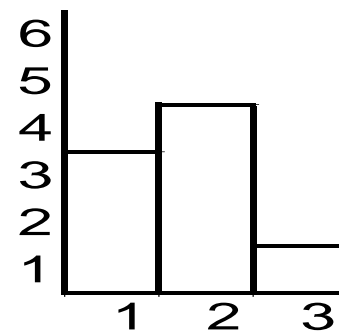
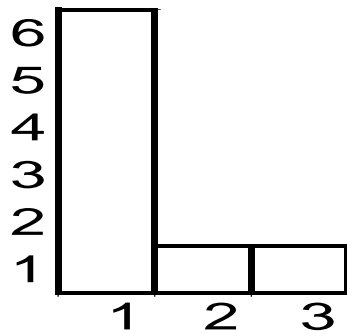
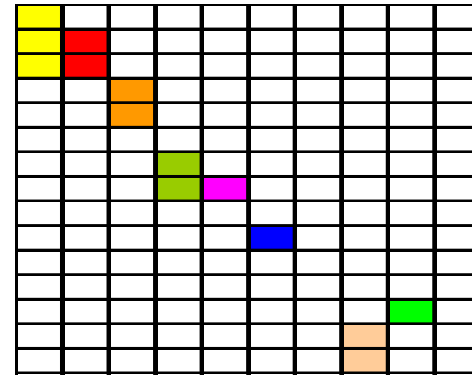
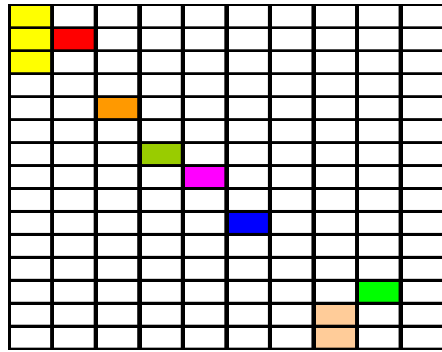
Crescimento populacional

Espectro de frequências

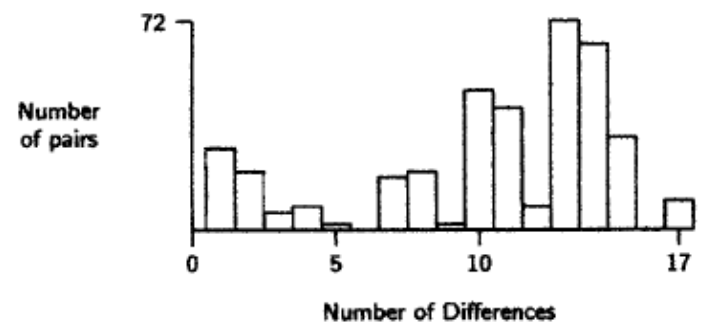
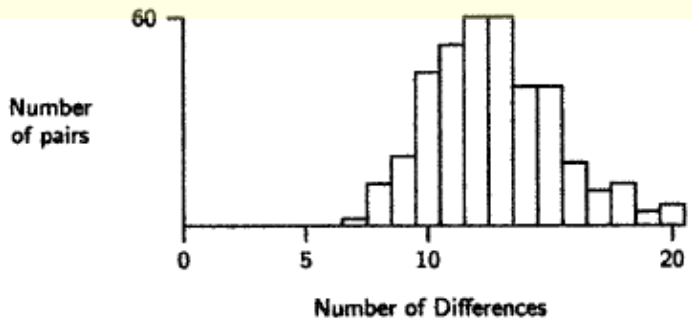
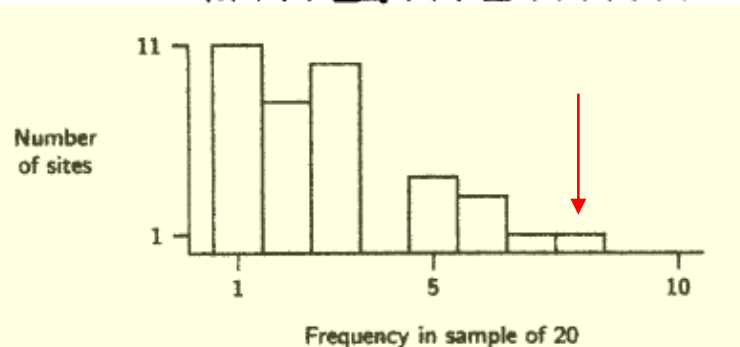
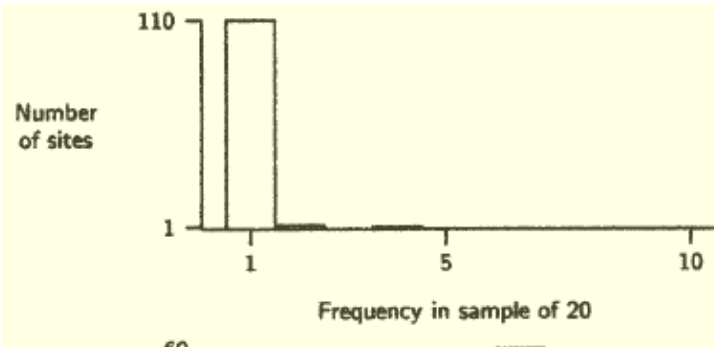
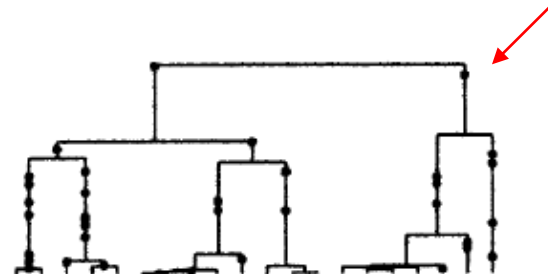
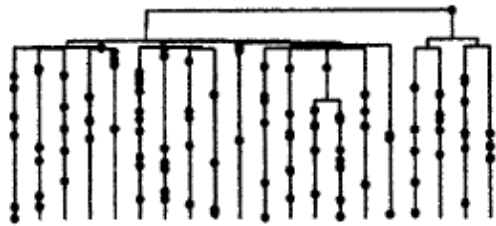


Crescimento populacional

Espectro de frequências

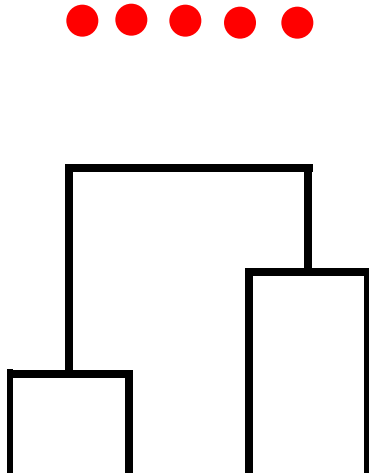


Crescimento populacional



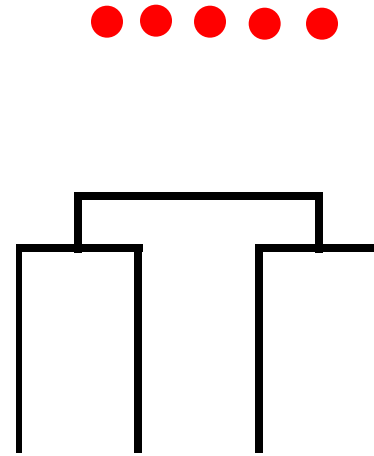
Crescimento populacional

- O D de Tajima



$$\theta_w = S / \sum_{i=1}^{n-1} \frac{1}{i} = 2,73$$
$$\pi = \theta \pi = 2,25$$

$$D = \frac{\pi - \theta}{sd(\pi - \theta)} = \frac{2,25 - 2,73}{sd(\pi - \theta)} = \frac{-0,48}{sd(\pi - \theta)}$$

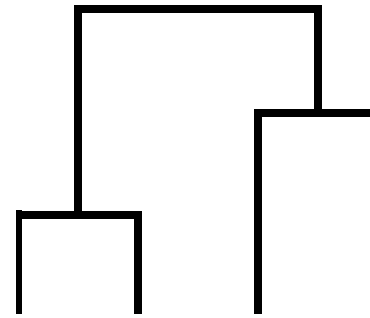
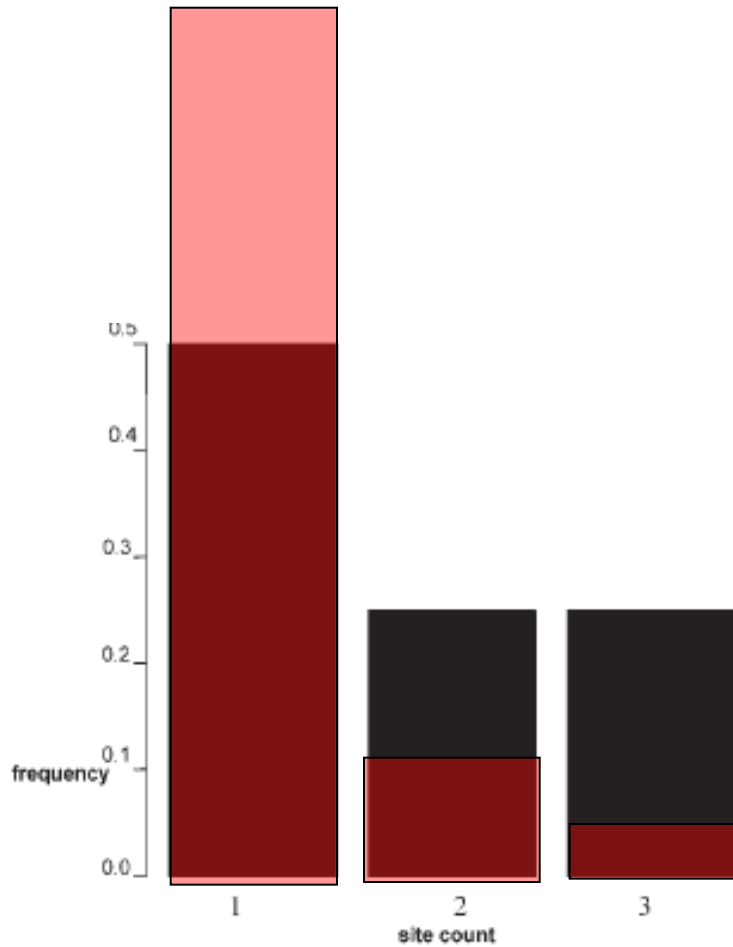


$$\theta_w = S / \sum_{i=1}^{n-1} \frac{1}{i} = 2,73$$
$$\pi = \theta \pi = 1,875$$

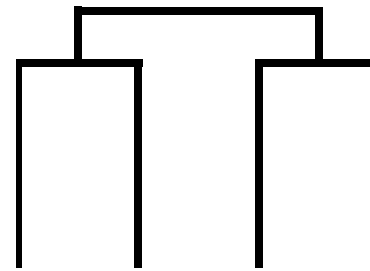
$$D = \frac{\pi - \theta}{sd(\pi - \theta)} = \frac{1,875 - 2,73}{sd(\pi - \theta)} = \frac{-0,855}{sd(\pi - \theta)}$$

Crescimento populacional

- O D de Tajima



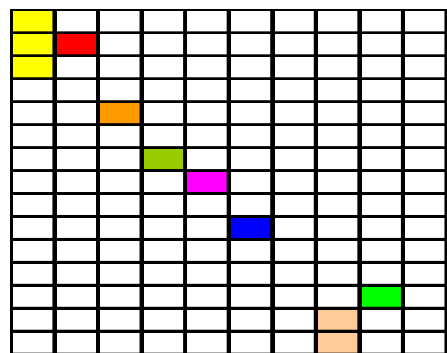
$D \sim 0$



$D < 0$

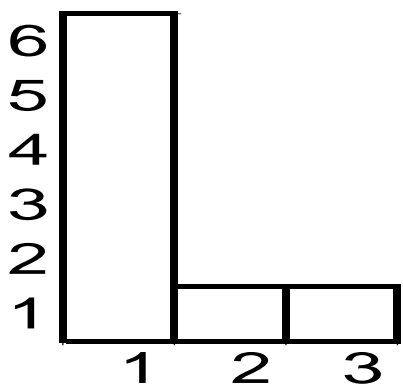
Crescimento populacional

Teste de D

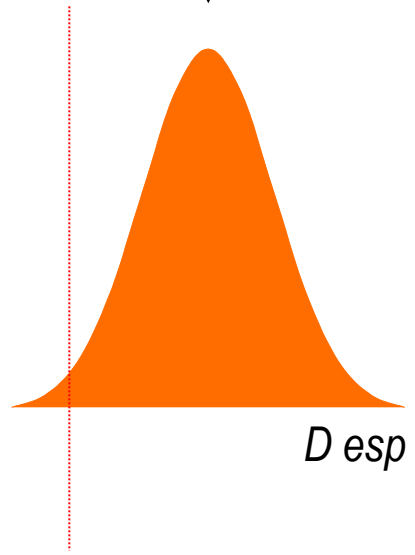


θ, π, S

Simulações/WF, π, θ, S



D_{obs}

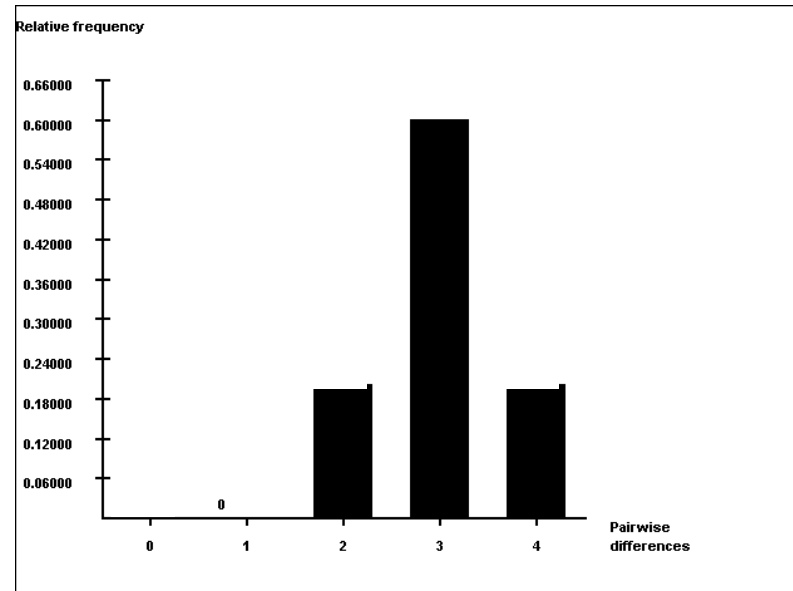
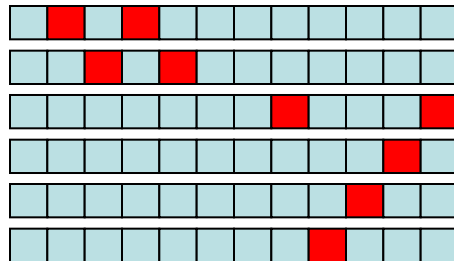
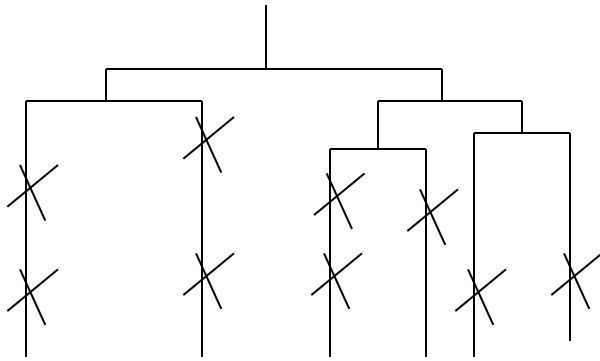


D_{esp}

Crescimento populacional

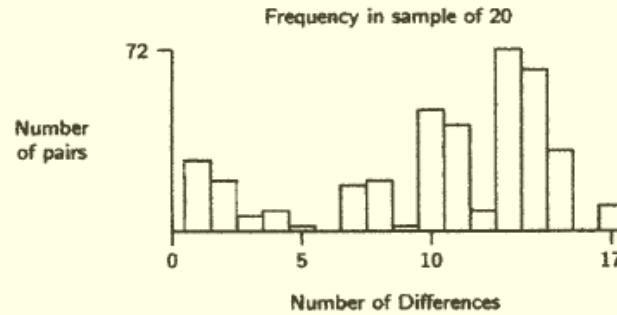
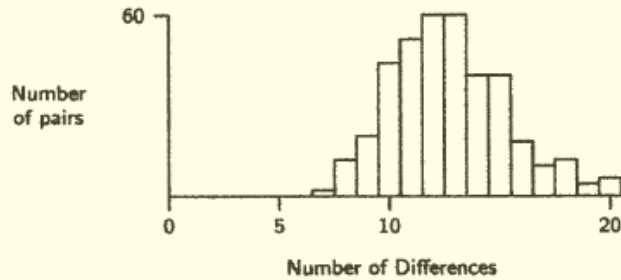
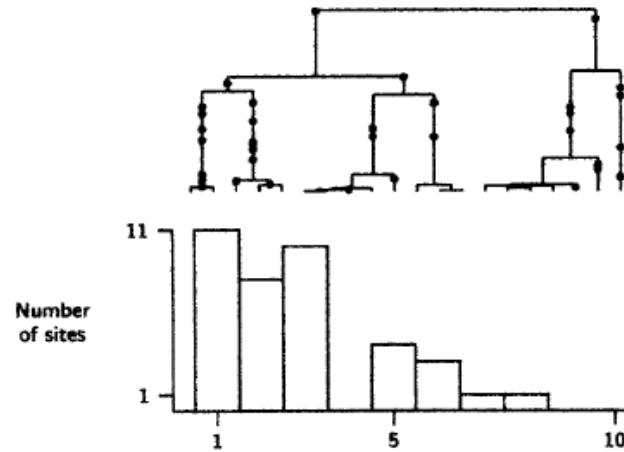
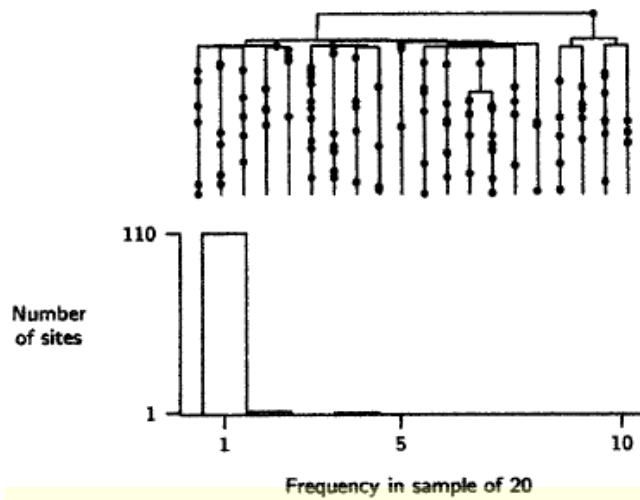
Número de diferenças entre pares de sequências comparadas (*mismatch distributions*).

Média= π



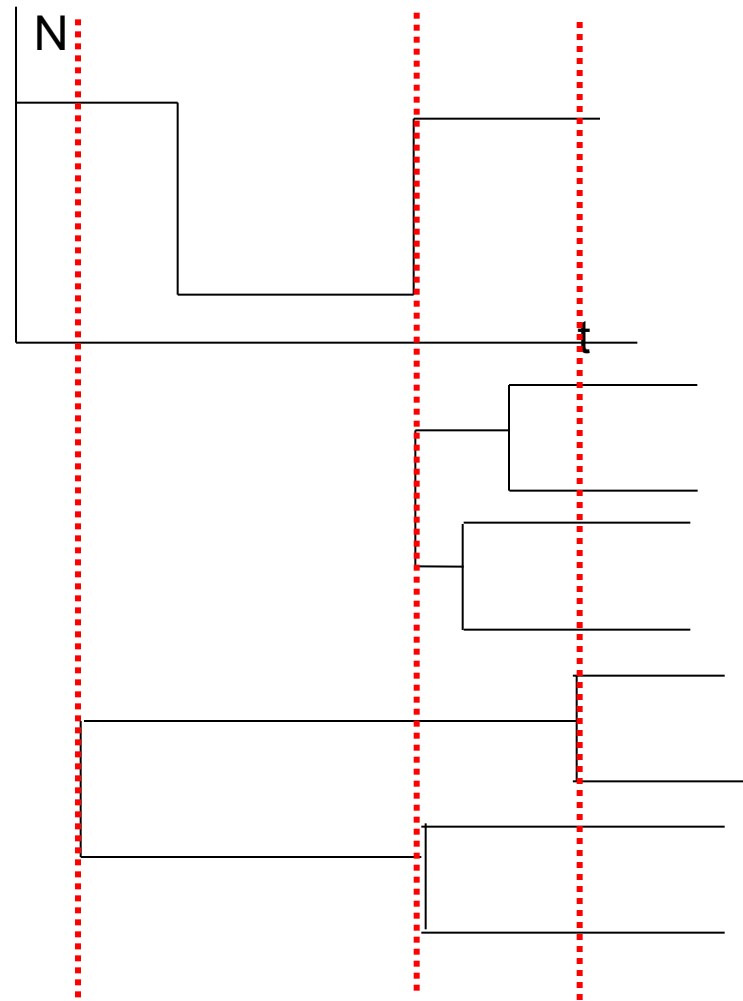
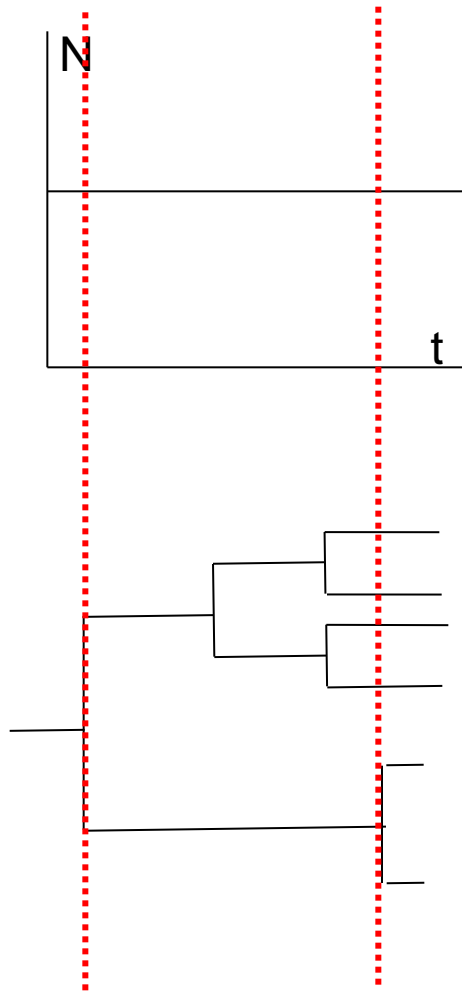
Unweighted mean pairwise difference: 3.000

Crescimento populacional

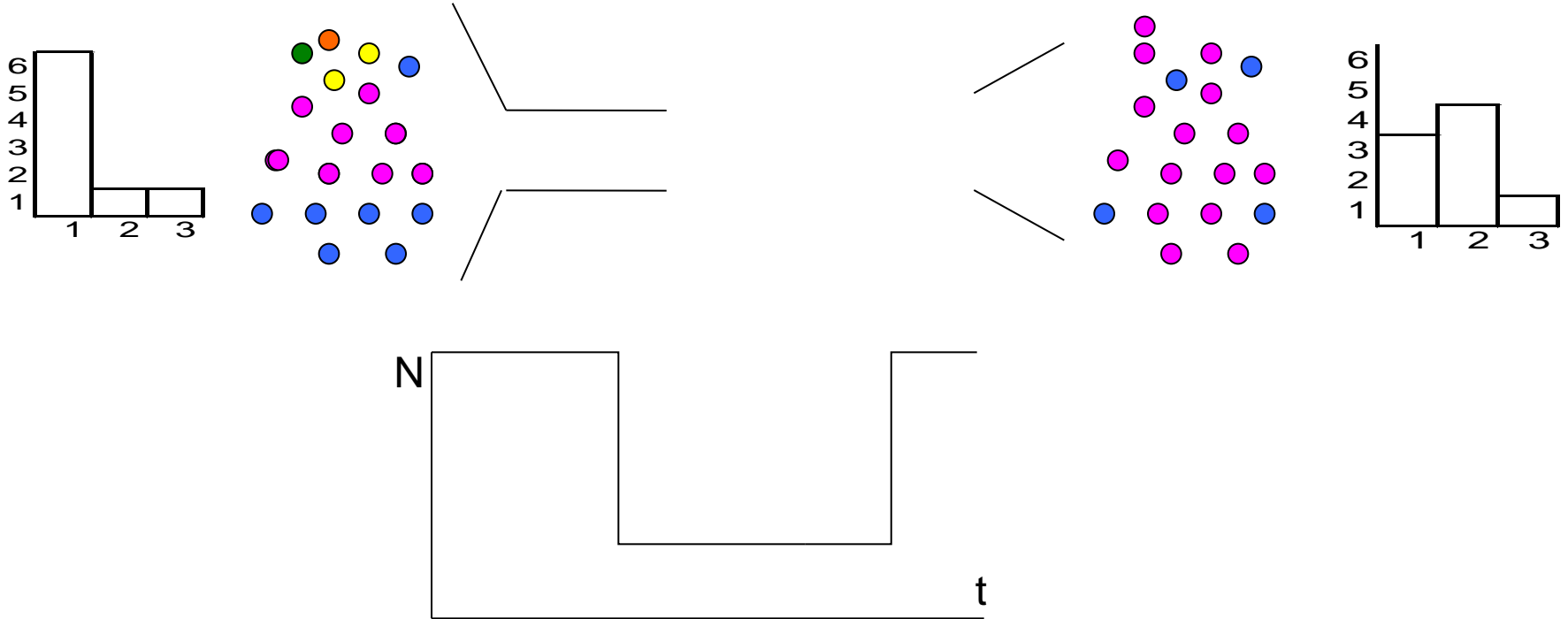


Diminuição da população

Topologia das genealogias



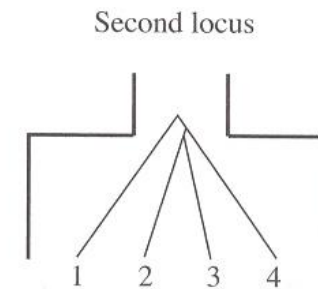
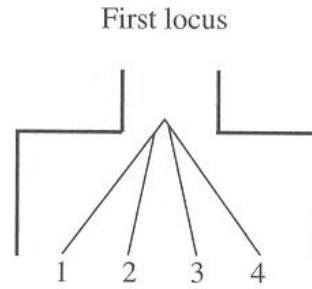
Diminuição da população



Alterações no tamanho

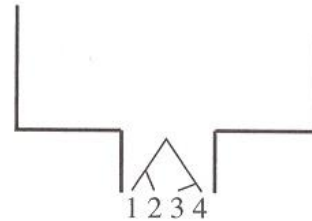
(a)

Population growth:
Small $\text{Var}[S]$
Excess low freq SNPs



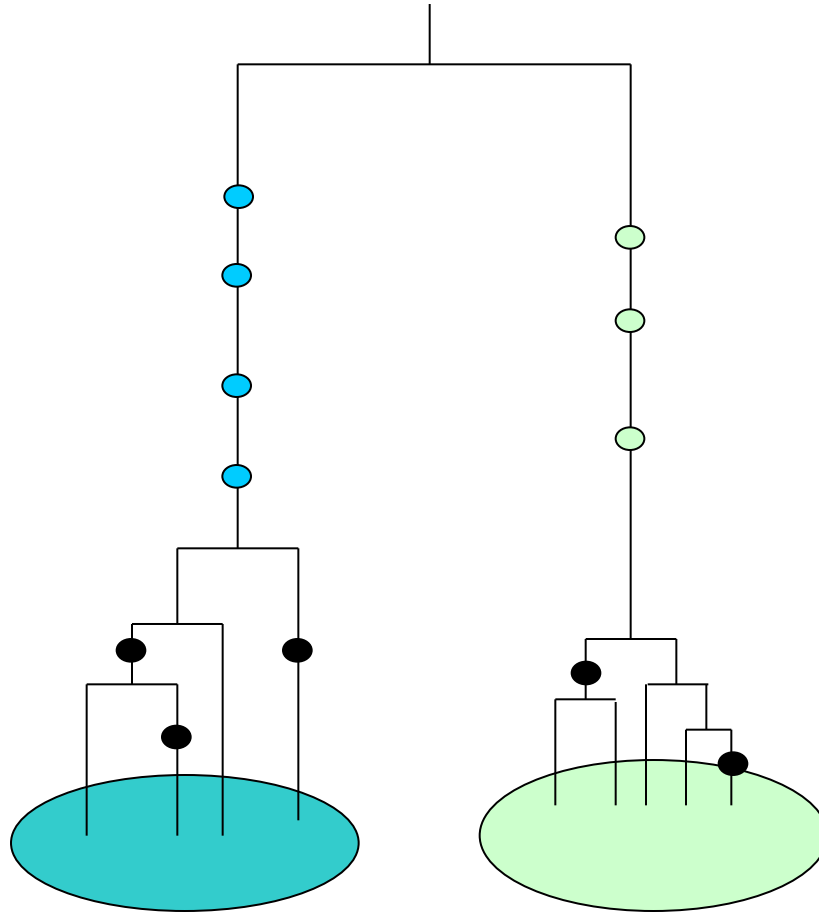
(b)

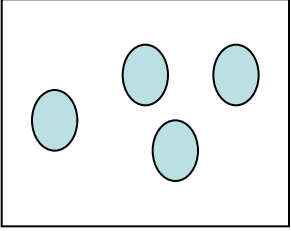
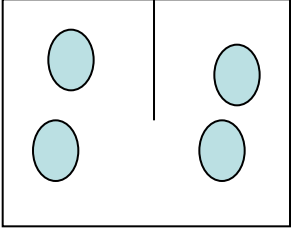
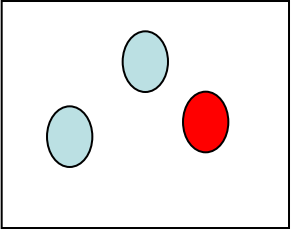
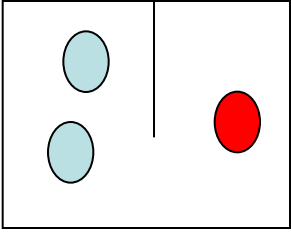
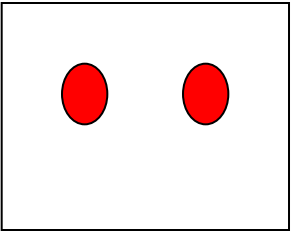
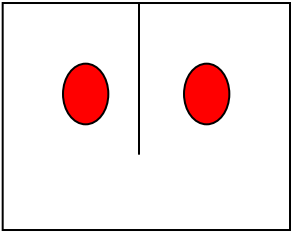
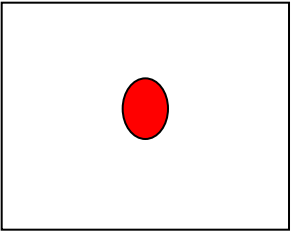
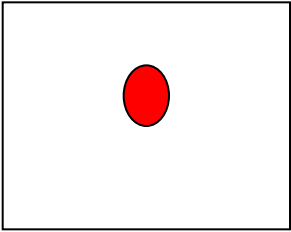
Population decline:
Large $\text{Var}[S]$
Excess middle freq SNPs



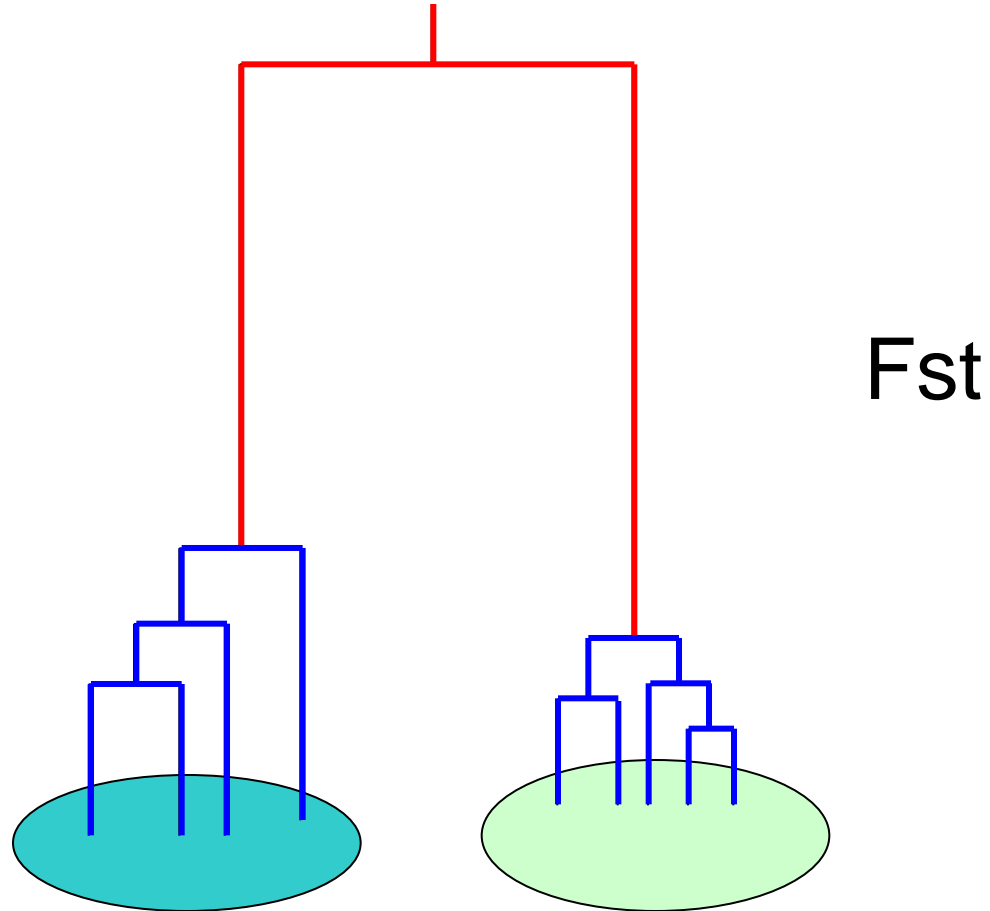
Subdivisão populacional

Subdivisão da população

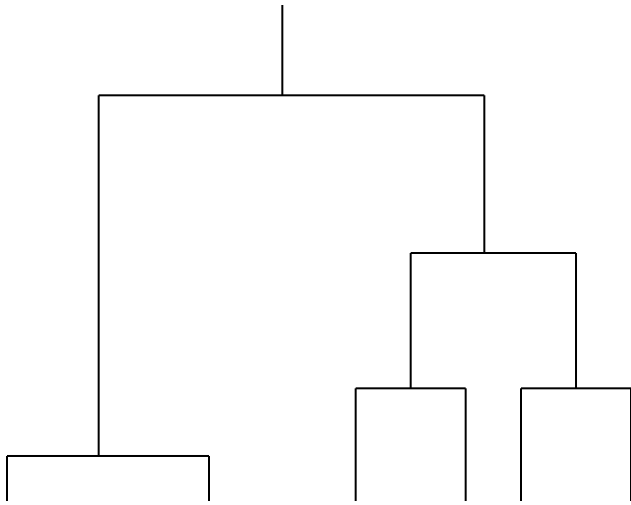




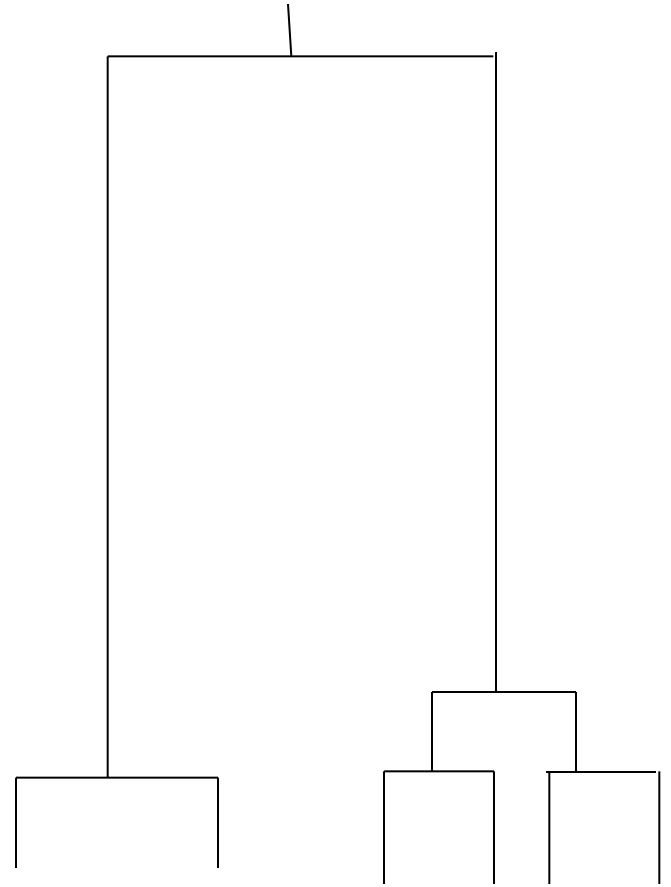
Subdivisão da população



Subdivisão populacional

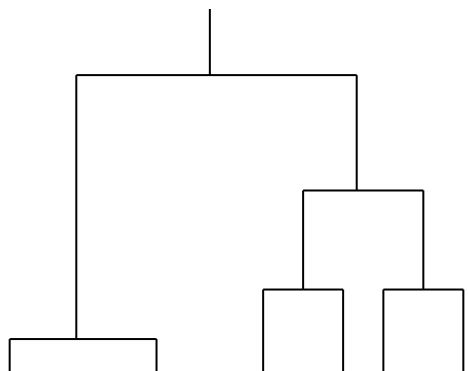


PANMIXIA

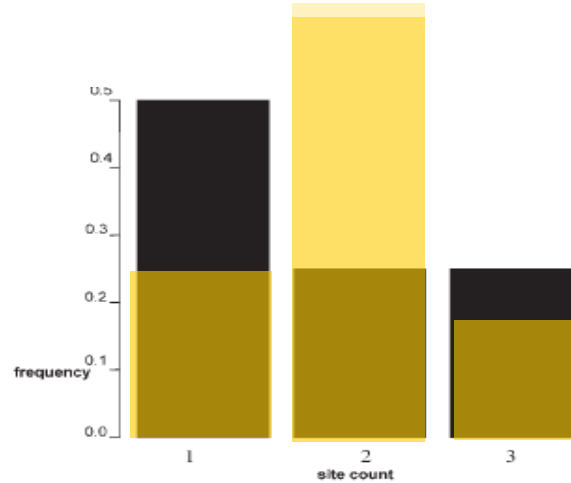
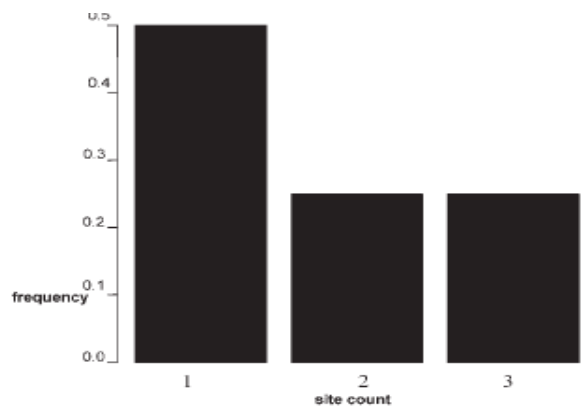
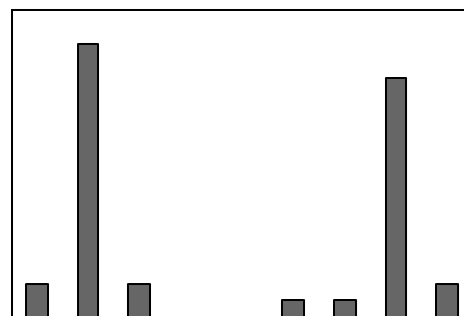
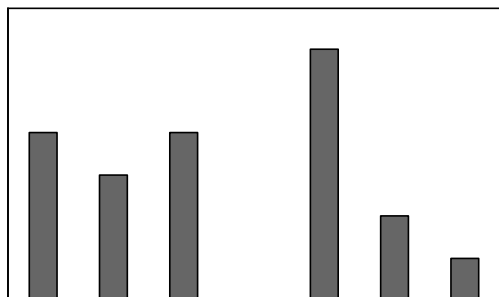
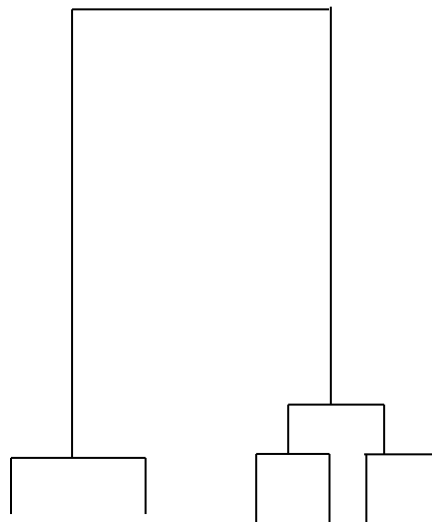


SUBESTRUTURAÇÃO

$D \sim 0$



$D > 0$



Alterações no tamanho e subdivisão

