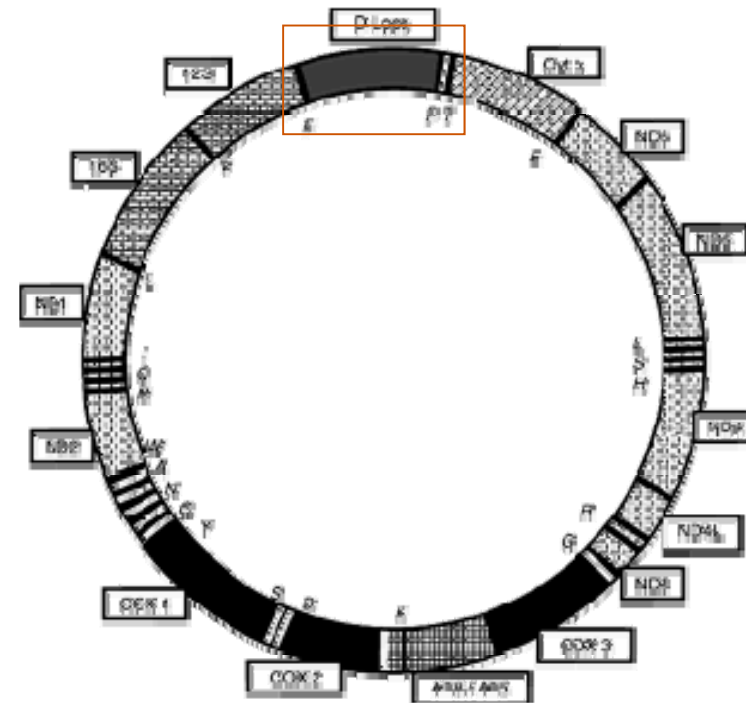
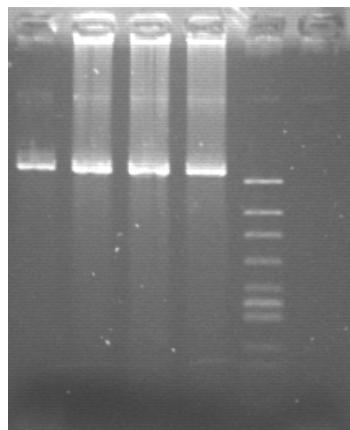


Estudo da diversidade com sequências de DNA

estudo do DNA por sequenciação

extracção de DNA

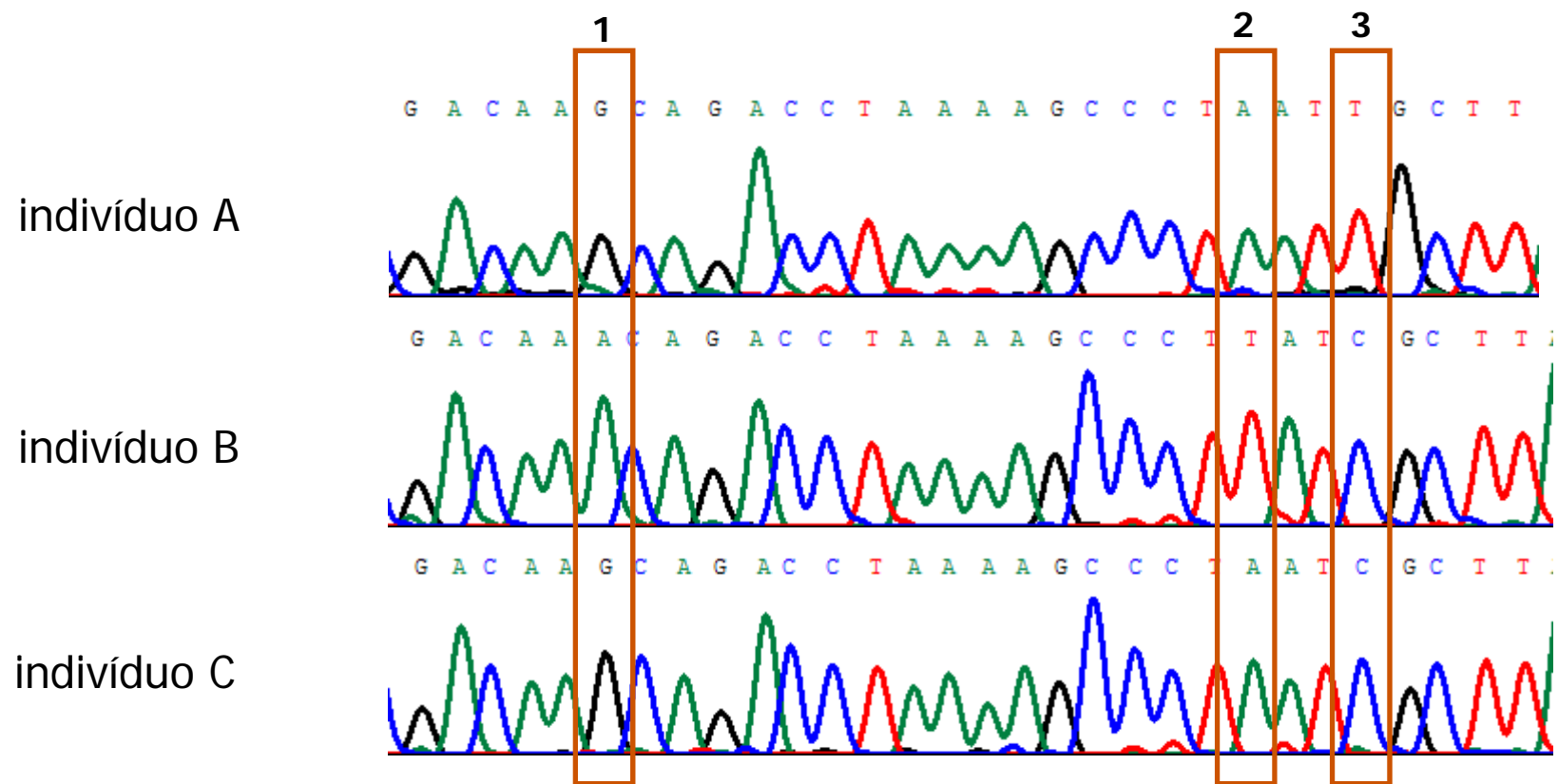
PCR do fragmento desejado



sequenciação

estudo do DNA por sequenciação

✓ sequenciação



Alguns índices de diversidade:

π (diversidade nucleotídica):

Número médio de diferenças (por site) entre duas sequências retiradas ao acaso de uma amostra.

H (diversidade haplotípica):

Probabilidade de, retirando aleatoriamente dois haplótipos de uma amostra, estes serem diferentes

S (número de sítios variáveis)

θ (population mutation parameter)

Nível esperado de variabilidade assumindo equilíbrio entre mutação e deriva e ausência de selecção. Estabelece uma relação entre o efectivo populacional e a taxa de mutação.

$$\theta = 4N_e\mu$$

autossomas

$$\theta = 3N_e\mu$$

cromossoma X

$$\theta = N_e\mu$$

cromossoma Y, mitocondrial

Em vez de uma só maneira de o calcular, existem diferentes estimadores que aproximam este valor (exemplo: θ de Watterson (1975)). Usa-se este valor para estimar o N_e .

PROGRAMAS QUE CALCULAM PARÂMETROS DE DIVERSIDADE

DNAsp

Arlequin

SITES

```
Output of: ChioCytb.fas
DNA Polymorphism

Input Data File: C:\...Aula P2\ChioCytb.fas
Number of sequences: 124   Number of sequences used: 124
Selected region: 1-700   Number of sites: 700
Total number of sites (excluding sites with gaps / missing data): 700

Number of polymorphic (segregating) sites, S: 34
Total number of mutations, Eta: 34

Number of Haplotypes, h: 30
Haplotype (gene) diversity, Hd: 0.834
  Variance of Haplotype diversity: 0.00087
  Standard Deviation of Haplotype diversity: 0.030

Nucleotide diversity, Pi: 0.00733
Theta (per site) from Eta: 0.00901

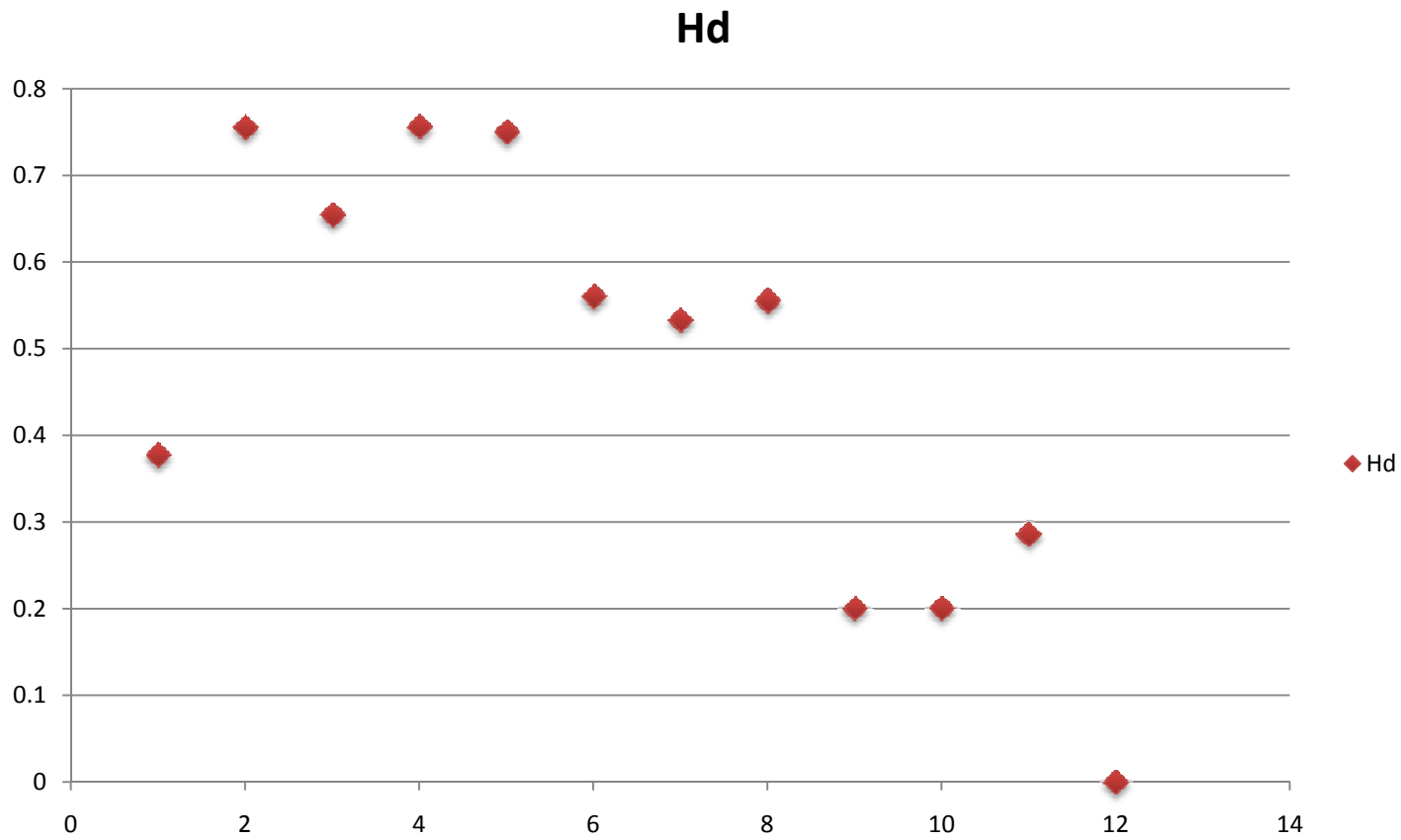
Theta (per site) from S, Theta-W: 0.00901
  Variance of theta (no recombination): 0.0000066
  Standard deviation of theta (no recombination): 0.00256
  Variance of theta (free recombination): 0.0000024
  Standard deviation of theta (free recombination): 0.00154

Finite Sites Model
  Theta (per site) from Pi: 0.00740
  Theta (per site) from S: 0.00925
  Theta (per site) from Eta: 0.00911

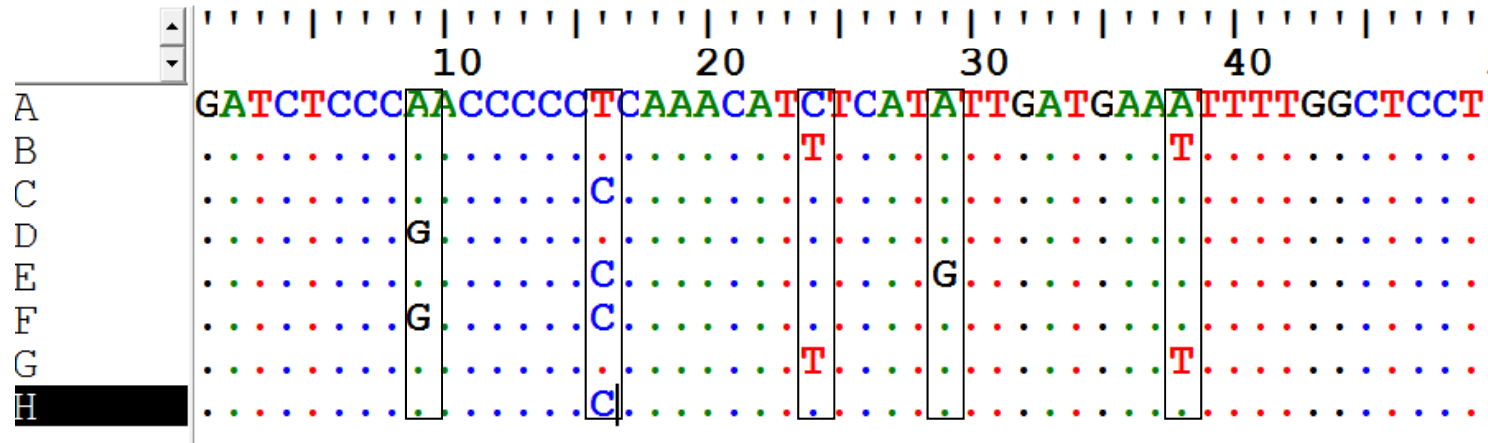
Average number of nucleotide differences, k: 5.130
  Stochastic variance of k (no recombination), Vst(k): 6.165
  Sampling variance of k (no recombination), Vs(k): 0.101
  Total variance of k (no recombination), V(k): 6.266
  Stochastic variance of k (free recombination), Vst(k): 1.710
  Sampling variance of k (free recombination), Vs(k): 0.028
  Total variance of k (free recombination), V(k): 1.738

Theta (per sequence) from S, Theta-W: 6.304
  Variance of theta (no recombination): 3.224
  Variance of theta (free recombination): 1.169
```

- A diversidade diminui nas populações mais a norte.

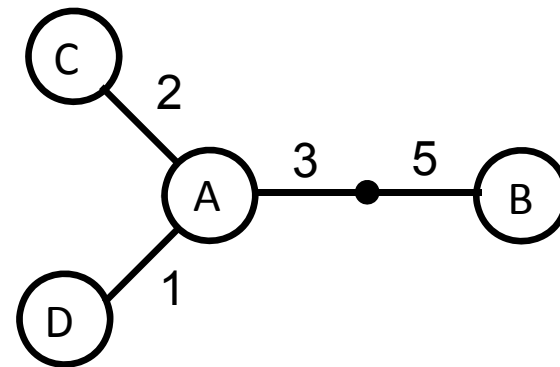
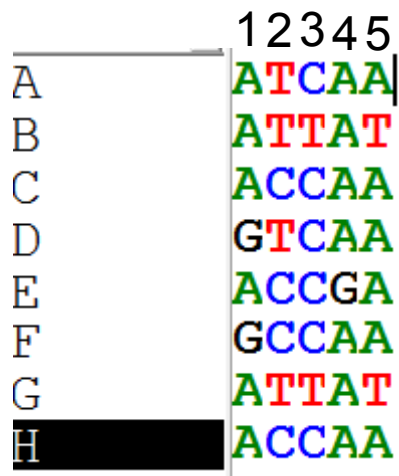
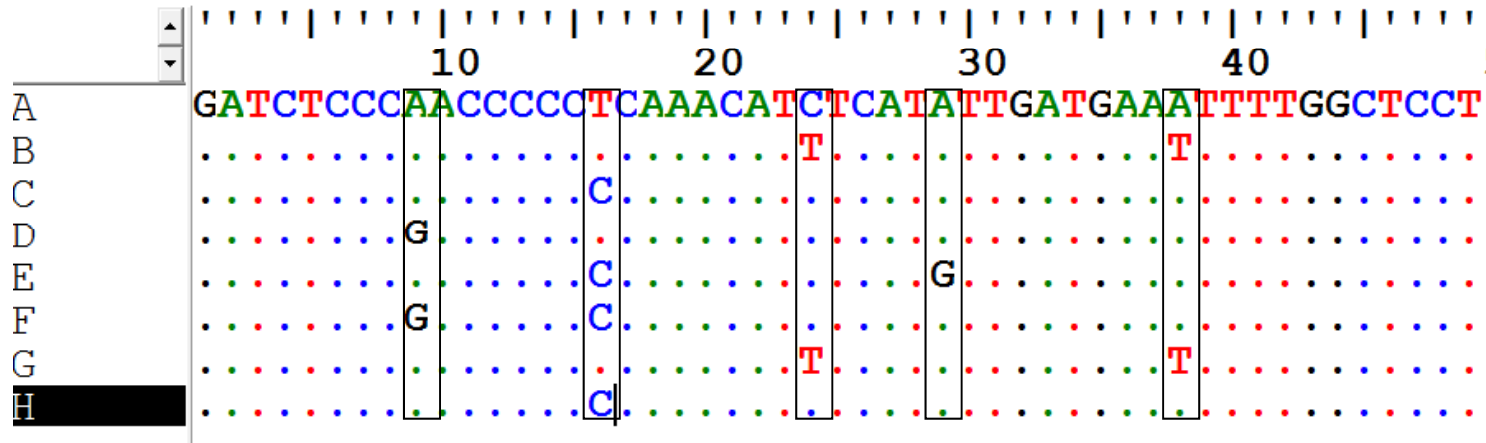


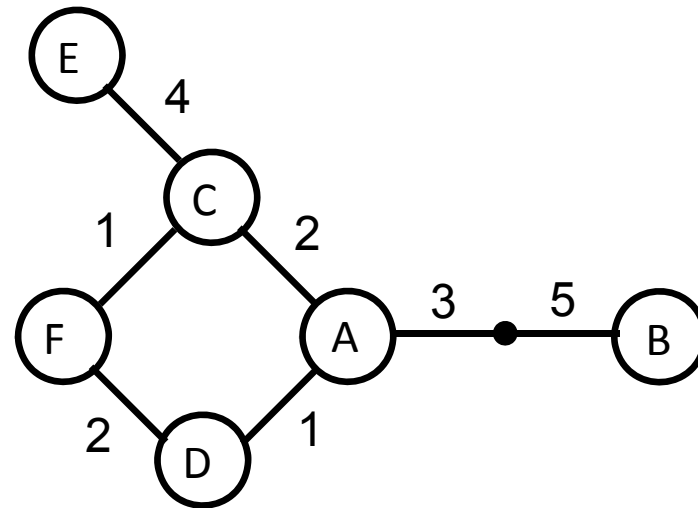
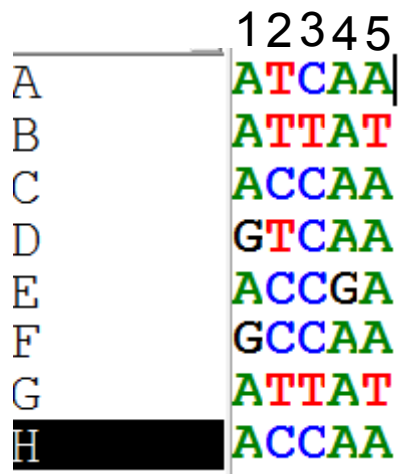
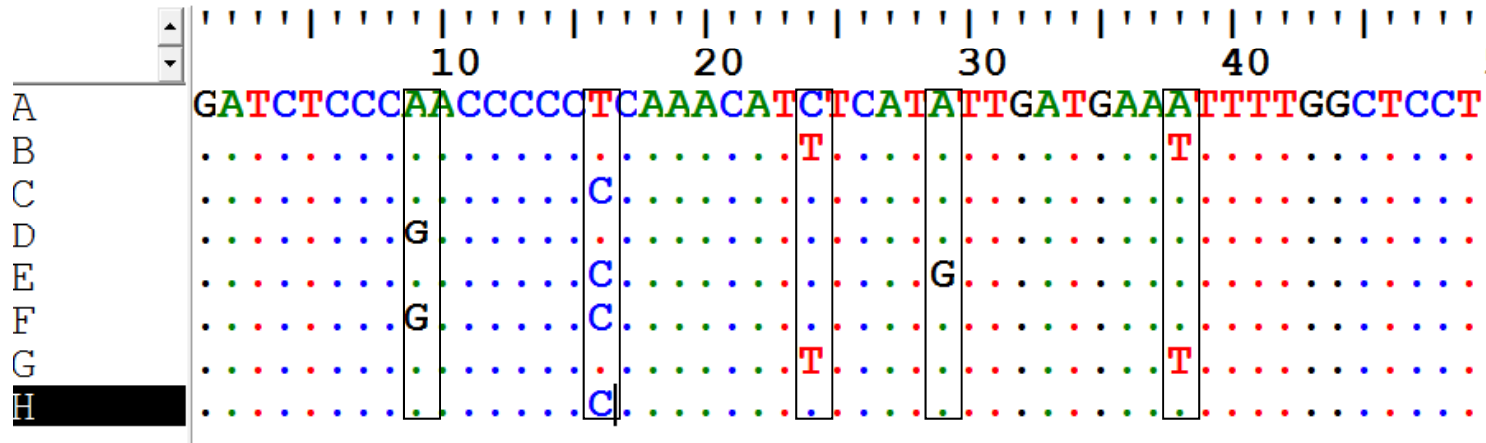
Construção de redes de haplótipos

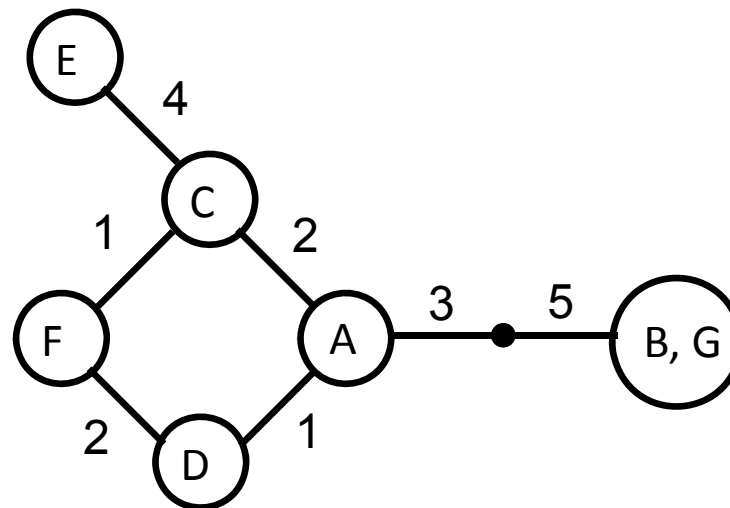
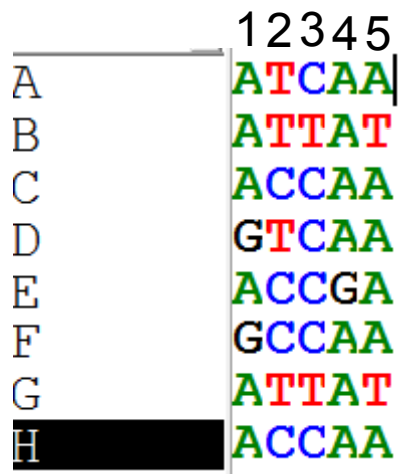
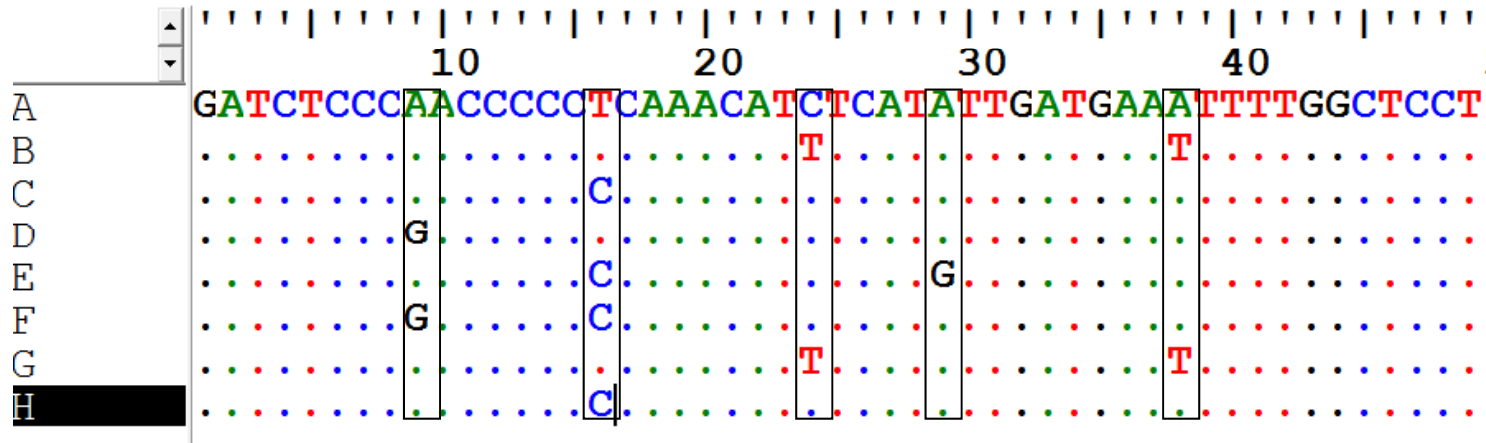


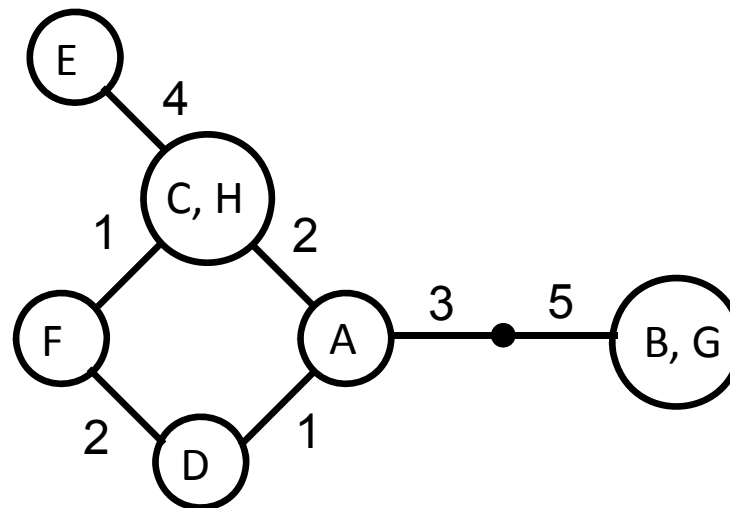
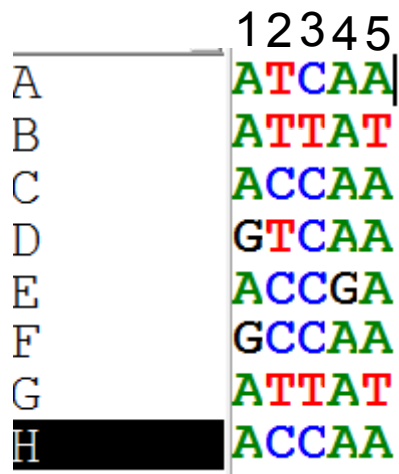
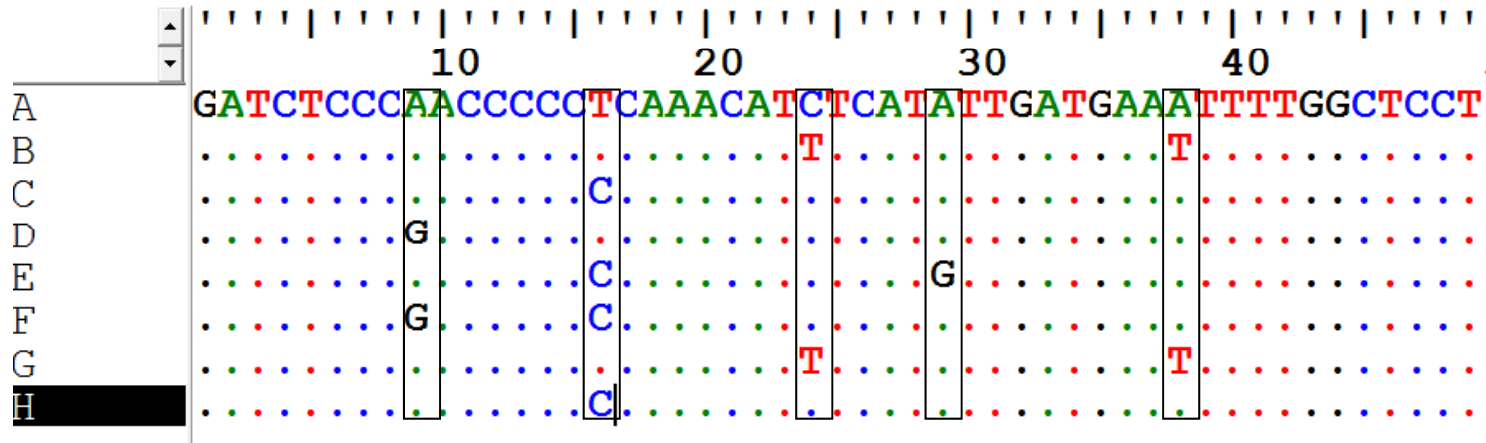
	1	2	3	4	5
A	A	T	C	A	A
B	A	T	T	A	T
C	A	C	C	A	A
D	G	T	C	A	A
E	A	C	C	G	A
F	G	C	C	A	A
G	A	T	T	A	T
H	A	C	C	A	A

A









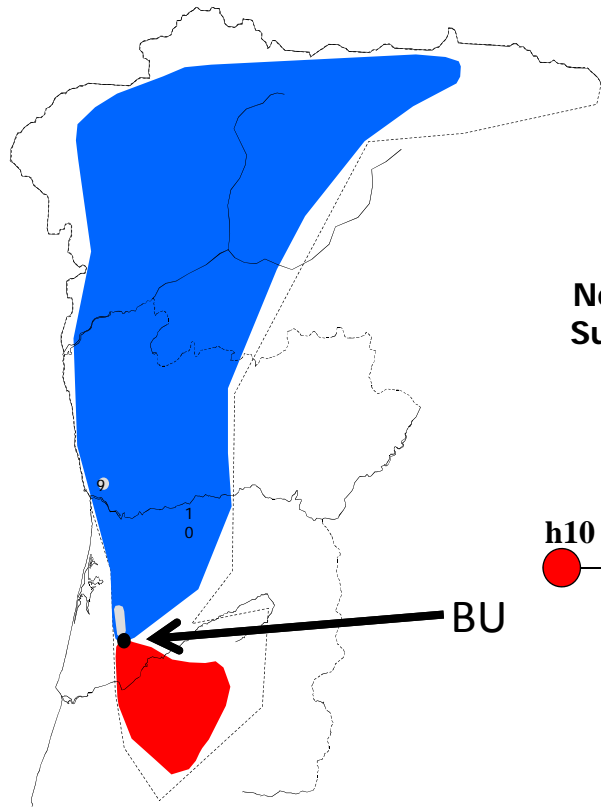
Softwares que fazem redes de haplótipos:

-Network

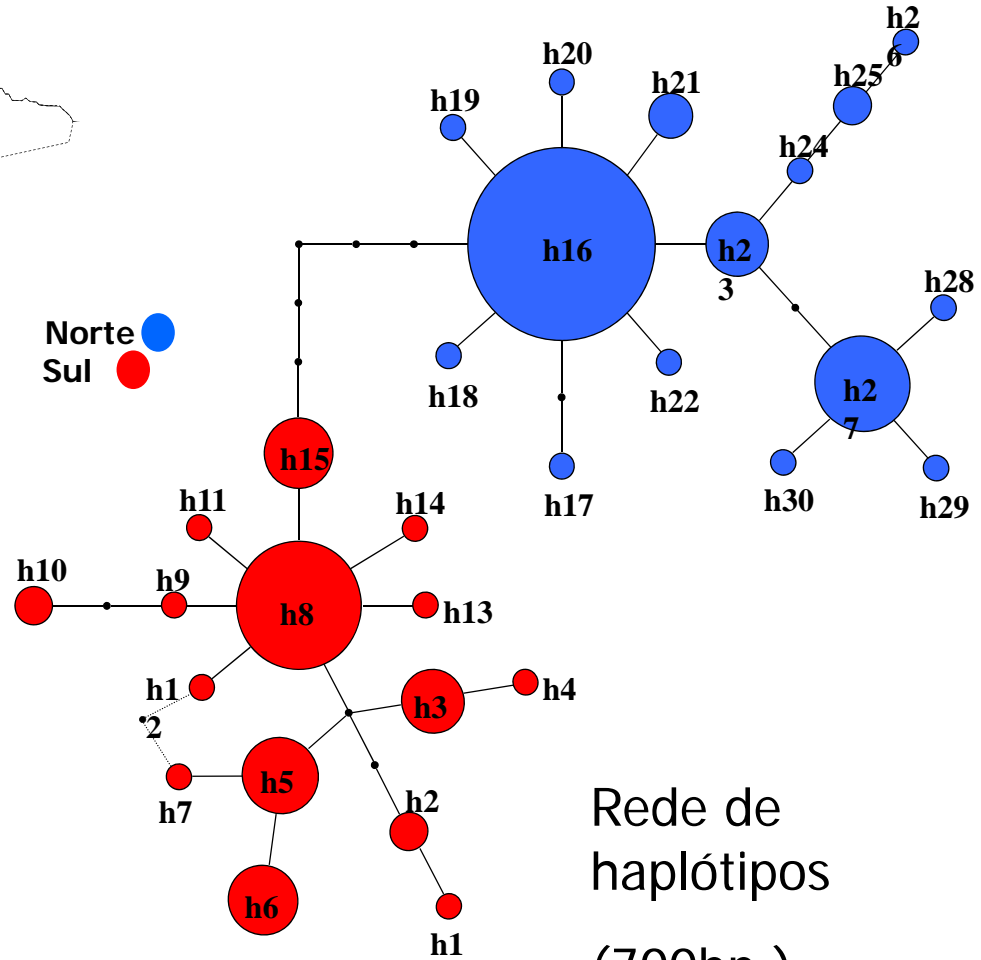
-TCS

-(Arlequin)...

(usam métodos diferentes)

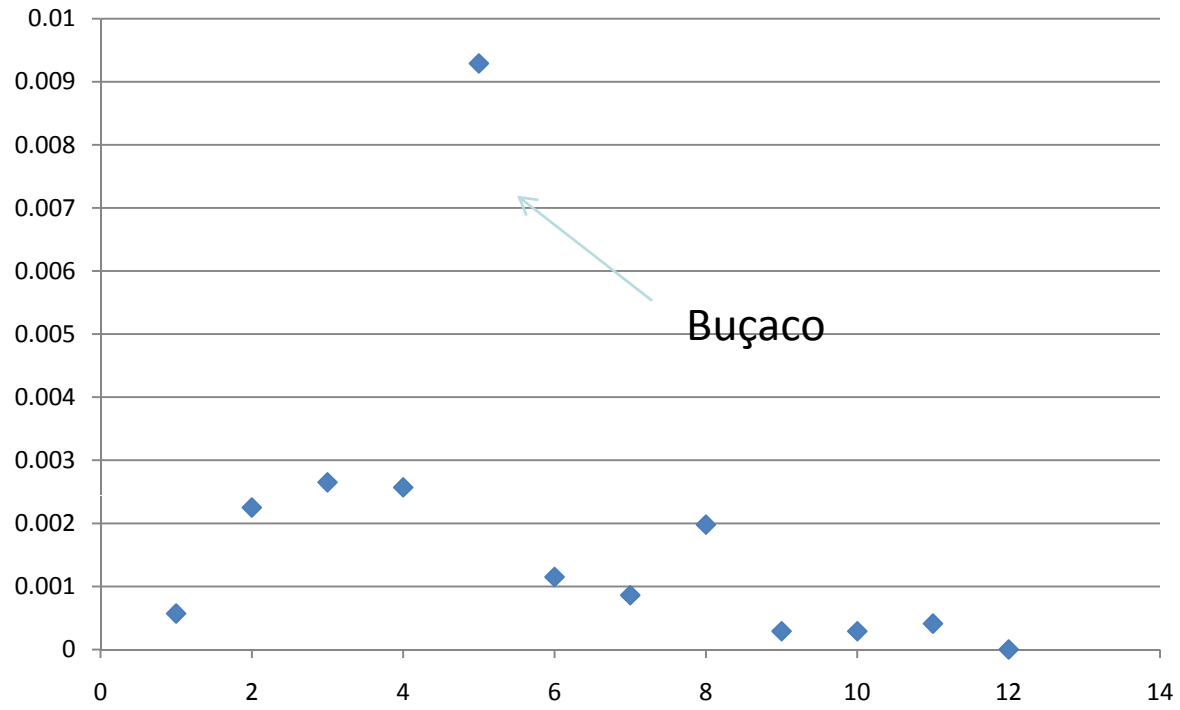


Distribuição geográfica dos grupos



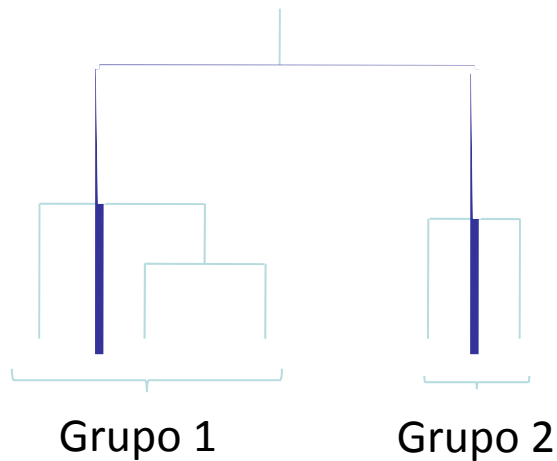
Rede de haplótipos (700bp)

Pi



Subestruturação ao nível de sequências do DNA

Distância entre populações



D_{xy}: distância média entre elementos dos dois grupos.

D_a: distância média entre elementos dos dois grupos, corrigida para a distância observada dentro de cada grupo.

Resultados

```
Output: DNA Div. Between Populations
DNA Divergence Between Populations

Input Data File: E:\... \chiosubdiv.nex
Selected region: 1-700   Number of sites: 700
Total sites [excluding alignment gaps]: 700

Population 1: norte
Number of sequences: 76
Number of polymorphic sites: 15
Total number of mutations: 15
Average number of nucleotide differences, k: 1,424
Nucleotide diversity, Pi(1): 0,00203

Population 2: sul
Number of sequences: 48
Number of polymorphic sites: 16
Total number of mutations: 16
Average number of nucleotide differences, k: 2,381
Nucleotide diversity, Pi(2): 0,00340

Total data:
Number of sequences: 124
Number of polymorphic sites: 34
Total number of mutations: 34
Average number of nucleotide differences, k: 5,130
Nucleotide diversity, Pi(t): 0,00733

Between populations:
Number of fixed differences: 6
Mutations polymorphic in population 1, but monomorphic in population 2: 12
Mutations polymorphic in population 2, but monomorphic in population 1: 13
Shared Mutations: 3

Average number of nucleotide differences between populations: 8,875
Average number of nuc. subs. per site between populations, Dxy: 0,01268
Number of net nuc. subs. per site between populations, Da: 0,00996
```

$D_{xy} = 1,268\%$

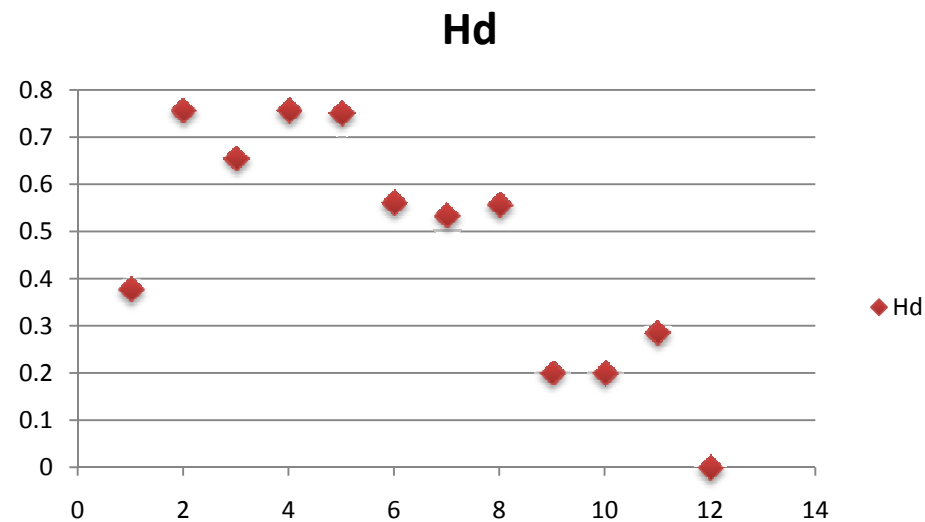
$D_a = 0,996\%$

Nº de posições fixas = 6

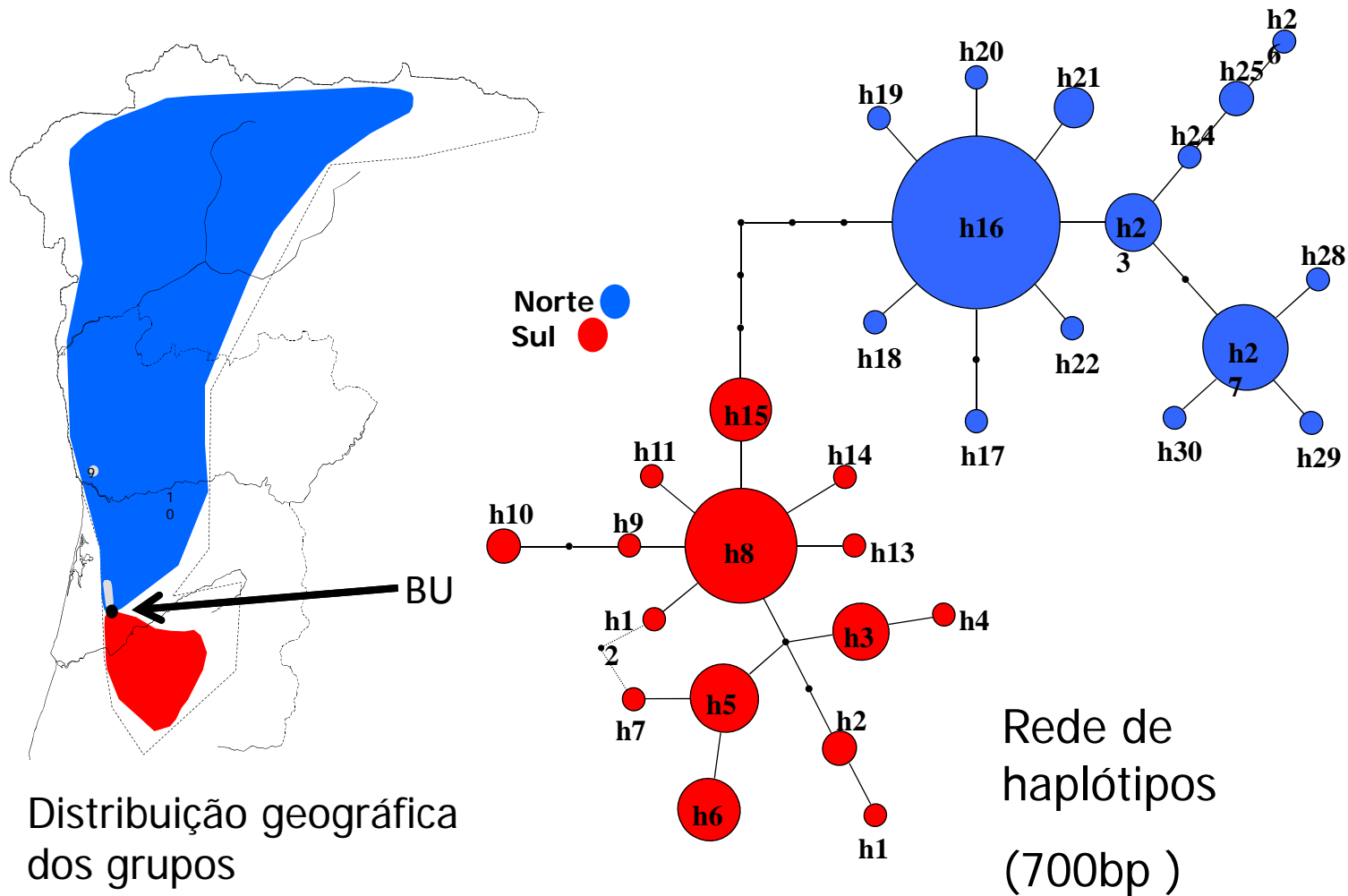
Etc. etc.

Recapitulando...

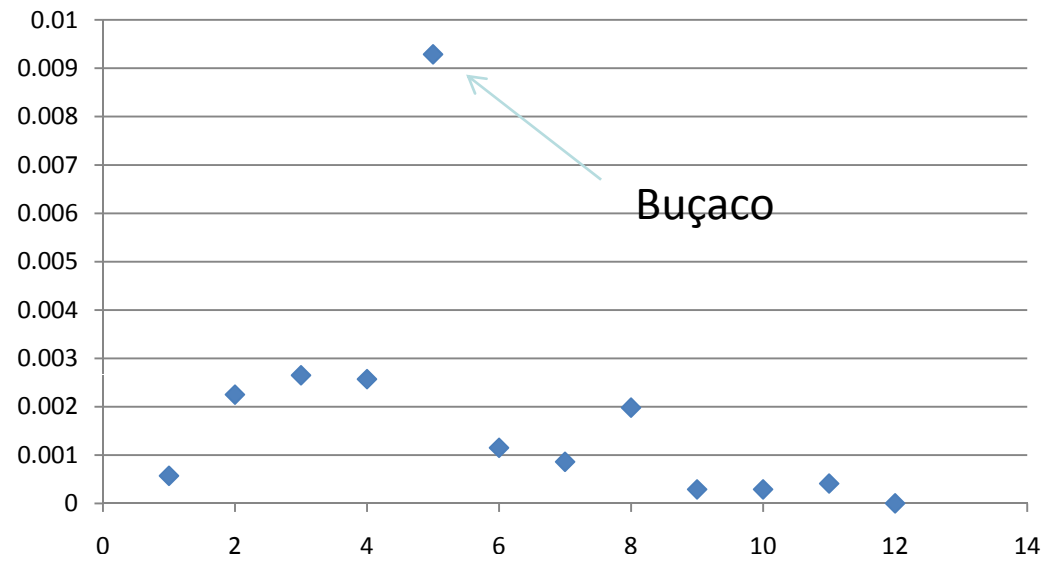
- Calculámos diferentes índices de diversidade para populações de *C. lusitanica*
- Verificámos que, tal como nos marcadores nucleares, a diversidade diminui nas populações mais a norte.



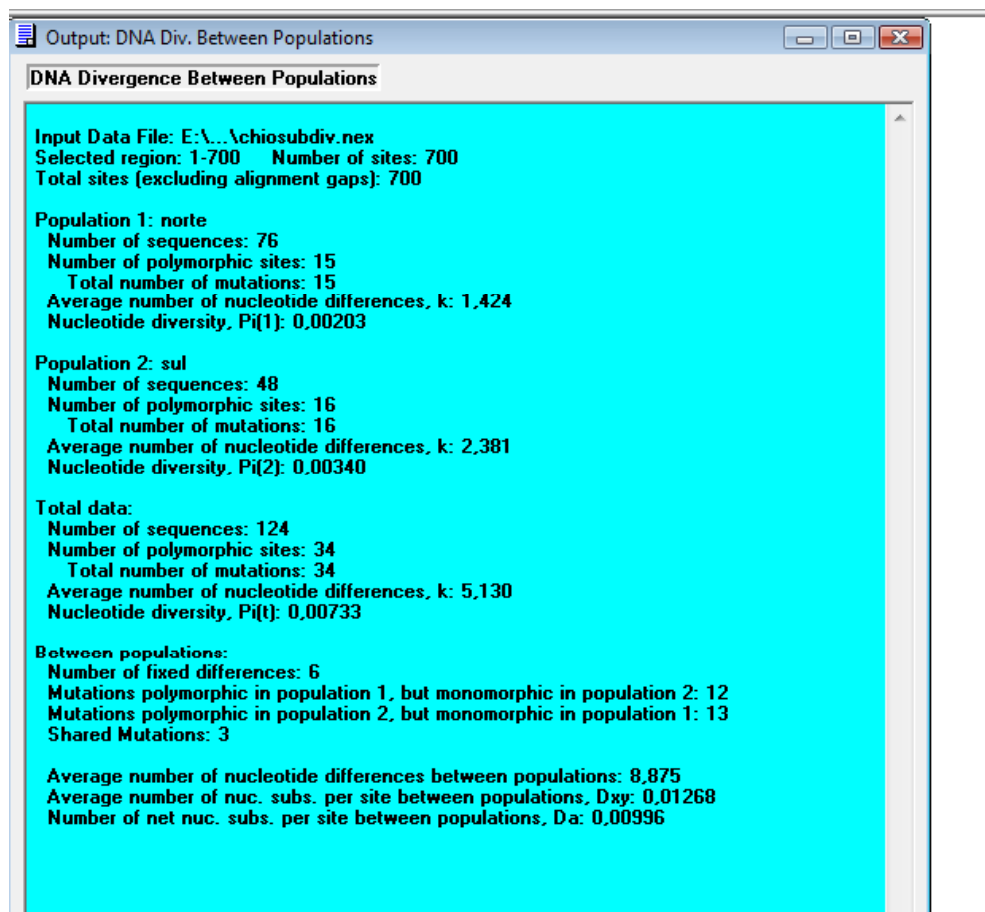
- Verificámos que existem dois grandes grupos de haplótipos que ocupam áreas geográficas diferentes.



Pi



- Iniciámos ainda a análise de subestruturação calculando a distância genética entre os grupos.



```
Output: DNA Div. Between Populations
DNA Divergence Between Populations
Input Data File: E:\...Achiosubdiv.nex
Selected region: 1-700  Number of sites: 700
Total sites (excluding alignment gaps): 700

Population 1: norte
Number of sequences: 76
Number of polymorphic sites: 15
Total number of mutations: 15
Average number of nucleotide differences, k: 1,424
Nucleotide diversity, Pi(1): 0,00203

Population 2: sul
Number of sequences: 48
Number of polymorphic sites: 16
Total number of mutations: 16
Average number of nucleotide differences, k: 2,381
Nucleotide diversity, Pi(2): 0,00340

Total data:
Number of sequences: 124
Number of polymorphic sites: 34
Total number of mutations: 34
Average number of nucleotide differences, k: 5,130
Nucleotide diversity, Pi(t): 0,00733

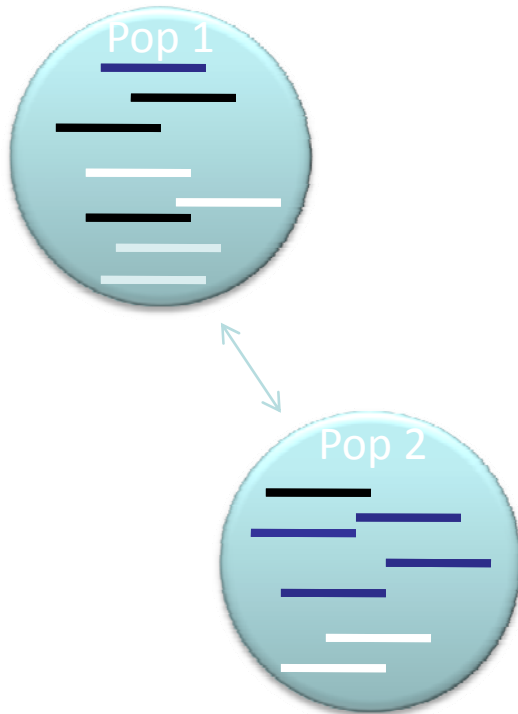
Between populations:
Number of fixed differences: 6
Mutations polymorphic in population 1, but monomorphic in population 2: 12
Mutations polymorphic in population 2, but monomorphic in population 1: 13
Shared Mutations: 3

Average number of nucleotide differences between populations: 8,875
Average number of nuc. subs. per site between populations, Dxy: 0,01268
Number of net nuc. subs. per site between populations, Da: 0,00996
```

Subestruturação ao nível de sequências do DNA

Diferenciação entre populações

As estatísticas F (**Fst**, **Fit**, **Fis**) foram desenvolvidas a pensar em dados alélicos (como proteínas), mas podemos fazer um paralelo com sequências de DNA.



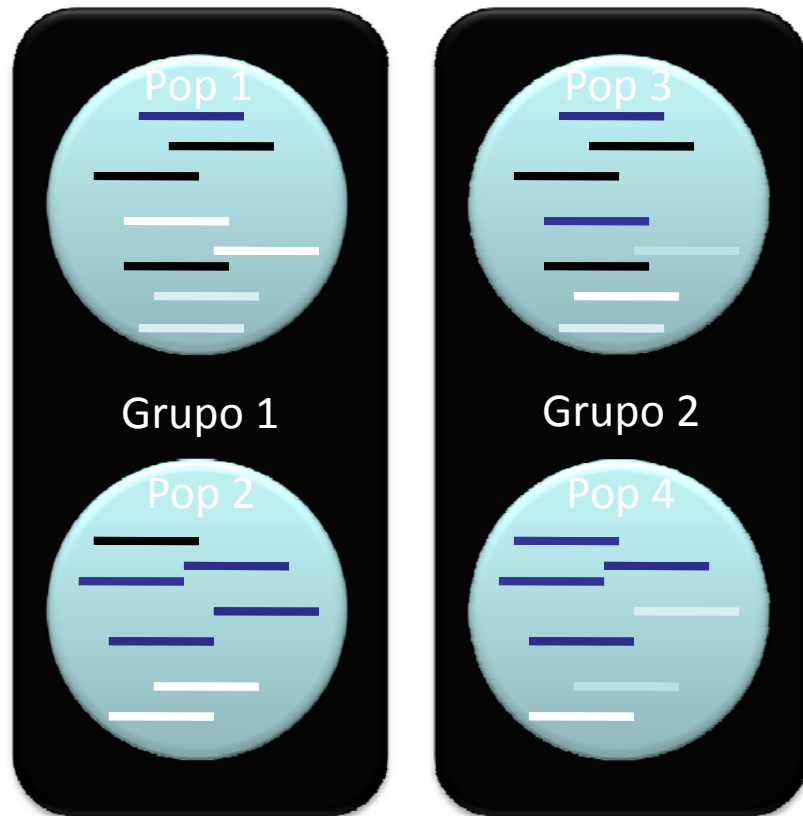
Existem diversas medidas análogas ao Fst para sequências de DNA. Algumas incorporam informação sobre o número de mutações (distância) entre alelos (ex. **Fst** de Hudson et al. (1992)), outras contemplam apenas frequências (como as medidas “clássicas”).

No fundo, o que estes valores nos medem é a proporção de variação que está entre as unidades que consideramos (ex. populações) por oposição à proporção que está dentro dessas unidades.

O DNAsp e o Arlequin calculam-nos estes valores de Fst (mas para isto temos primeiro de definir populações).

Diferenciação entre populações

E se quisermos pensar a mais do que um nível?



Qual a % de variação explicada por diferenças:

entre grupos?

entre populações dentro de cada grupo?

dentro das populações?

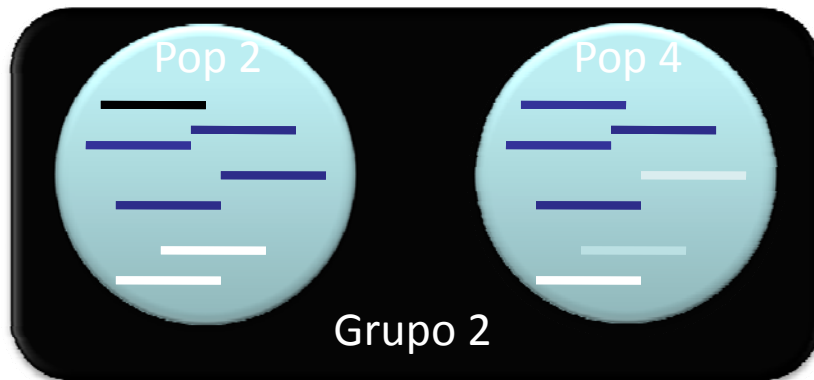
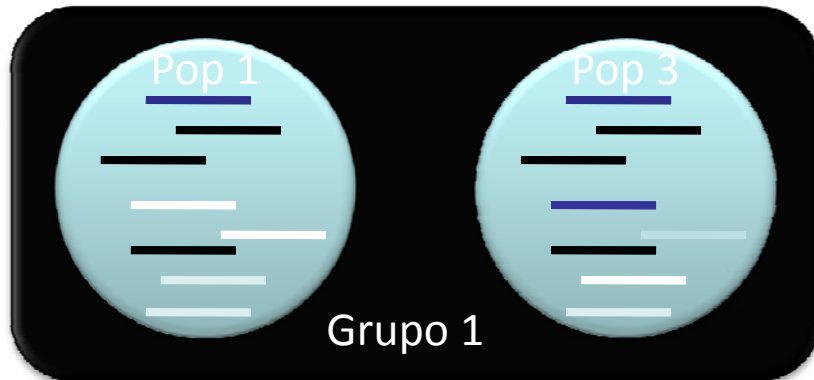
AMOVA – analysis of molecular variance
(Excoffier et al. 1992)

Pode ser feita não só para sequências de DNA mas também para outros dados (proteínas, microssatélites, etc.).

Software Arlequin

Diferenciação entre populações

E se quisermos pensar a mais do que um nível?



Qual a % de variação explicada por diferenças:

entre grupos?

entre populações dentro de cada grupo?

dentro das populações?

AMOVA – analysis of molecular variance
(Excoffier et al. 1992)

Pode ser feita não só para sequências de DNA mas também para outros dados (proteínas, microssatélites, etc.).

Software Arlequin

Diferenciação entre populações

AMOVA design and results :

Reference: Weir, B.S. and Cockerham, C.C. 1984.
Excoffier, L., Smouse, P., and Quattro, J. 1992.
Weir, B. S., 1996.

Source of variation	d.f.	Sum of squares	Variance components	Percentage of variation
Among groups	7	558.019	9.18219 Va	84.23
Among populations within groups	10	25.025	0.44491 Vb	4.08
Within populations	51	64.953	1.27359 Vc	11.68
Total	68	647.997	10.90068	

Fixation Indices
FSC : 0.25890
FST : 0.88316
FCT : 0.84235

Significance tests (1023 permutations)

Vc and FST : P(rand. value < obs. value) = 0.00000
P(rand. value = obs. value) = 0.00000
P(rand. value <= obs. value) = 0.00000+-0.00000

Vb and FSC : P(rand. value > obs. value) = 0.00880
P(rand. value = obs. value) = 0.00000
P(rand. value >= obs. value) = 0.00880+-0.00251

Va and FCT : P(rand. value > obs. value) = 0.00000
P(rand. value = obs. value) = 0.00000
P(rand. value >= obs. value) = 0.00000+-0.00000

Output do Arlequin:

- % de variação em cada nível hierárquico

- índices de fixação

- níveis de significância determinados por permutações

Resultados

AMOVA design and results :

Reference: Weir, B.S. and Cockerham, C.C. 1984.
Excoffier, L., Smouse, P., and Quattro, J. 1992.
Weir, B. S., 1996.

AMOVA norte vs. sul
(ignorando a população do Buçaco):

- % variação entre grupos: 77,91%
- % variação entre pops dentro dos grupos: 12,07%
- % variação dentro das pops: 10,02%

(todos os valores estimados para os índices de fixação são significativos)

Source of variation	d.f.	Sum of squares	Variance components	Percentage of variation
Among groups	1	185.455	3.34546 Va	77.91
Among populations within groups	14	56.013	0.51843 Vb	12.07
Within populations	99	42.602	0.43032 Vc	10.02
Total	114	284.070	4.29421	

Fixation Indices

FSC :	0.54643
FST :	0.89979
FCT :	0.77906

Significance tests (1023 permutations)

Vc and FST : P(rand. value < obs. value) = 0.00000
P(rand. value = obs. value) = 0.00000
P-value = 0.00000+-0.00000

Vb and FSC : P(rand. value > obs. value) = 0.00000
P(rand. value = obs. value) = 0.00000
P-value = 0.00000+-0.00000

Va and FCT : P(rand. value > obs. value) = 0.00000
P(rand. value = obs. value) = 0.00000
P-value = 0.00000+-0.00000