

Phylogenetics and Molecular Evolution/Filogenética e Evolução Molecular

Octávio S. Paulo

Computational Biology and Population Genomics Group (CoBiG2)

Análise intraespecífica de sequências

Sumário:

Análise de sequências ao nível intraespecífico, genealogias e coalescência. Redes intraespecíficas. A análise da história demográfica com base em sequências intraespecíficas.



**Ciências
ULisboa**

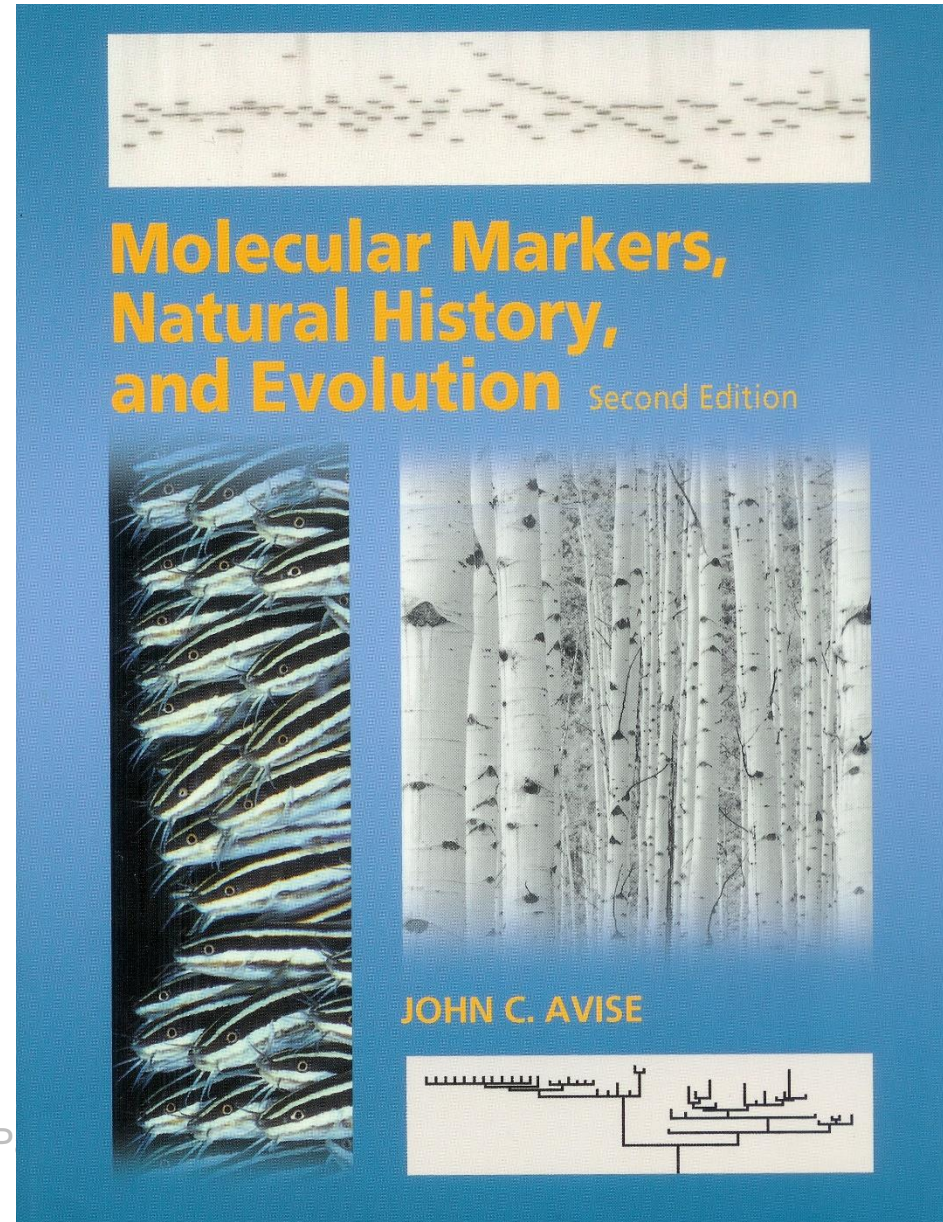
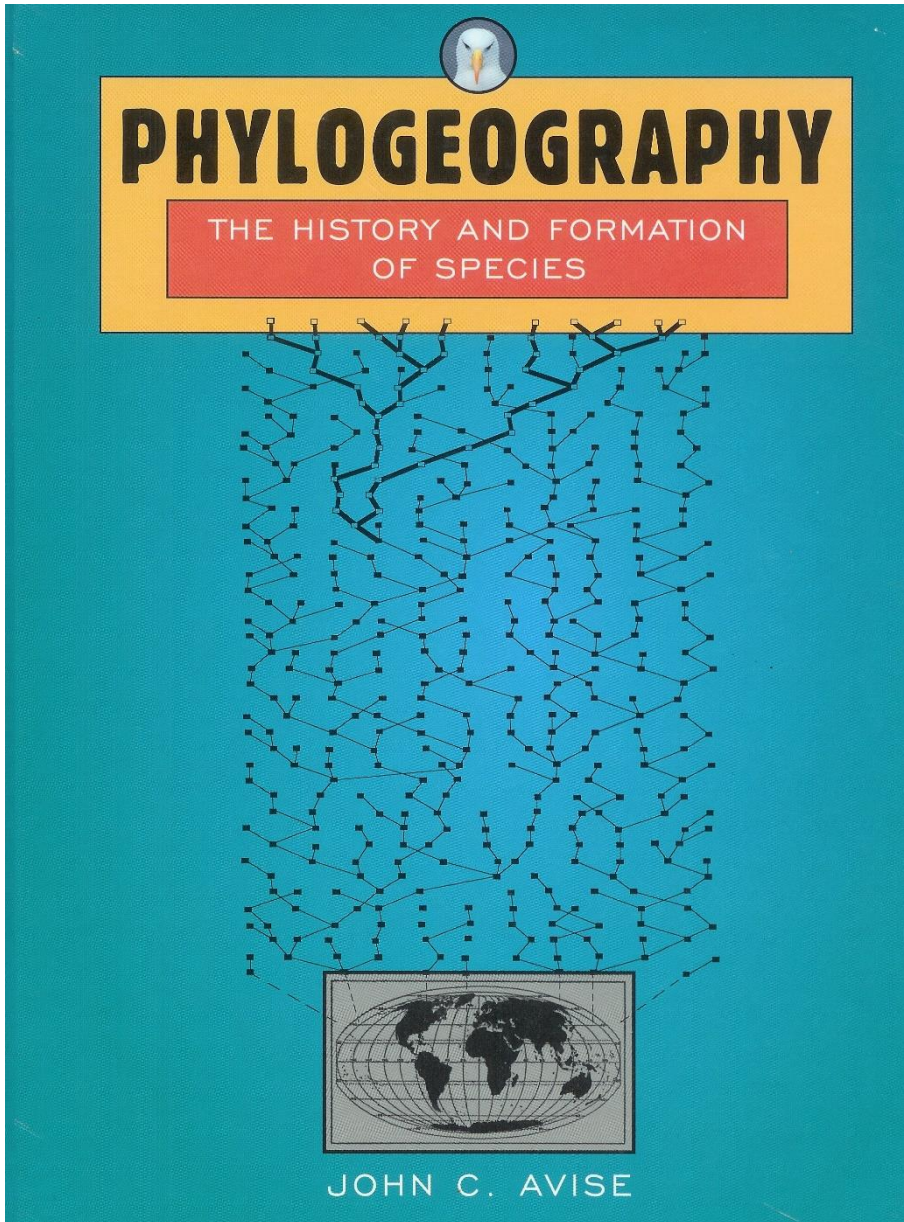
Faculdade
de Ciências
da Universidade
de Lisboa



**Computational
Biology & Population
Genomics Group**

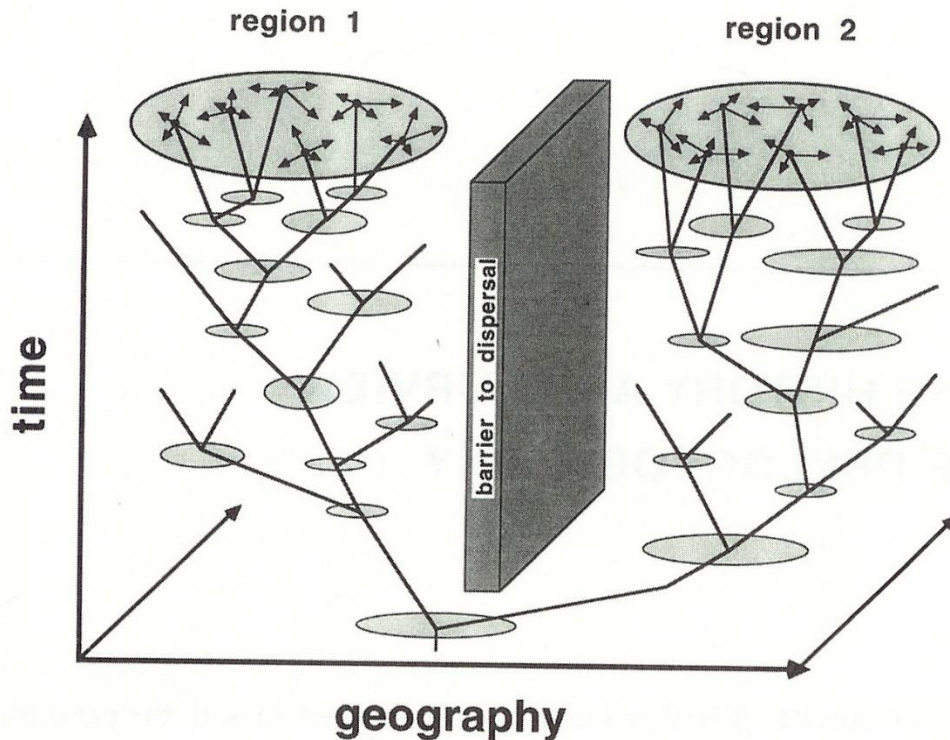


Phylogeography/Filogeografia

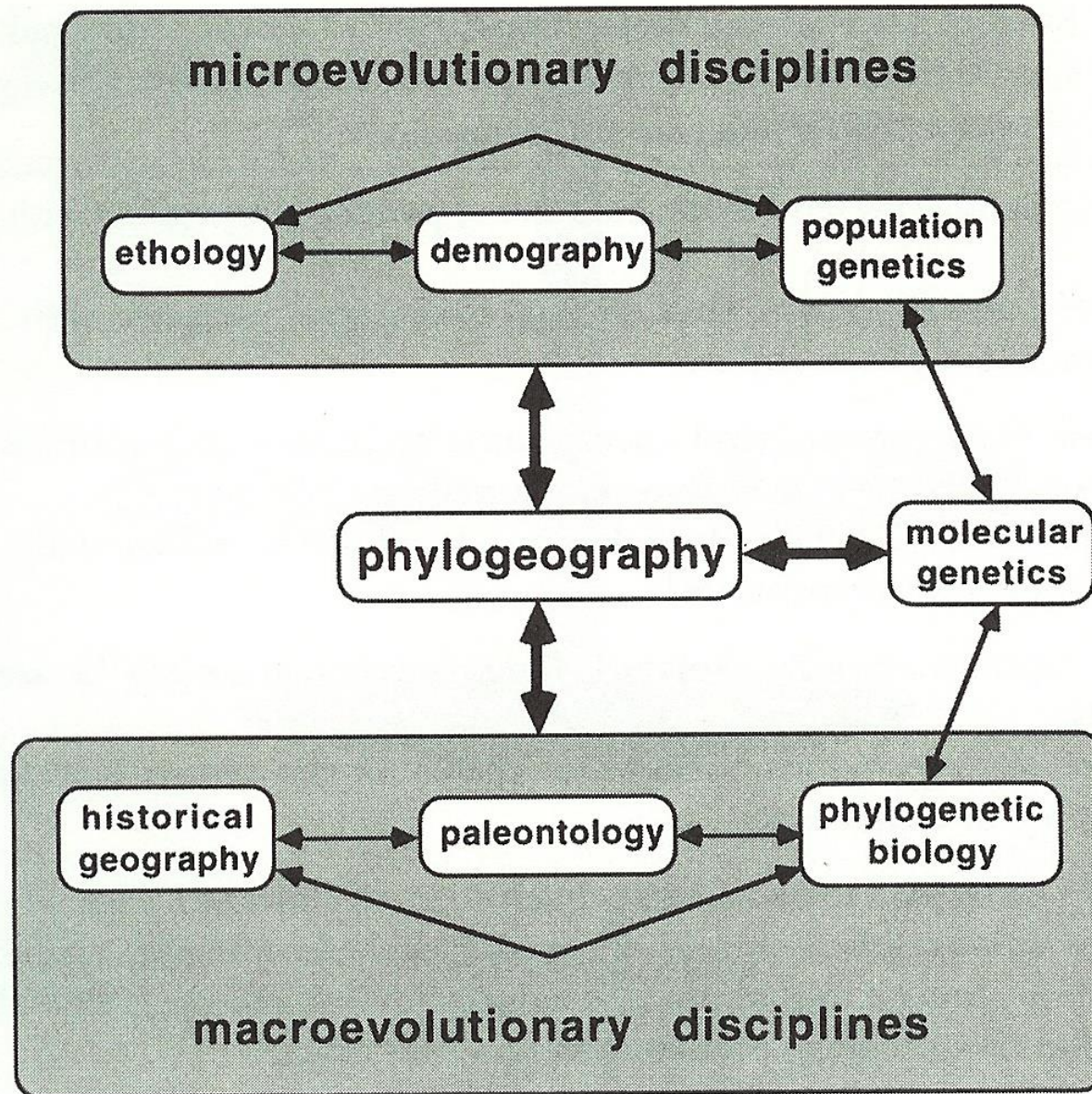


Phylogeography

Principles and processes governing the geographic distribution of genealogical lineages, especially those within and among closely related species



Genes e Geografia



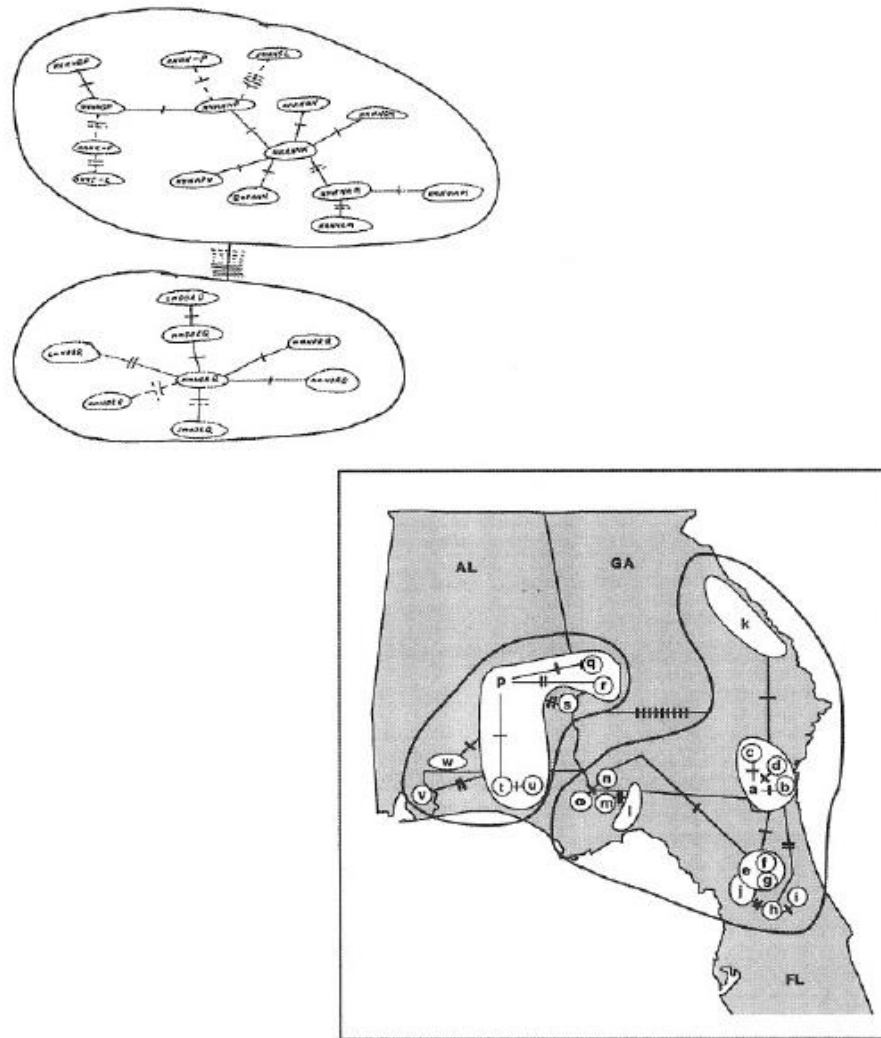
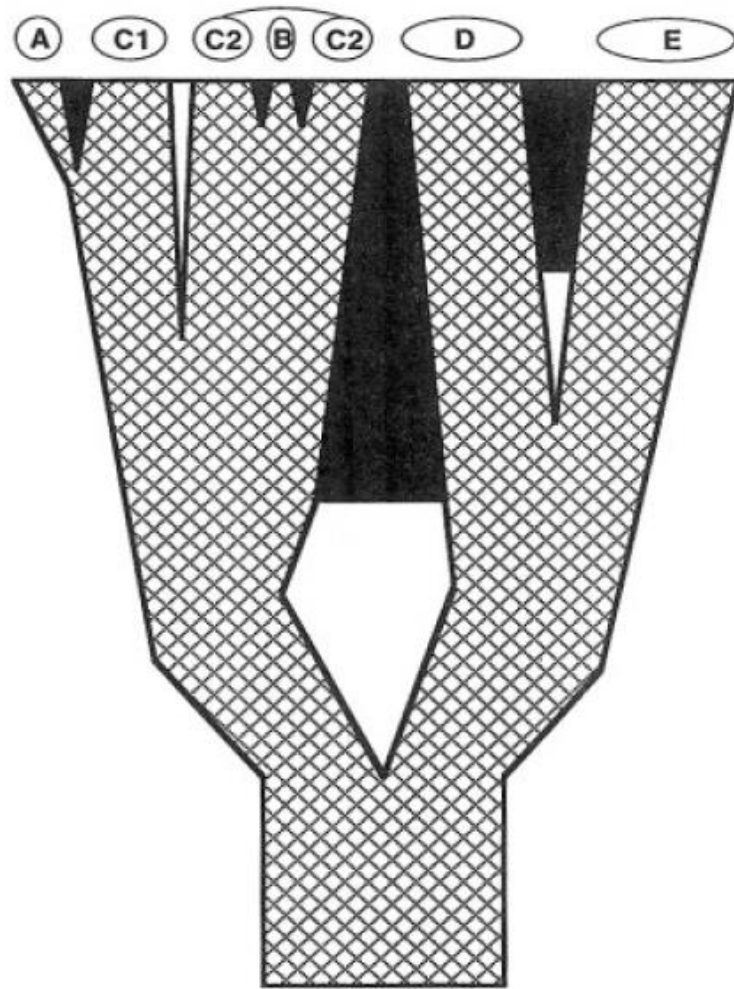
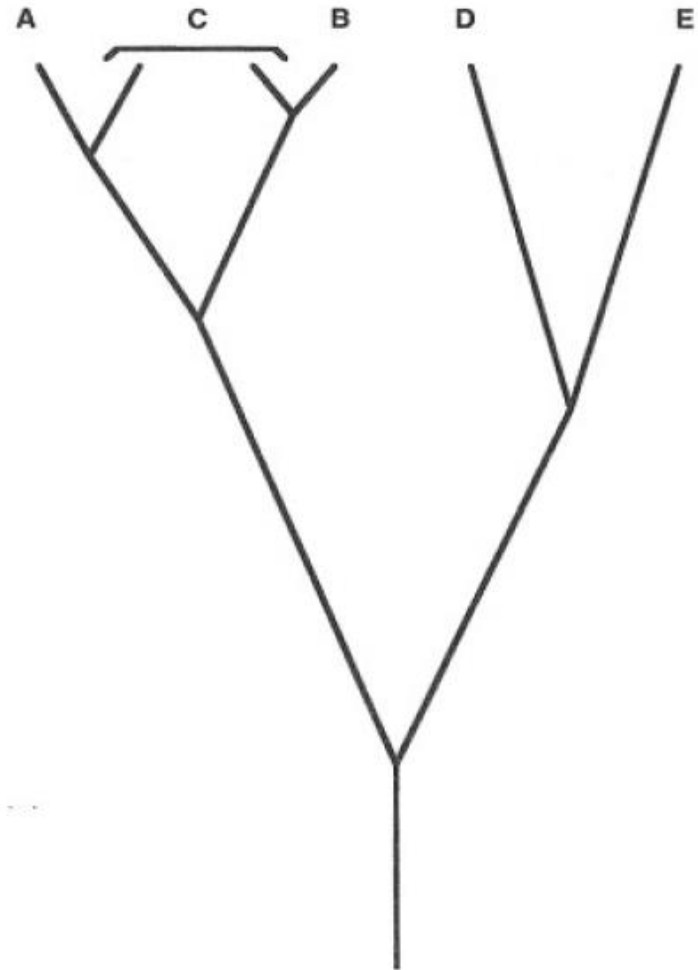


FIGURE 1.9 *Above:* Original hand-drawn phylogenetic network connecting 23 different mtDNA haplotypes in pocket gophers, as scribbled onto a piece of scratch paper from the RFLP data that eventually was to appear in Avise et al. (1979b). Uppercase letters provide coded summaries of the digestion profiles produced by six restriction enzymes. Slashes across network branches reflect the minimum numbers of inferred mutational steps along a pathway. Heavier lines encompass two distinctive mtDNA clades that differed by at least nine mutations. *Below:* Same network, now superimposed over the geographic sources of the samples in Alabama, Georgia, and Florida. Letters a–w indicate the 23 mtDNA haplotypes, and circles or extended ellipses encompass the geographic distributions observed for each haplotype.

Trees and lineages

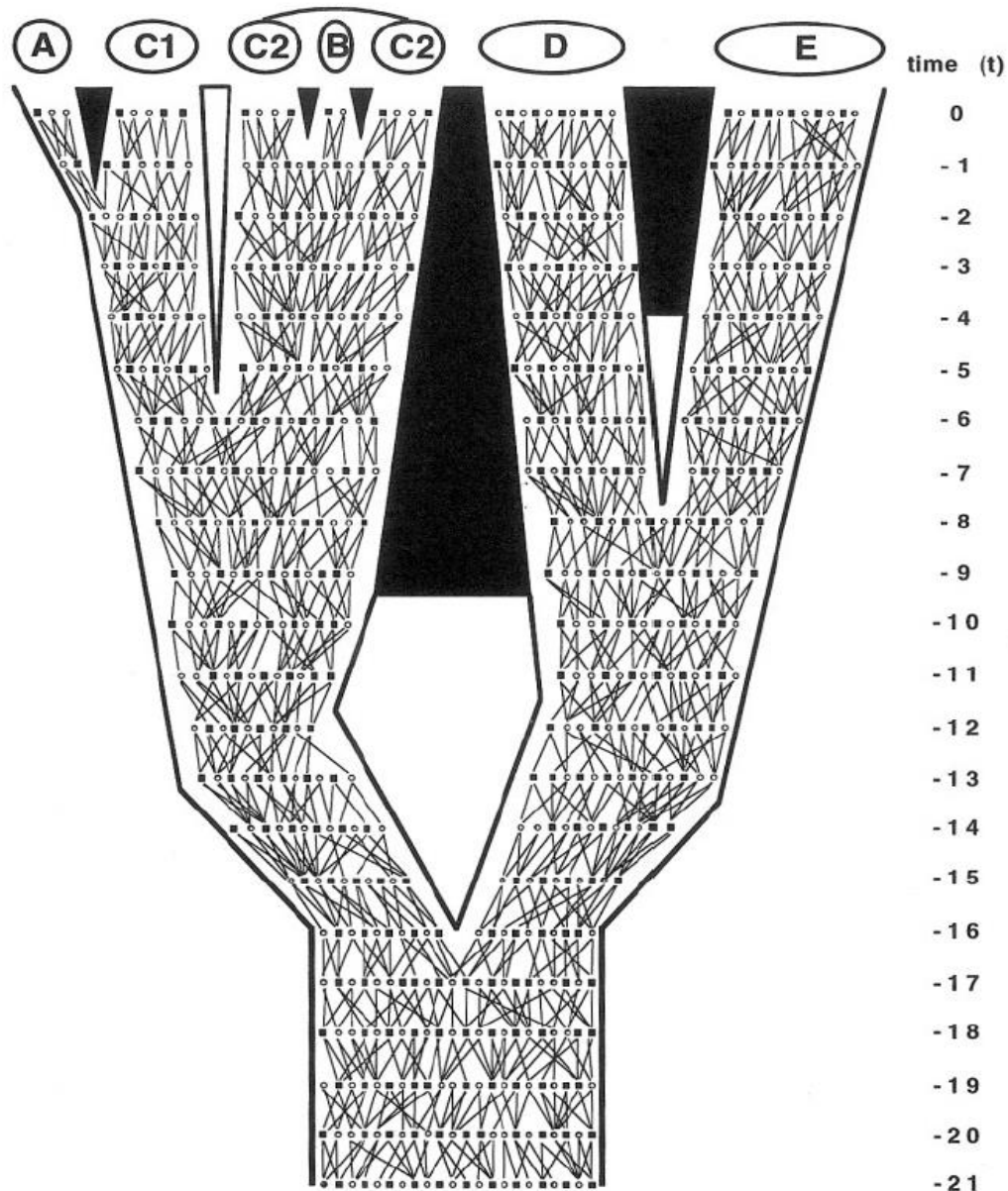


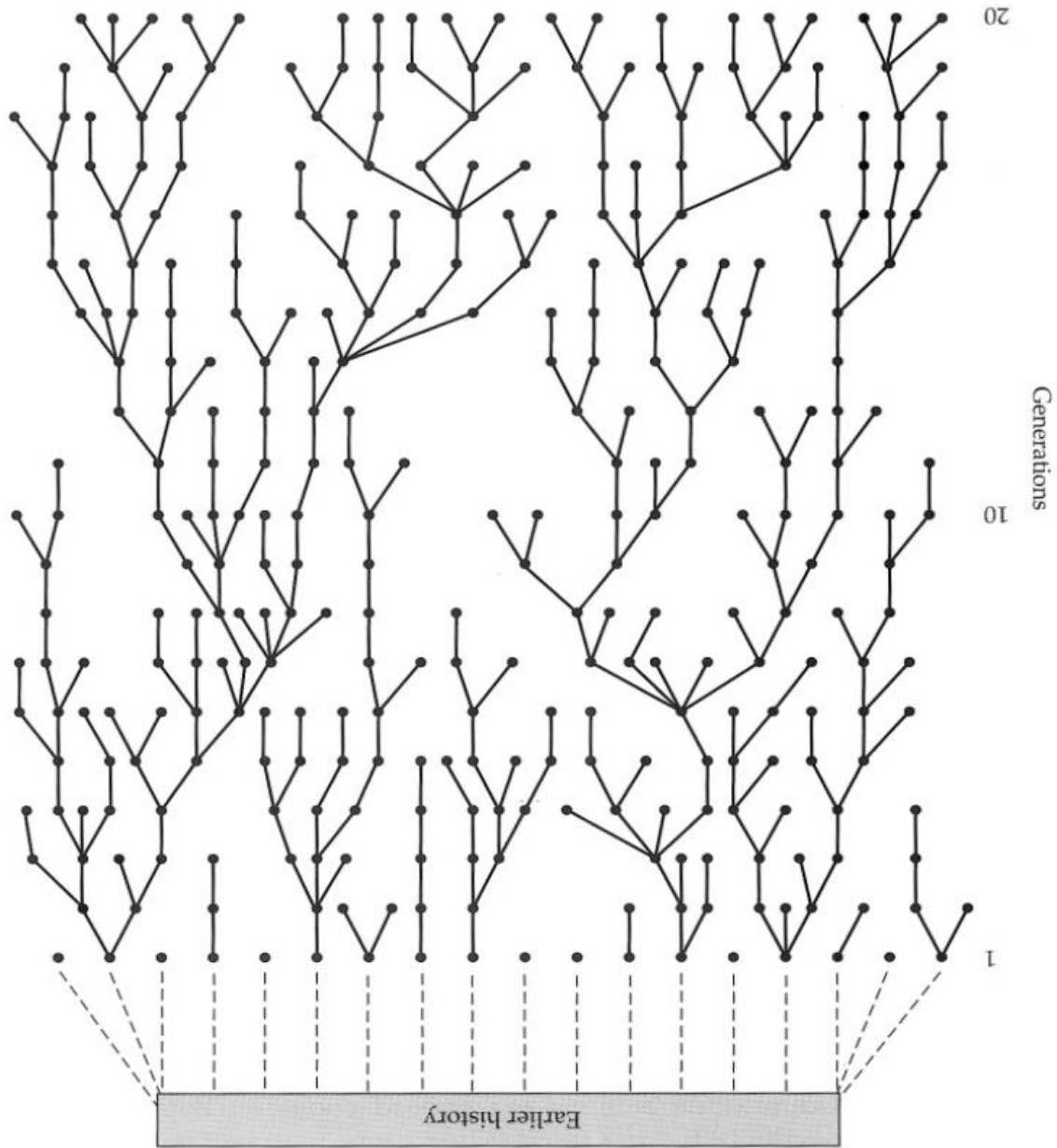
(a)



(b)

Lineages

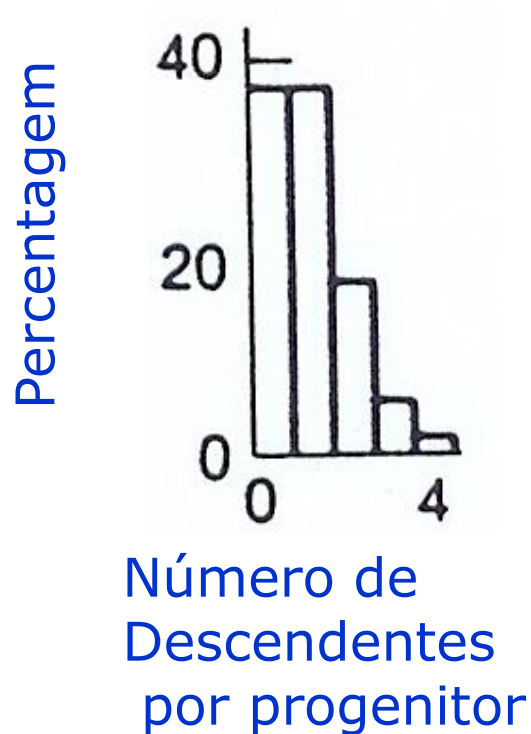




Distribuição de Poisson

Média = variância = 1

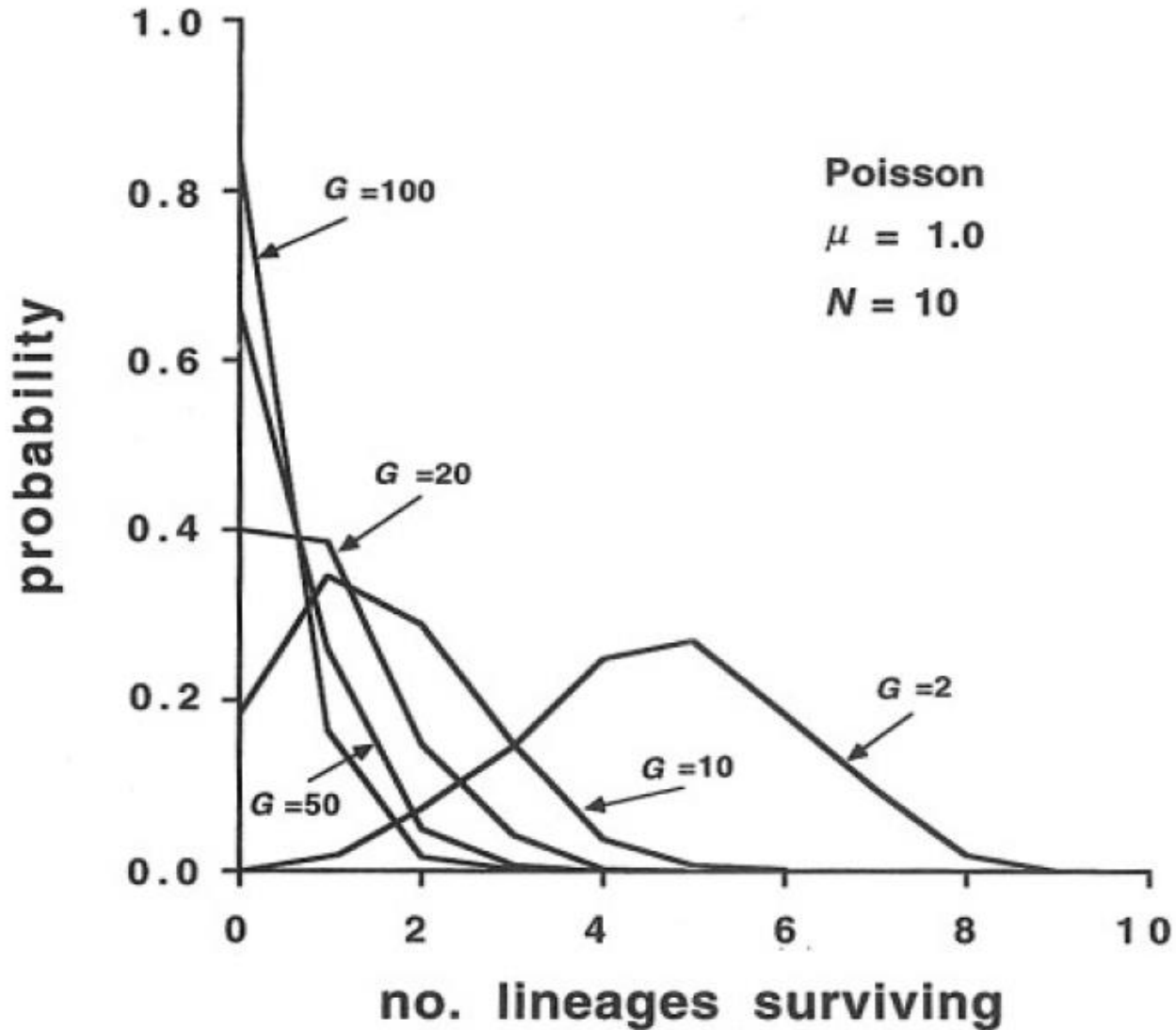
Probabilidade e Número
de descendentes por progenitor



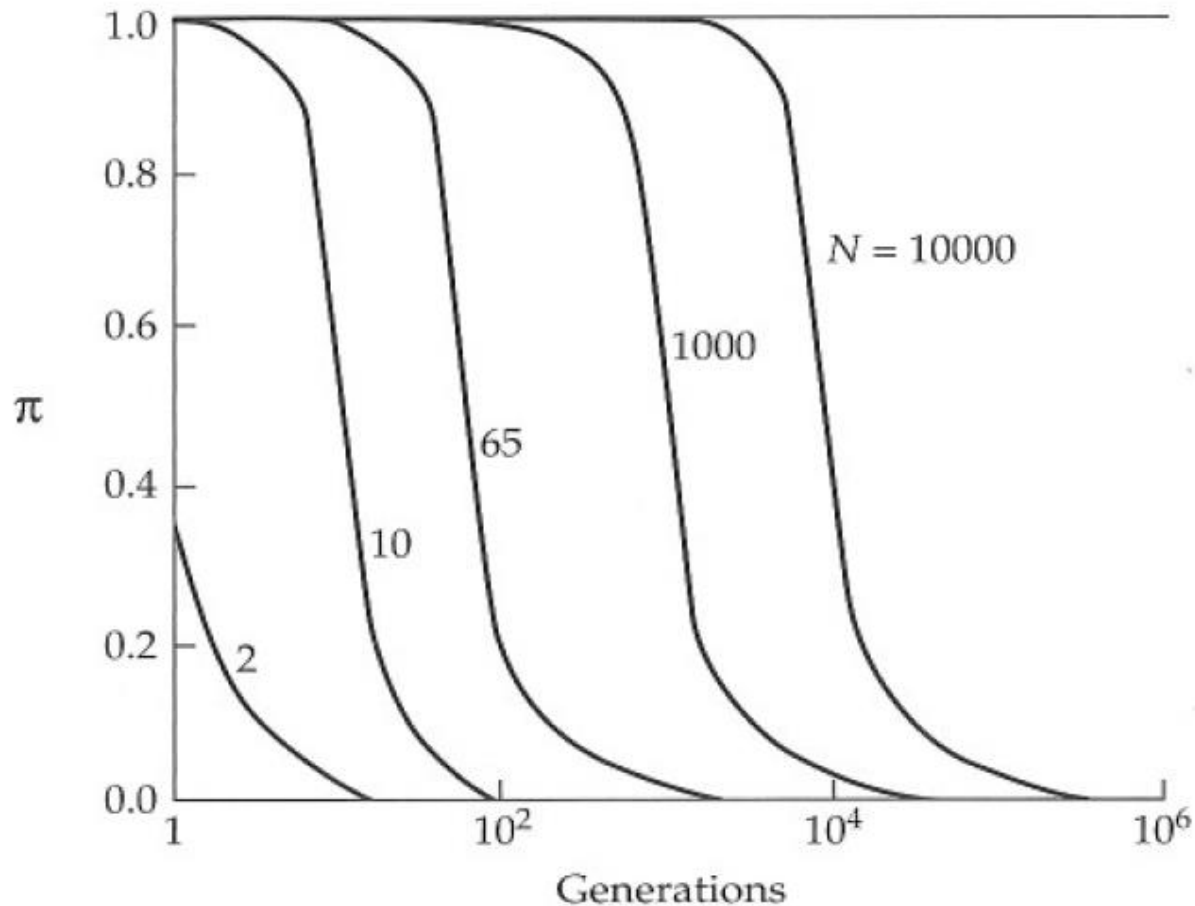
0.368 - 0
0.368 - 1
0.184 - 2
0.061 - 3
0.015 - 4
0.004 - 5

Probabilidade acumulada extinção
da linhagem ao fim de 100 gerações = 0.981

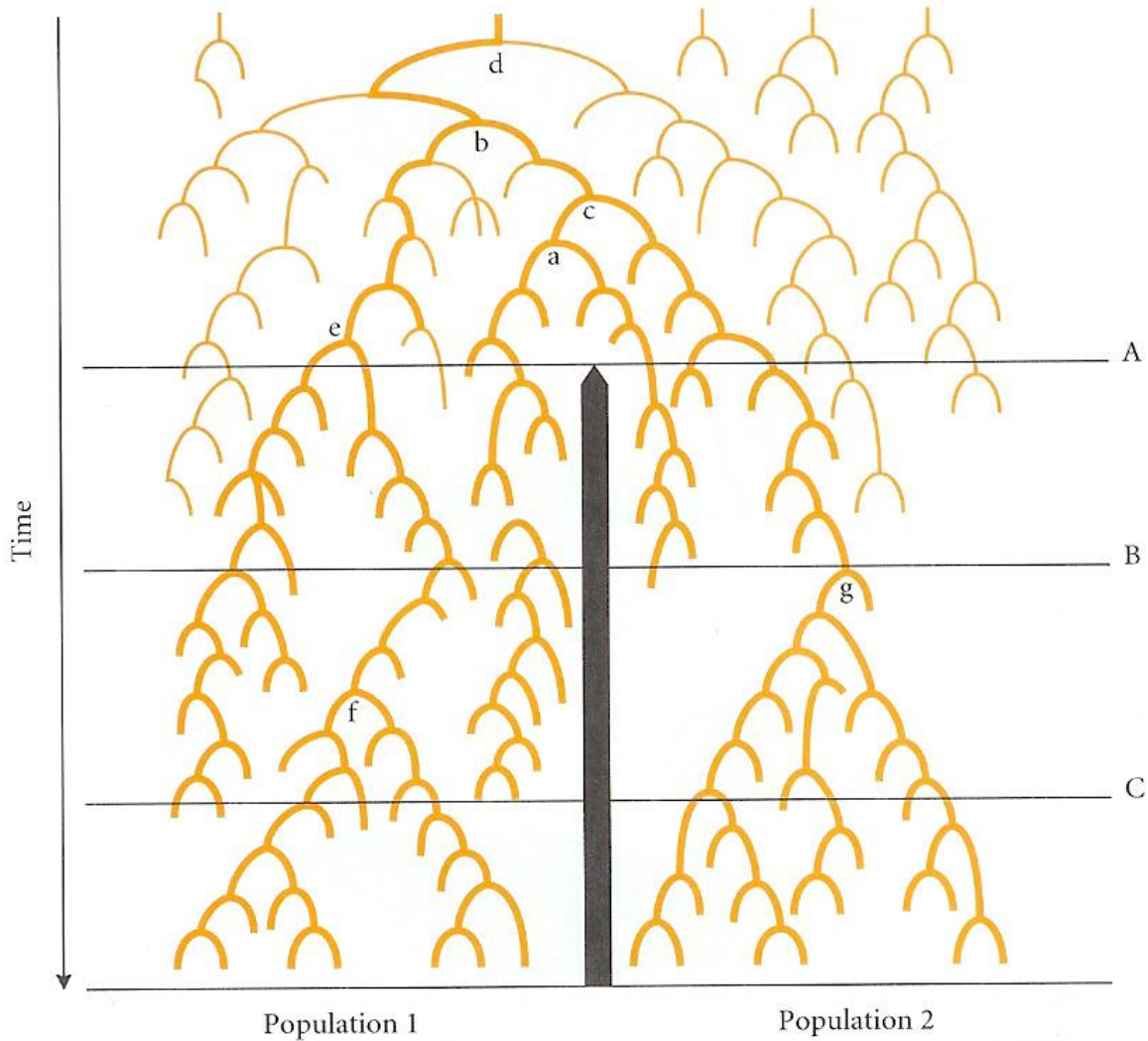
Frequency distribution of the probabilities of survival of founding lineages through G generations

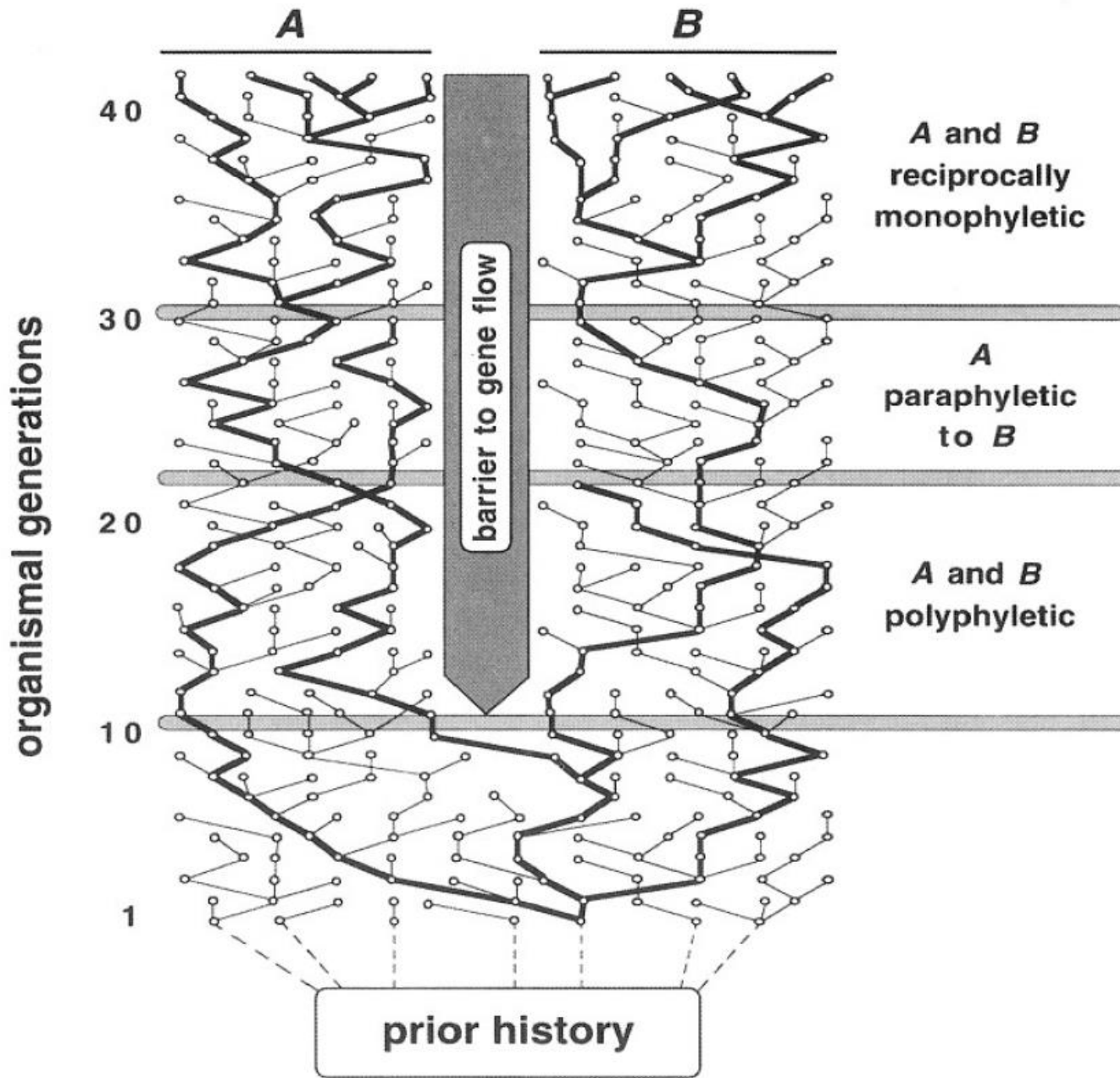


Probabilities of survival of two or more founding lineages

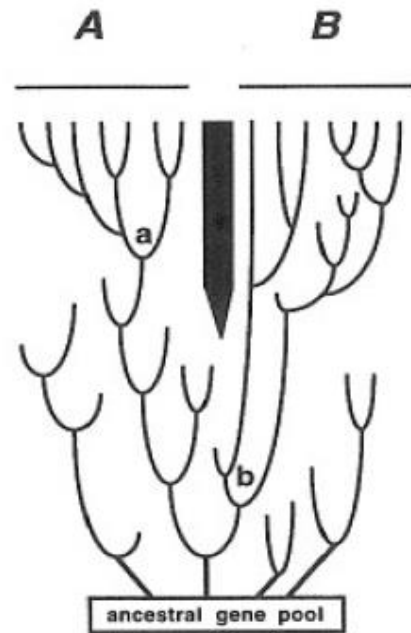


Lineage sorting

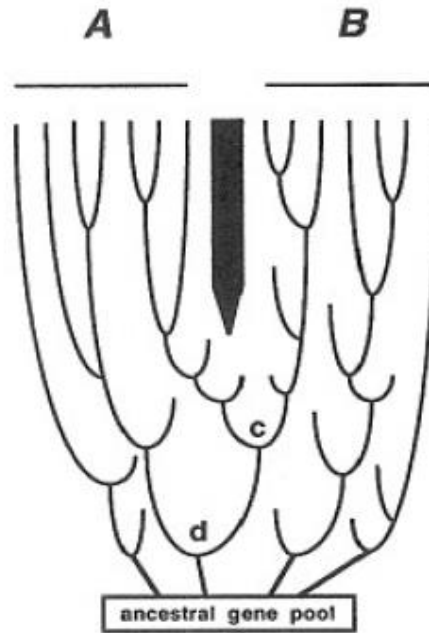




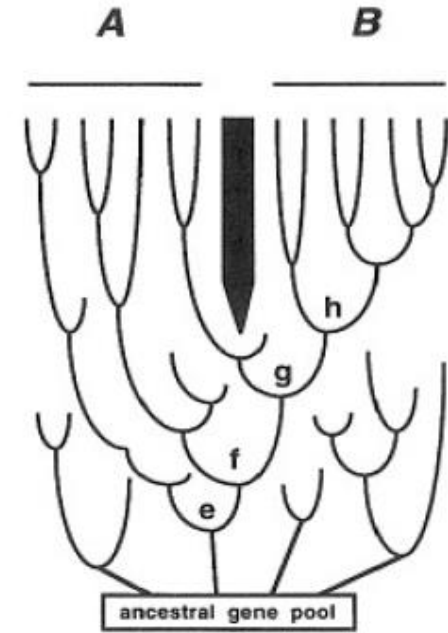
Lineage sorting



reciprocal monophyly



polyphyly

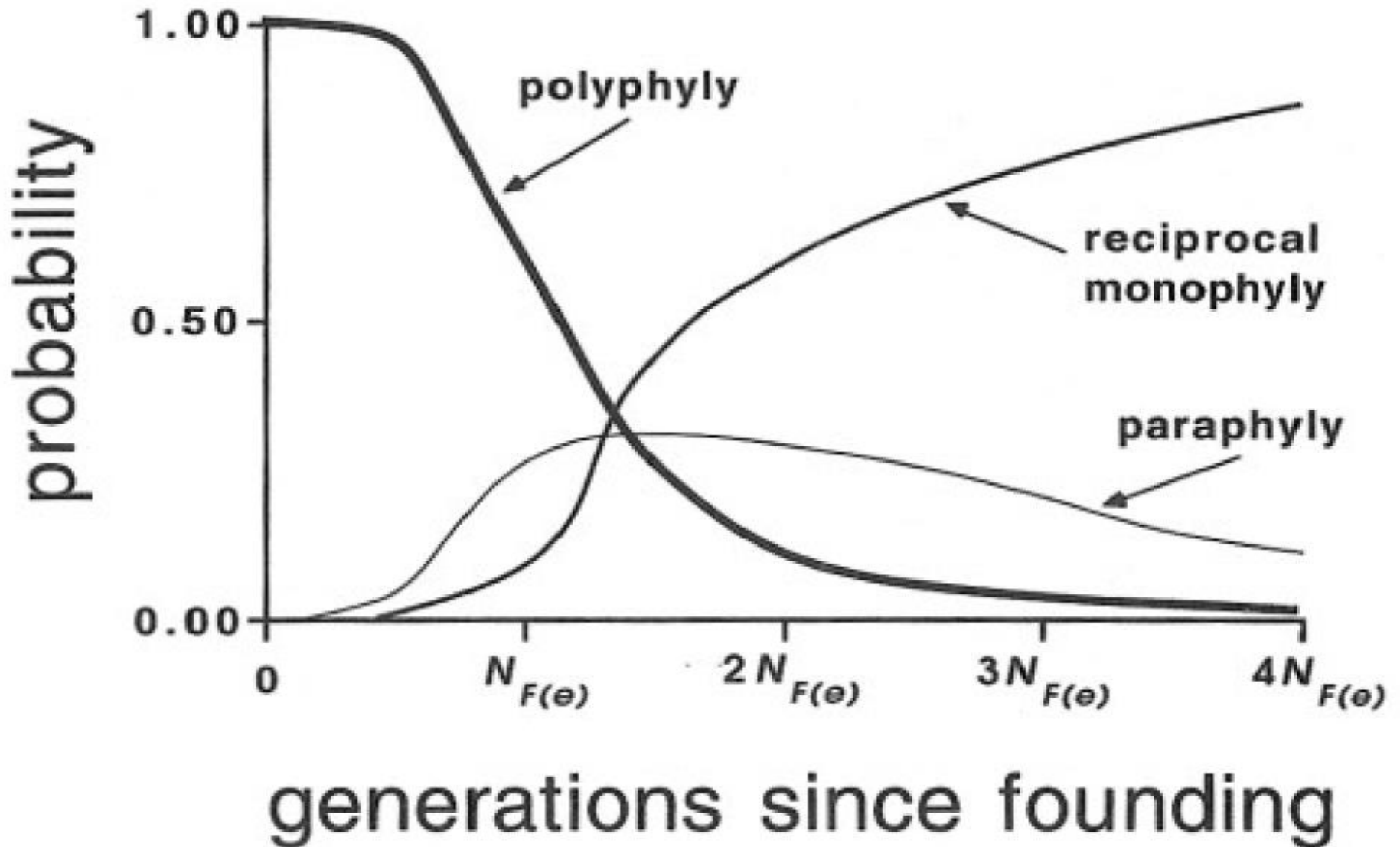


paraphyly

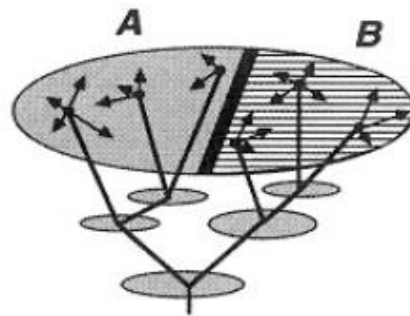
Lineage sorting

Phylogenetic Category	Phylogenetic Status	Distance Relationship
I	<i>A</i> and <i>B</i> monophyletic	$\max t_{AA} < \min t_{AB}$ and $\max t_{BB} < \min t_{AB}$
II	<i>A</i> and <i>B</i> polyphyletic	$\max t_{AA} > \min t_{AB}$ and $\max t_{BB} > \min t_{AB}$
IIIa	<i>A</i> paraphyletic with respect to <i>B</i>	$\max t_{AA} > \min t_{AB}$ and $\max t_{BB} < \min t_{AB}$
IIIb	<i>B</i> paraphyletic with respect to <i>A</i>	$\max t_{AA} < \min t_{AB}$ and $\max t_{BB} > \min t_{AB}$

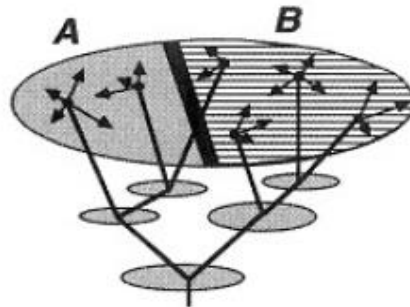
DNA Mitochondrial – Genes Nucleares



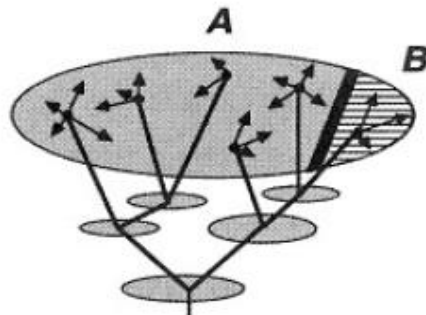
Lineage sorting



**reciprocal monophyly
for A and B**

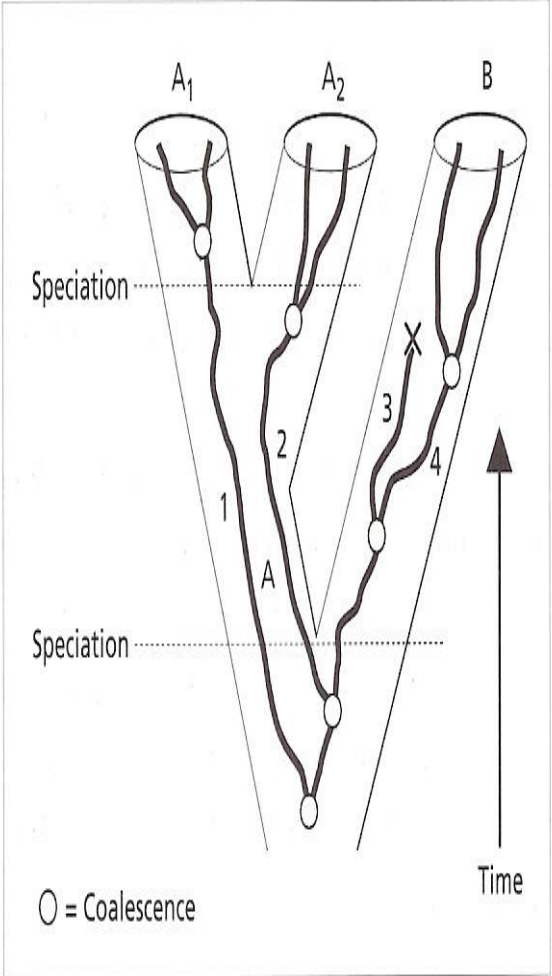
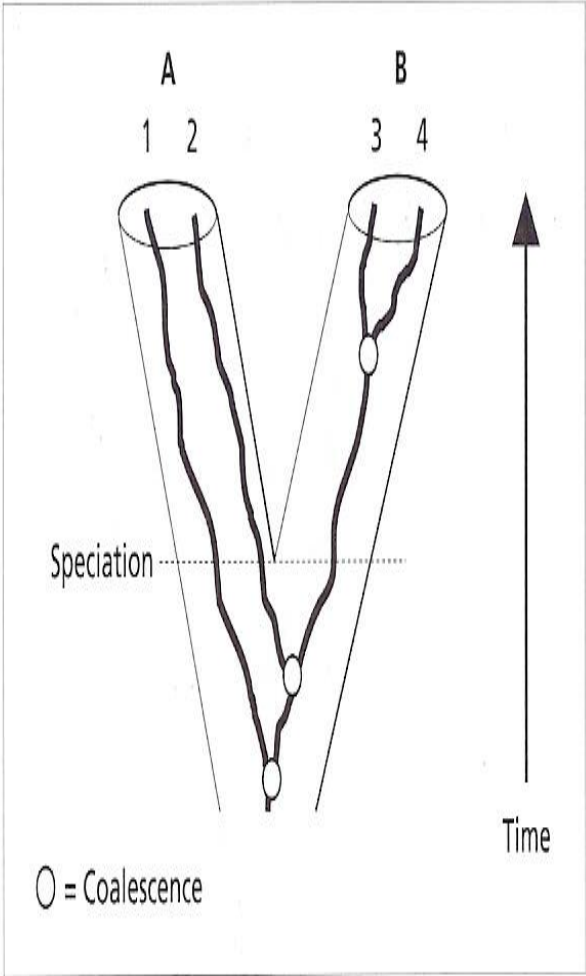
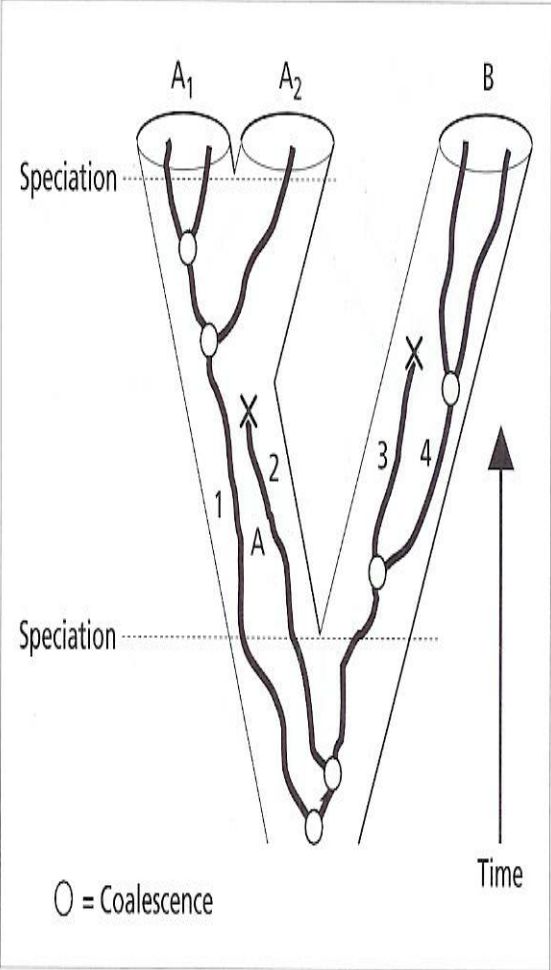


**polyphyly
for A and B**

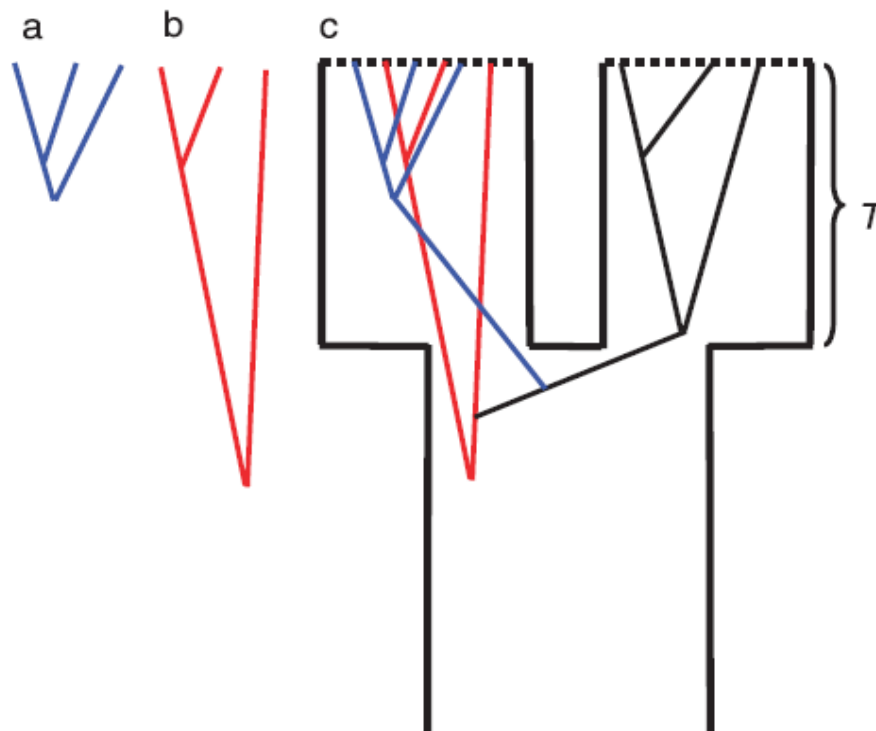


**A paraphyletic
with respect to B**

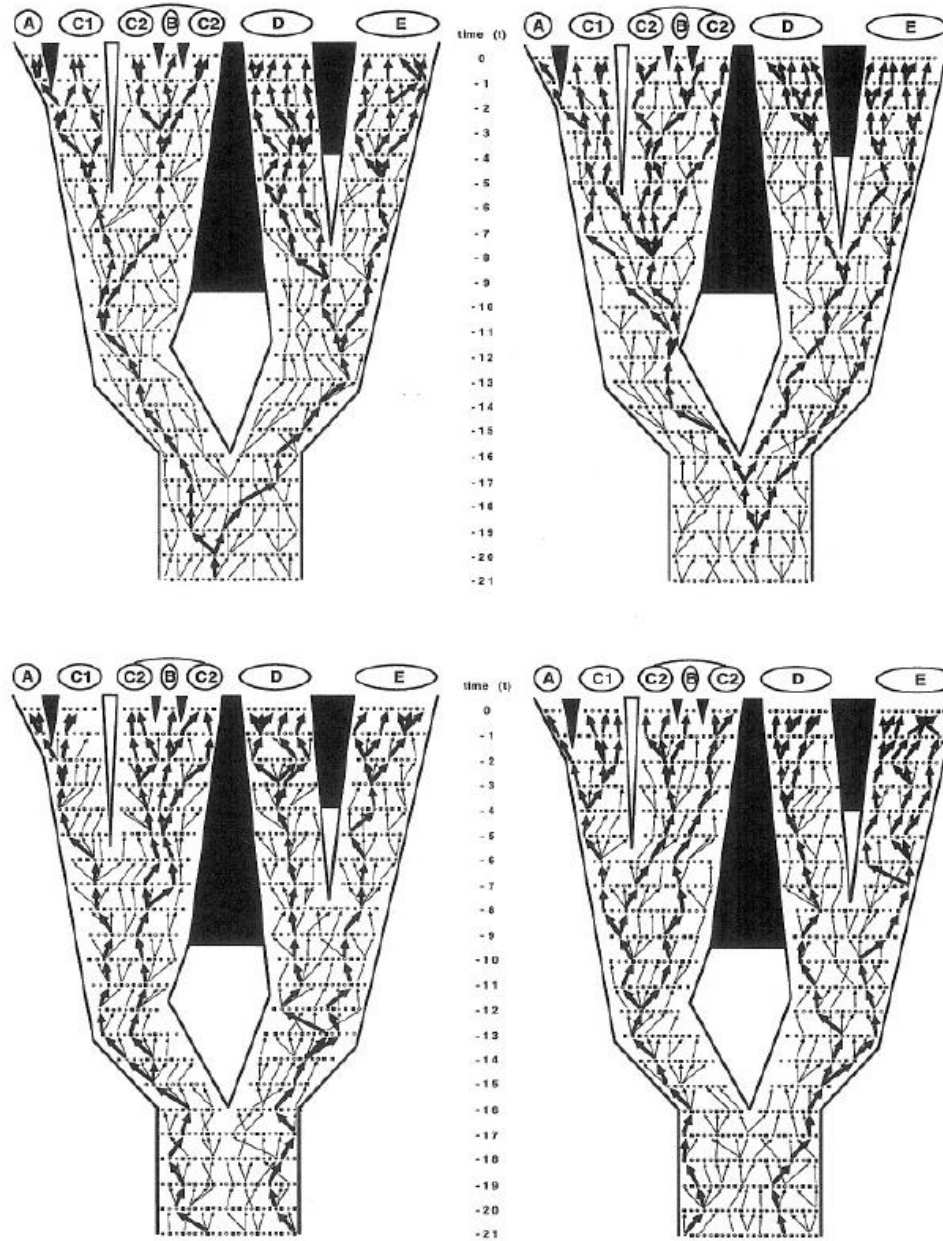
Arvores de genes e de espécies podem contar histórias diferentes



Genes trees



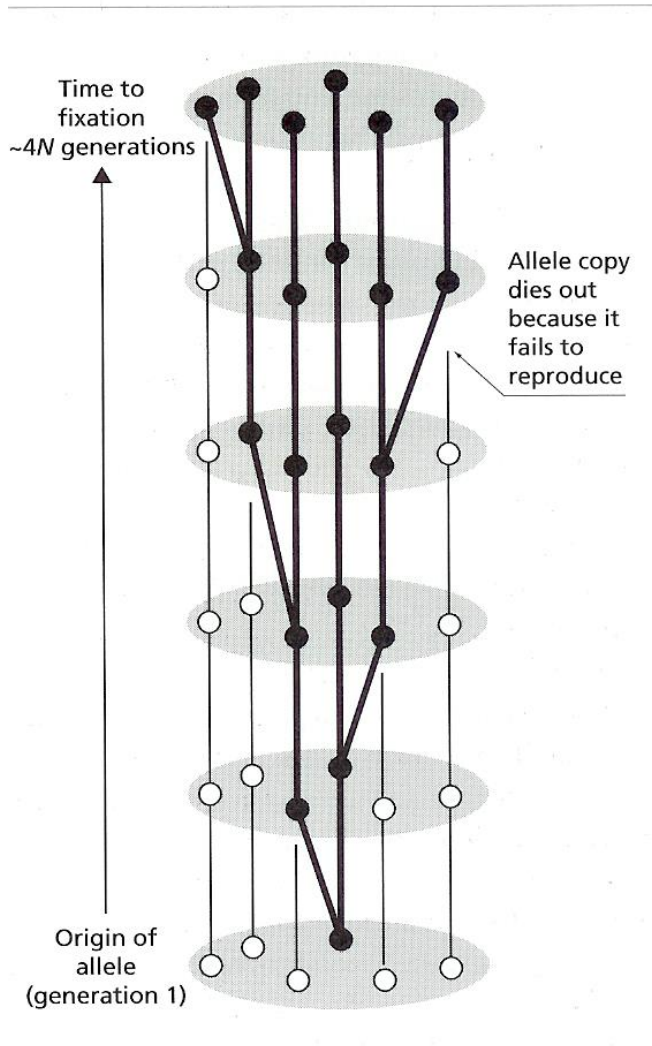
Lineage sorting



Genomic compartments

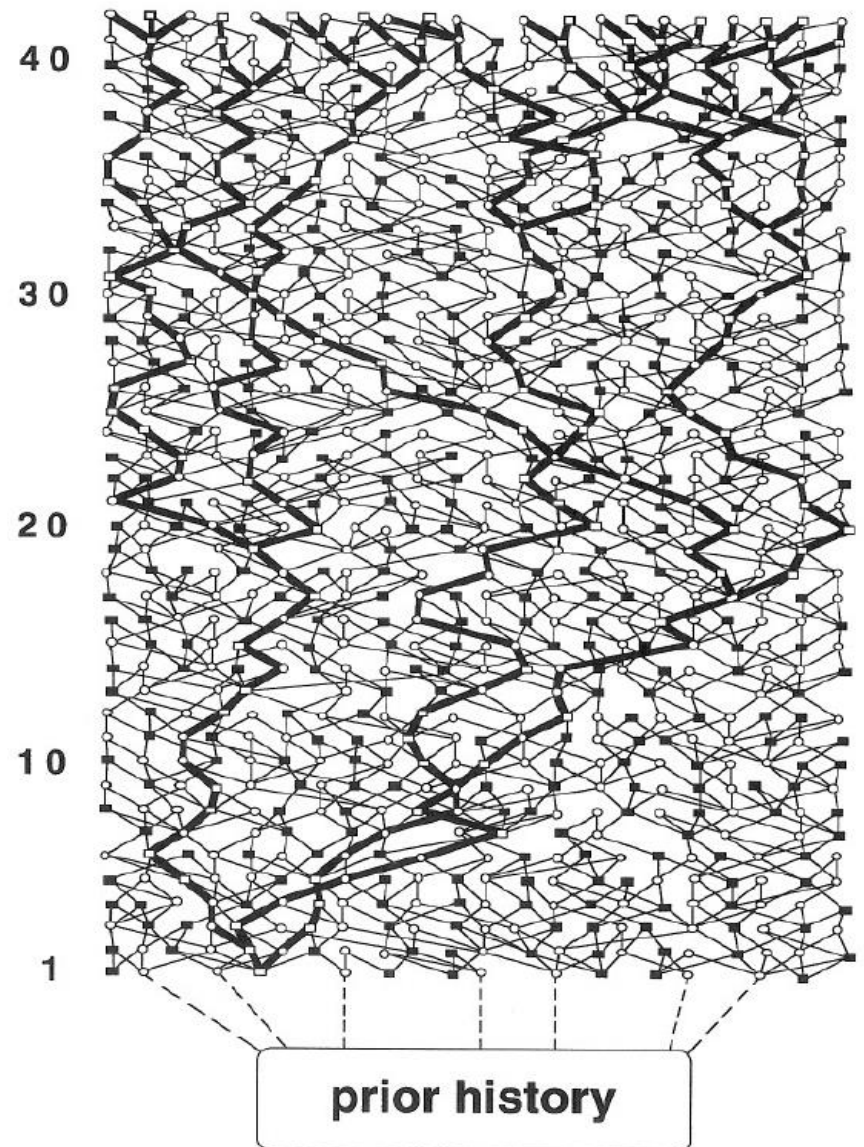
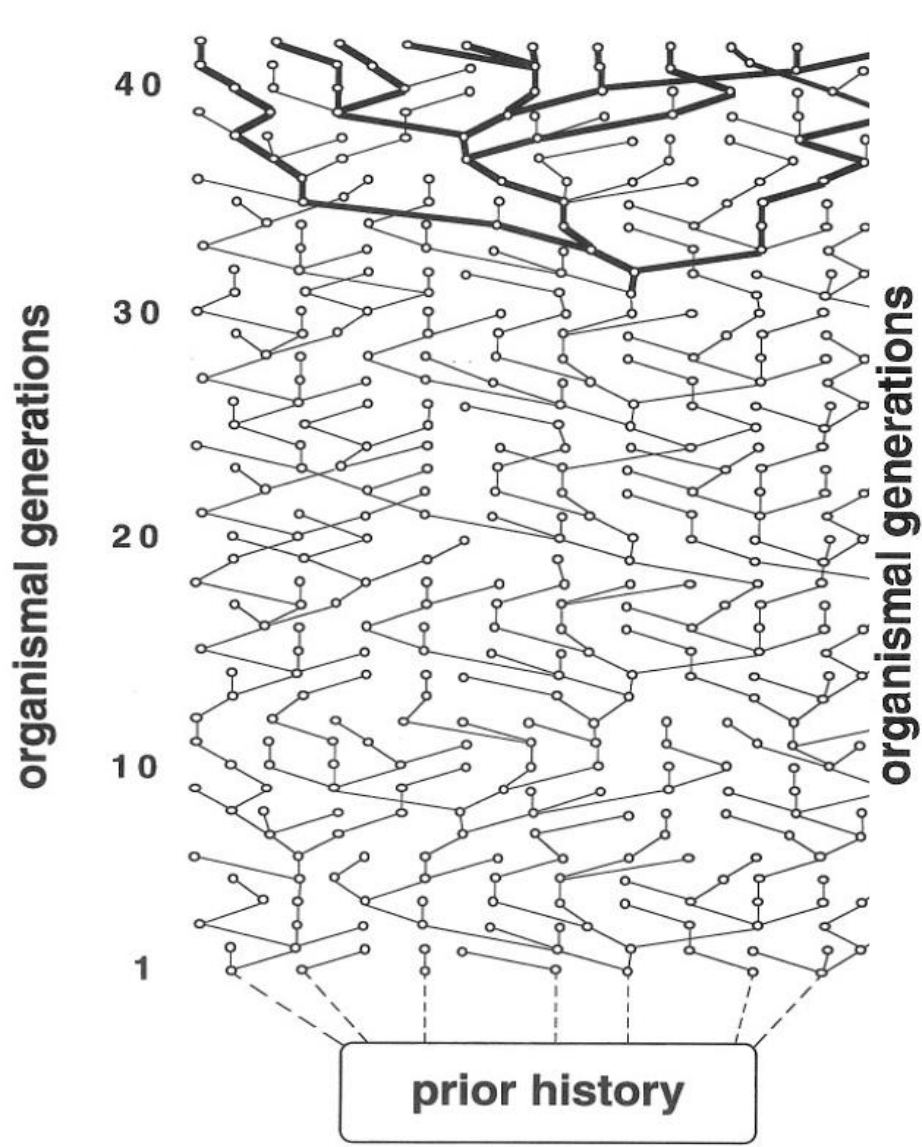
Feature	Genomic compartment			
	Autosomes	X chromosomes	NR _Y	mtDNA
Location	Nuclear	Nuclear	Nuclear	Cytoplasmic
Inheritance	Bi-parental	Bi-parental	Uni-parental	Uni-parental
Ploidy	Diploid	Haploid–diploid	Haploid	Haploid
Relative N_e	4	3	1	1
Recombination rate	Variable	Variable	Zero	Zero
Mutation rate	Low	Low	Low	High

Ne nuclear e Ne mitochondrial

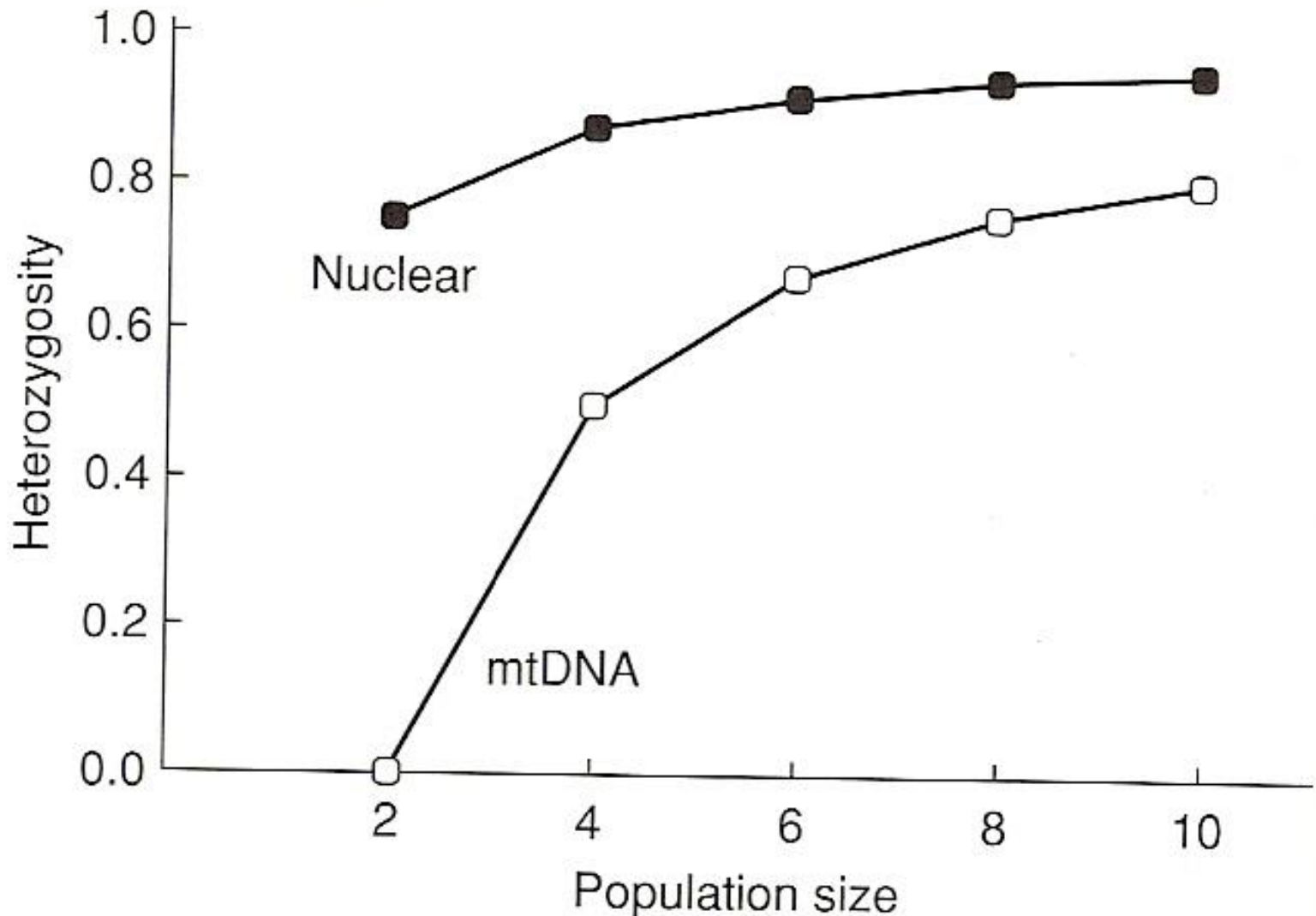


- Ne mitochondrial é um quarto do Ne nuclear.
- A fixação demora N gerações em vez de 4N gerações
- haploide e herdado maternalmente

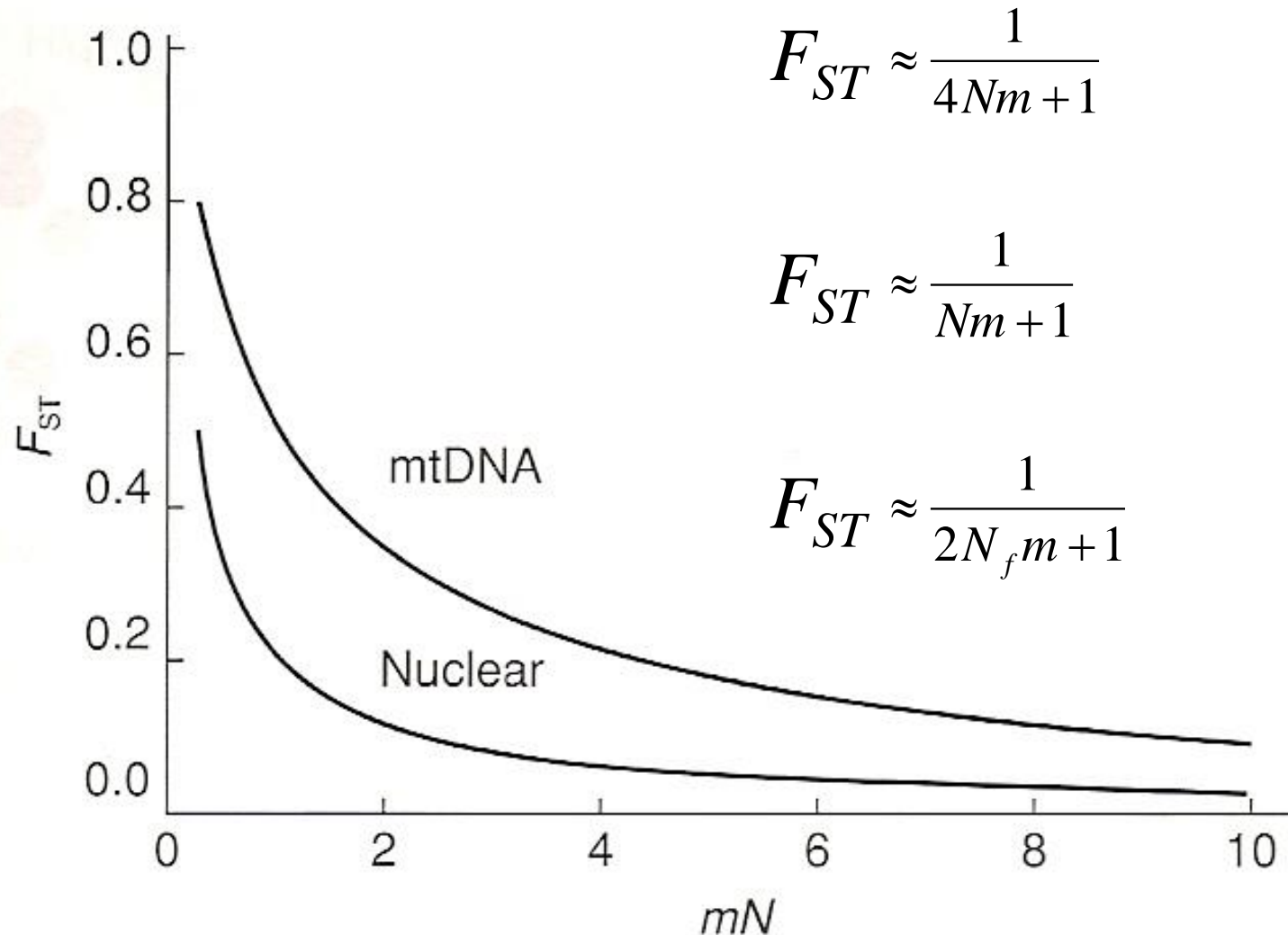
Lineage sorting - DNA Mitochondrial vs nuclear



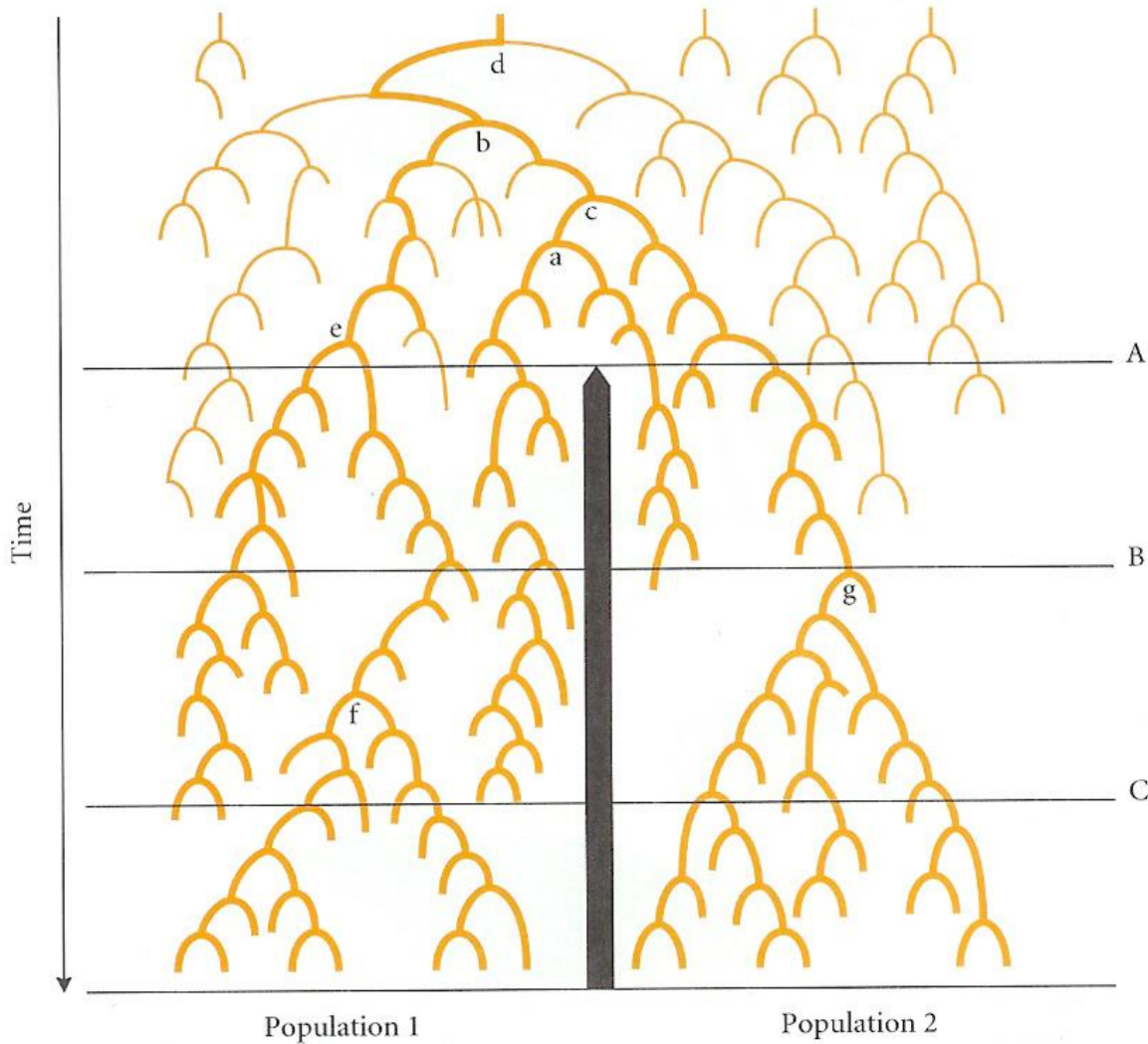
Bottleneck – Nuclear vs mtDNA



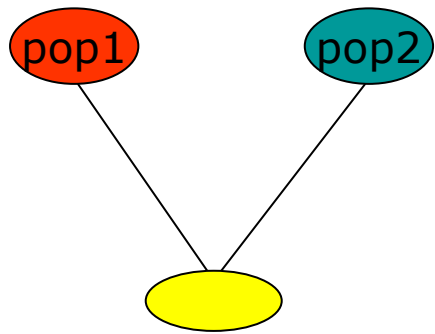
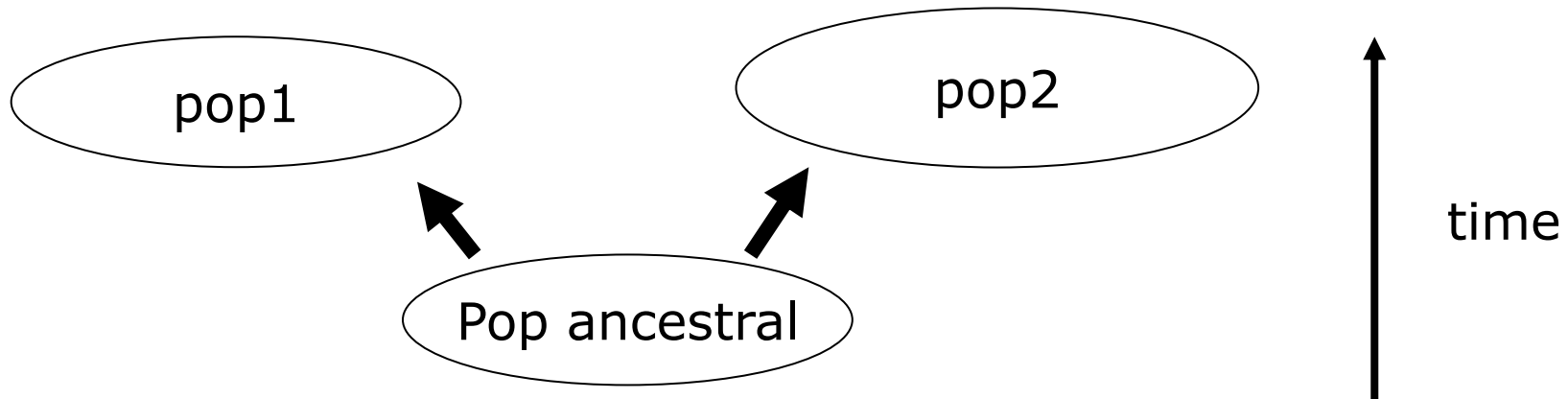
Fst nuclear vs mtDNA



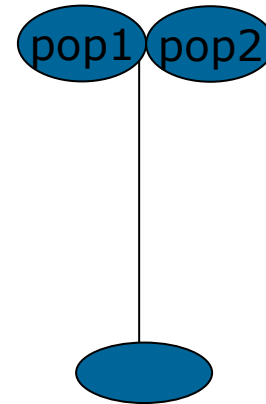
Diferentes estadios consuante os marcadores



mtDNA and nuclear

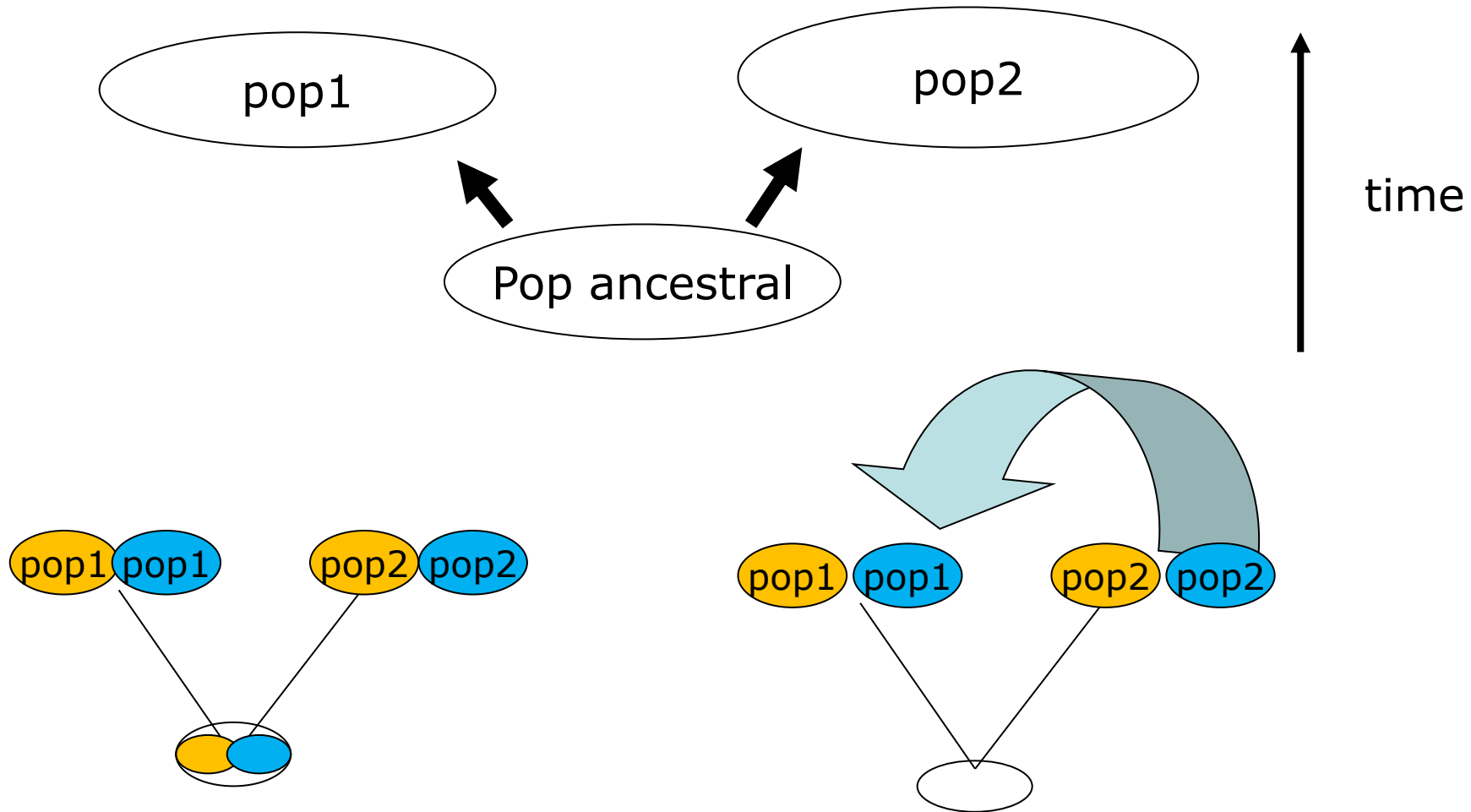


mtDNA

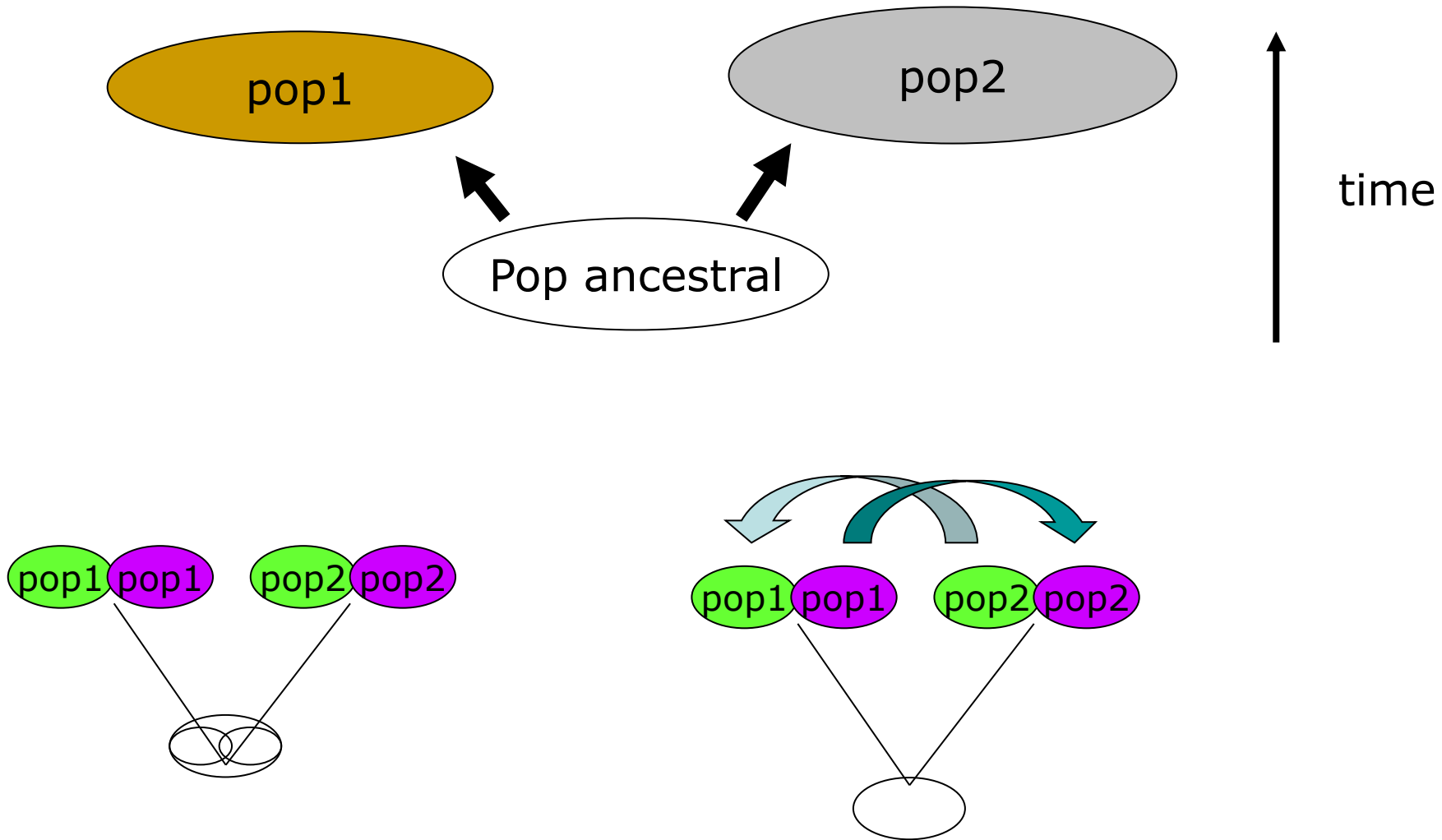


nuclear

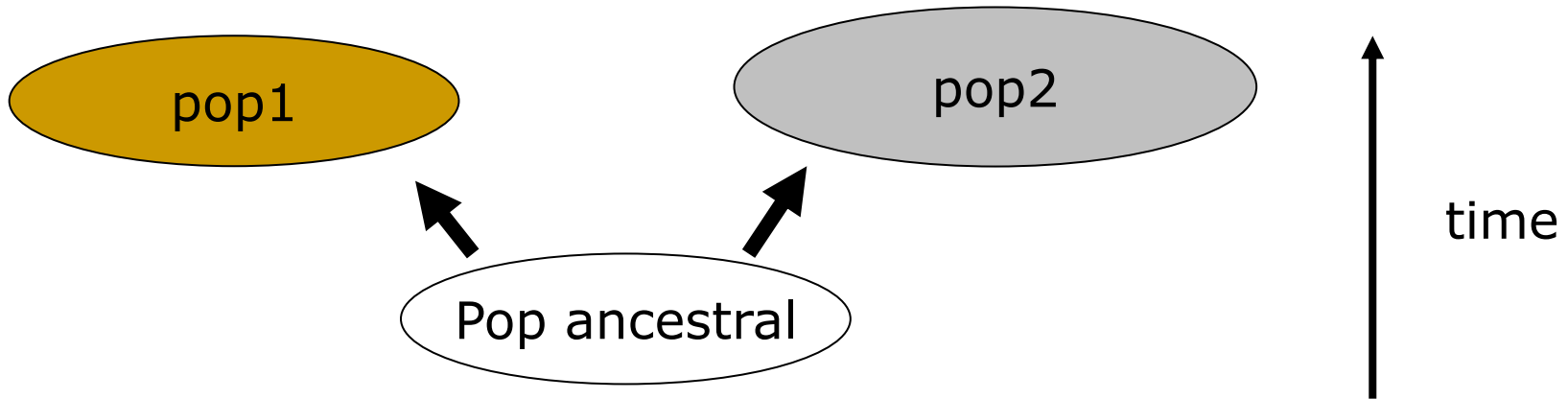
Relative efficacy of molecular markers



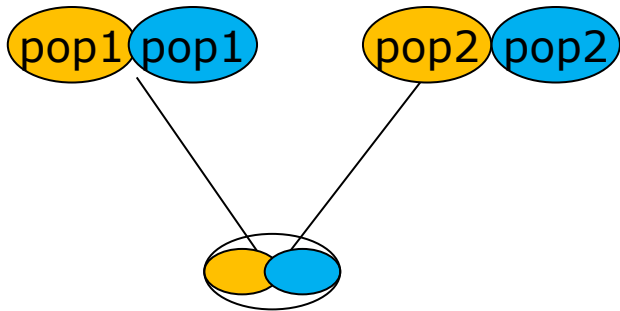
Relative efficacy of molecular markers



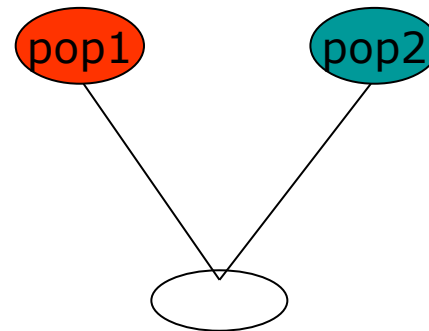
mtDNA and nuclear



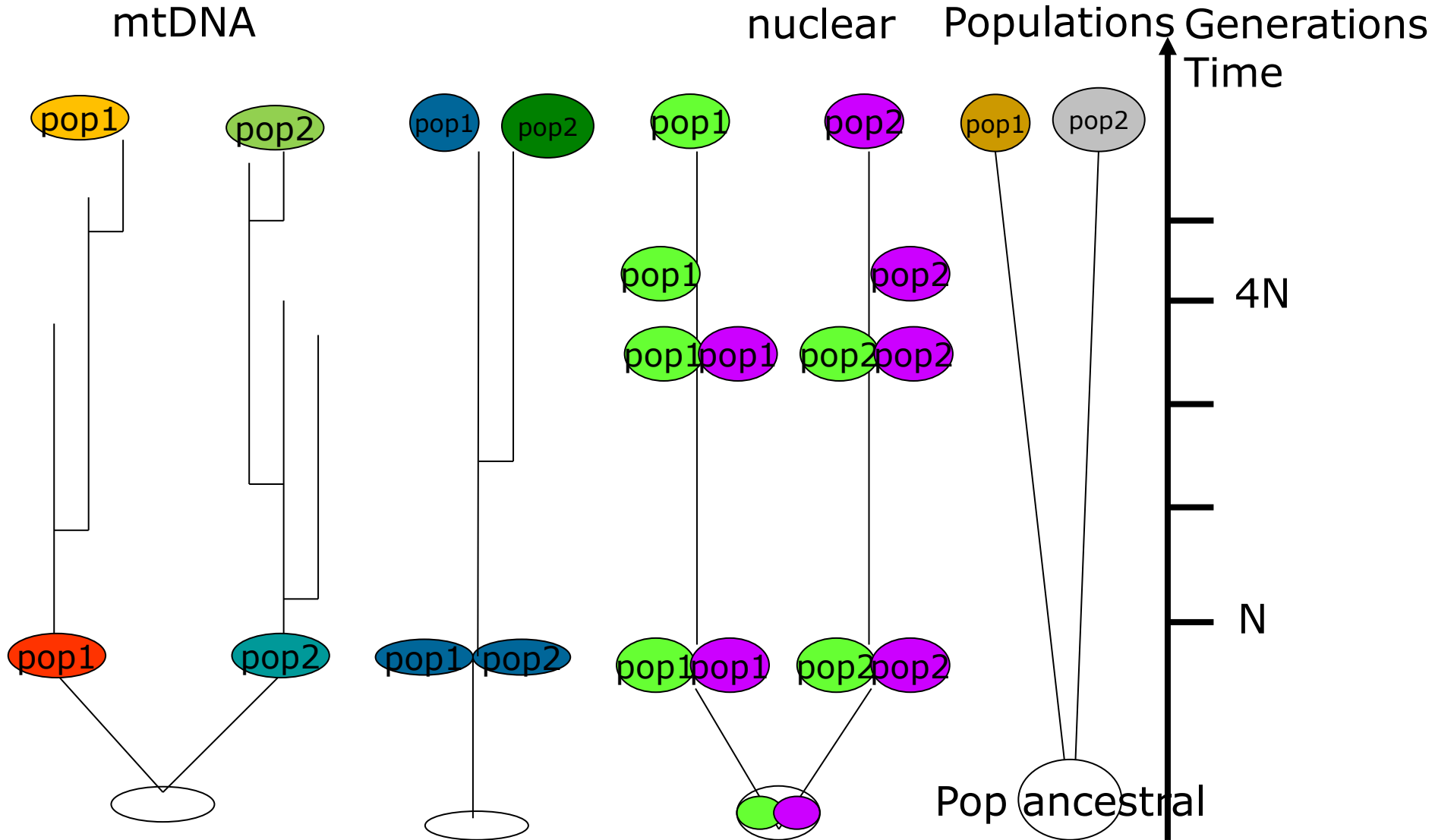
nuclear



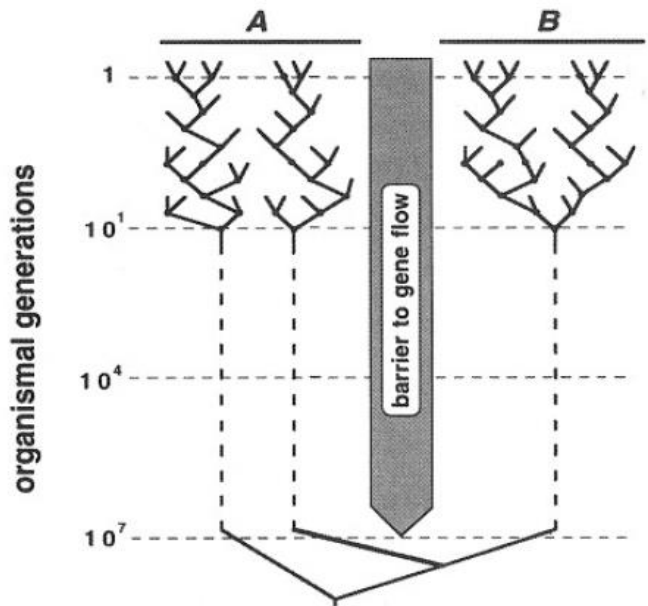
mtDNA



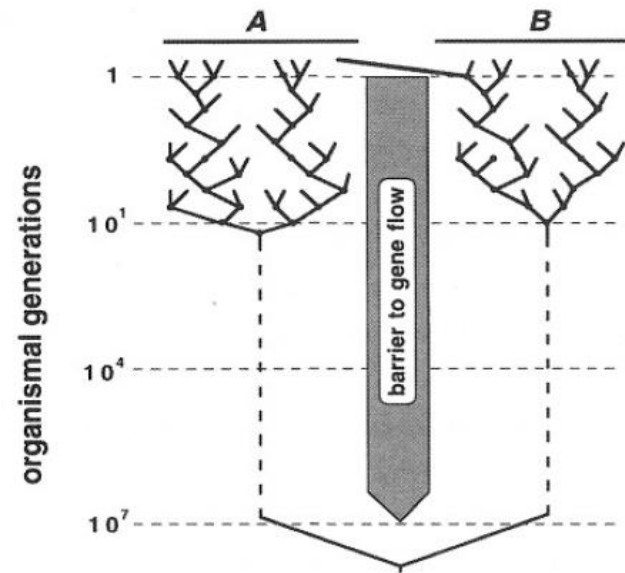
mtDNA and nuclear and time



Genealogical pattern



genealogical pattern due to incomplete lineage sorting



genealogical pattern due to secondary gene flow

Genealogical pattern

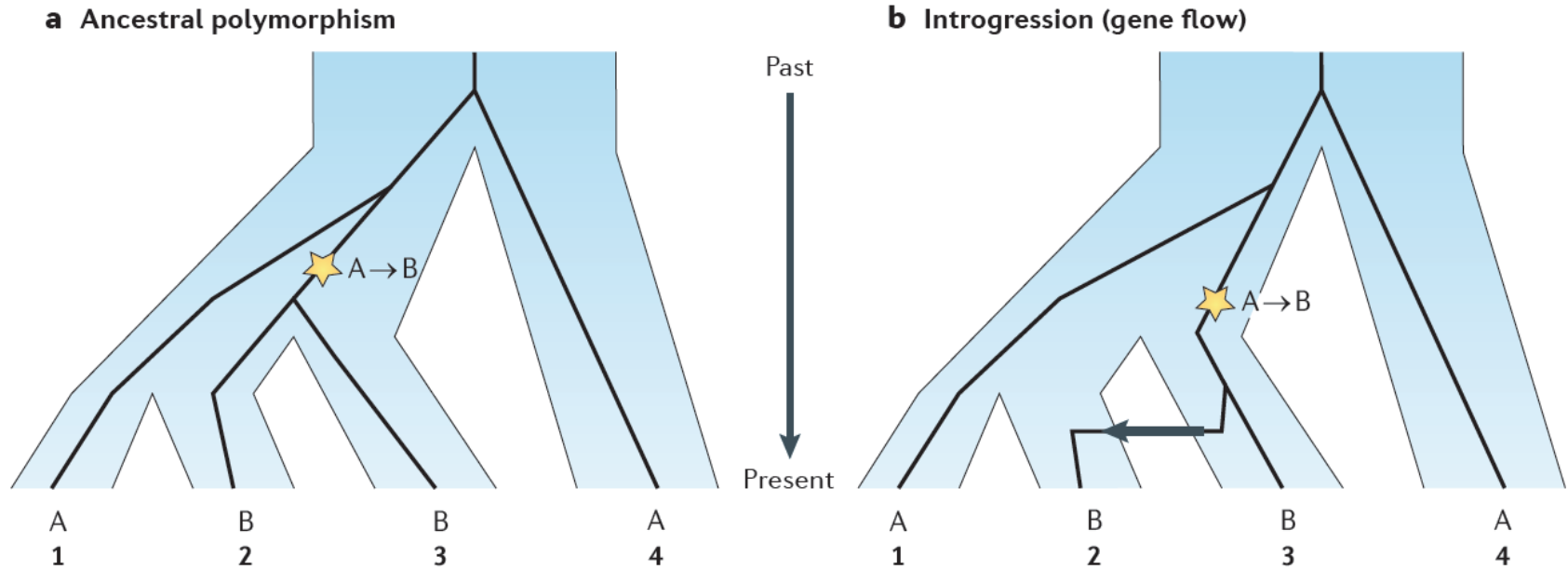


Figure 2 | **Disentangling ancestral polymorphism from gene flow (ABBA and BABA test).** The diagram shows the divergence of two sister populations (1 and 2), a third population (potential source of introgressed genes; 3) and an outgroup population (4) over time. The black line represents the gene tree of a given site, and the star represents a mutation from the ancestral state (allele A) to the derived state (allele B). The pattern ABBA can occur owing to an ancestral polymorphism (a): that is, coalescent of lineage from population 2 with lineage from population 3 in the ancestral population (population ancestral to populations 1, 2 and 3), or gene flow from population 3 to population 2 (b). Under a model with no gene flow, we expect that the pattern ABBA is as frequent as BABA owing to the fact that there is 50% chance that either the lineage from population 1 or from population 2 coalesces with lineage from population 3 in the population ancestral to populations 1, 2 and 3.

Sousa *et al.* 2013 NRG

mtDNA vs nuclear

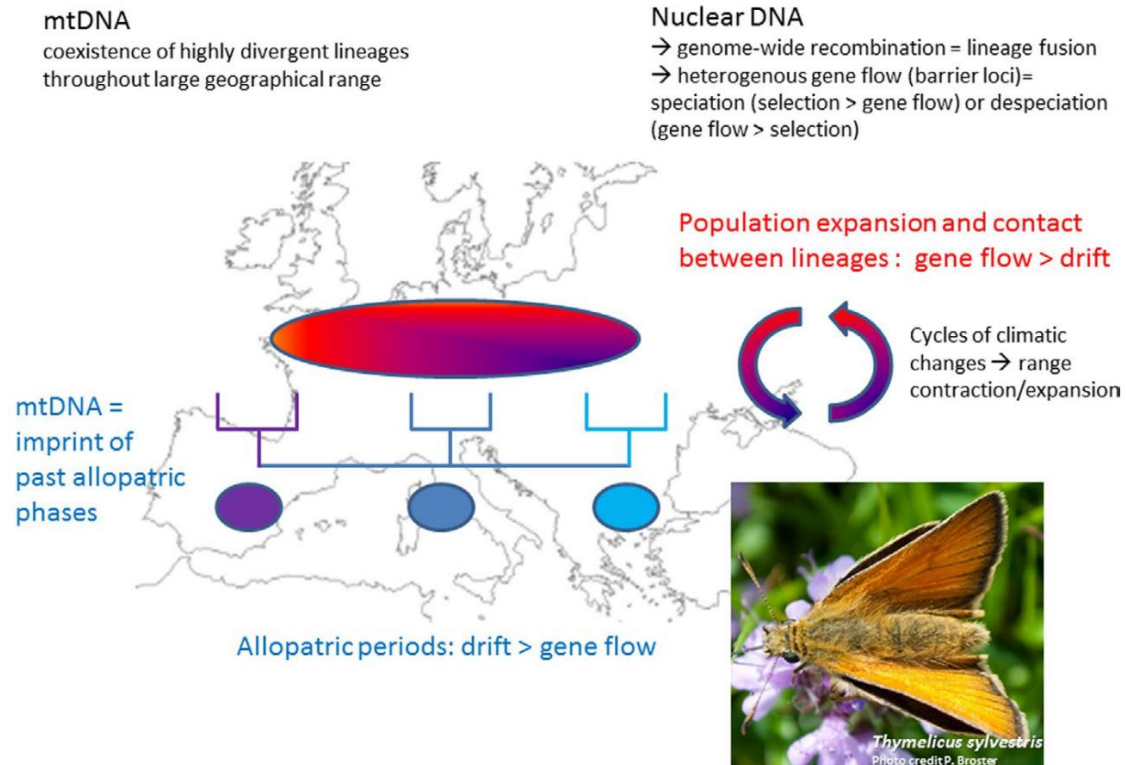
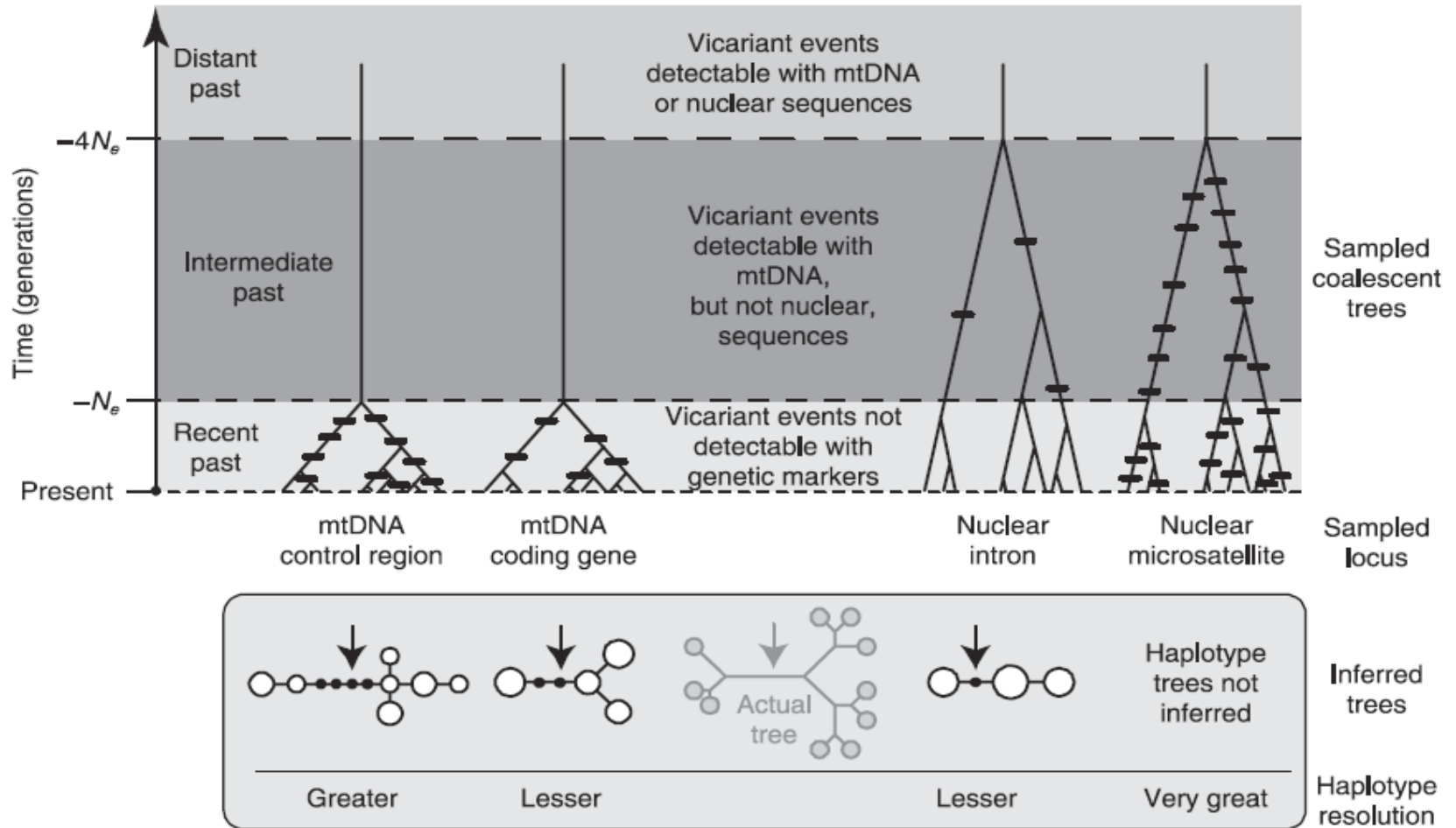
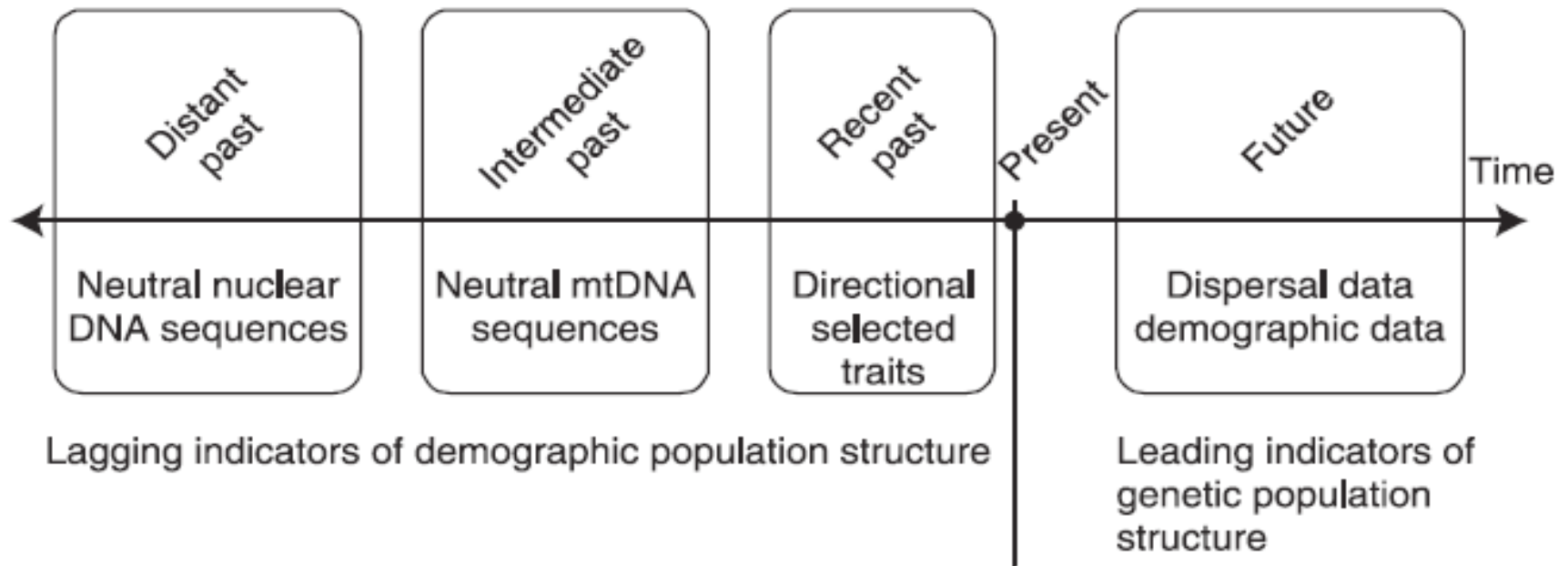


FIGURE 1 Climatic fluctuations have strong impact on species geographical distributions. The small skipper *Thymelicus sylvestris*, a widespread European butterfly, has experienced cycles of range contraction/expansion (Hijonosa et al., 2019). During Pleistocene glacial periods, the species was fragmented into small geographically isolated populations, and drift was the predominant evolutionary force, acting especially on mtDNA (small effective population size). During warming periods, populations expanded, and genome-wide recombination took place between previously diverging lineages, while nonrecombinant highly divergent mt lineages neutrally diffused over large geographical areas, gene flow overcoming drift. Another possible outcome of secondary contact (not observed in *T. sylvestris*) could be heterogeneous gene flow due to selection against hybrids if some isolation mechanisms (barrier loci) evolved during the allopatric phase, ultimately leading to speciation or despeciation, depending on the relative force of selection over gene flow. Photo credit: Peter Broster

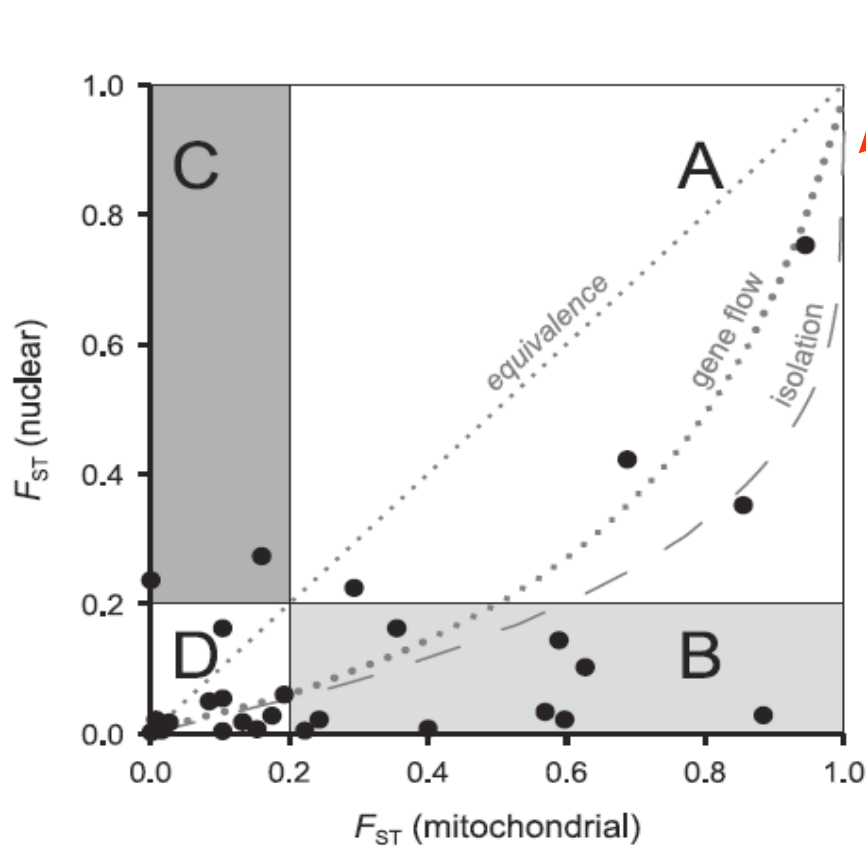
Relative efficacy of molecular markers



Relative efficacy of molecular markers



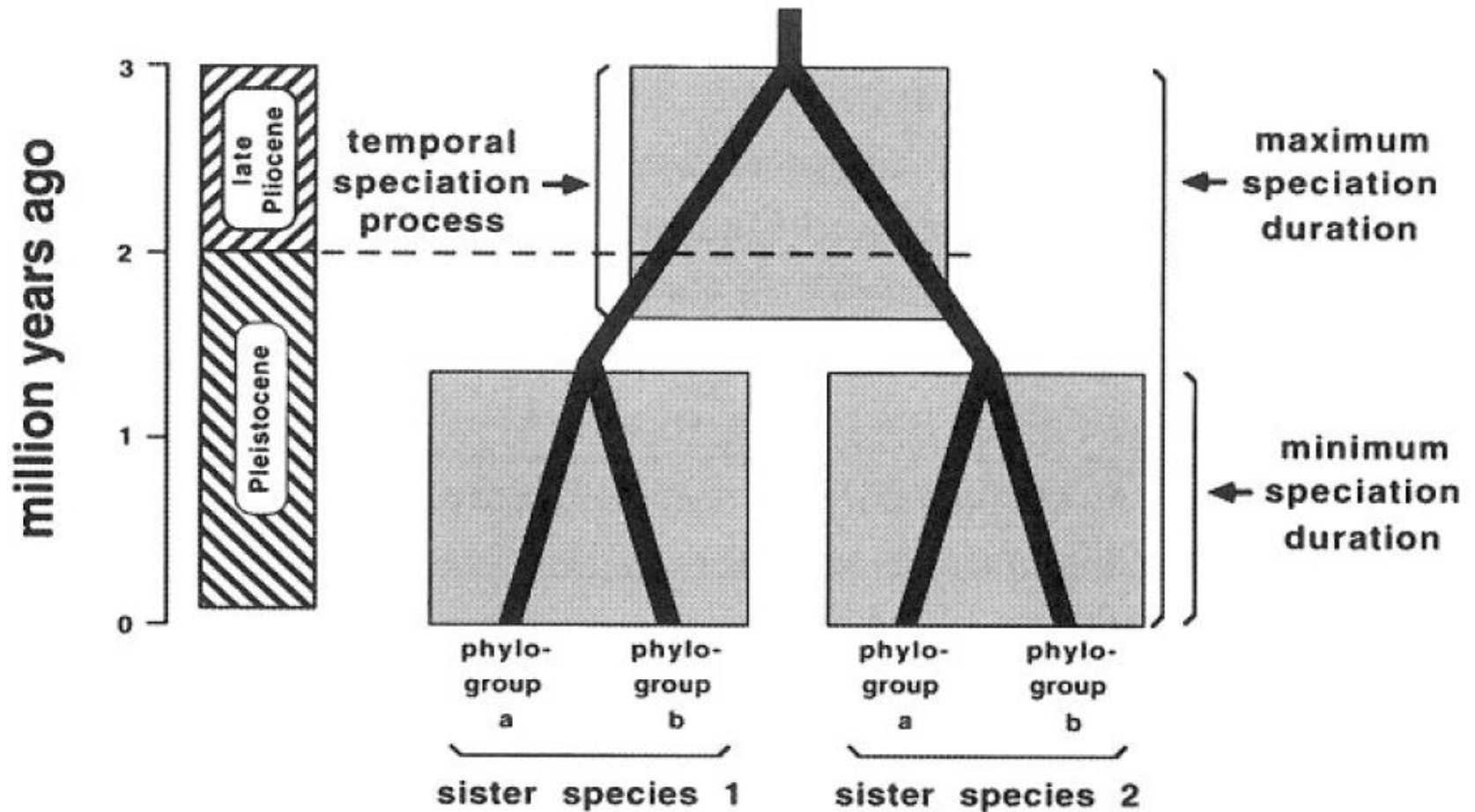
Relative efficacy of molecular markers



vicariance

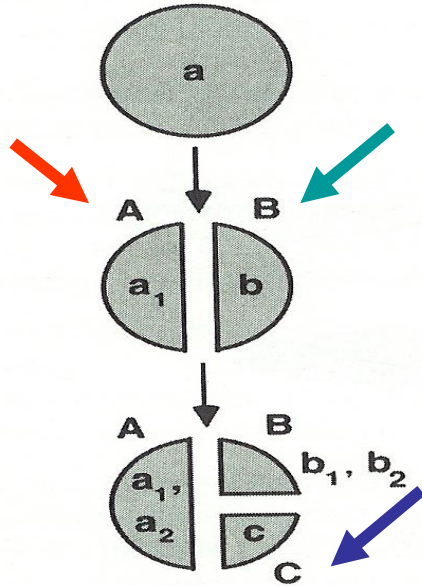
Fig. 1 Mitochondrial vs. nuclear estimates of F_{ST} for some published studies of birds (dots). Approximate regions of our four categories of phylogeographical results (Table 1) are shown with background shading; categories A and D indicate generally consistent results between mitochondrial and nuclear markers, category B results are consistent given differences in effective population size and coalescent times, and category C results are inconsistent (see text). Dotted lines show the expected relationships between mitochondrial and nuclear F_{ST} for cases in which they are identical (equivalence) and for gene flow at equilibrium under an island model of population structure (gene flow). Dashed line shows the expected temporal approach to equilibrium following vicariance (isolation). Space below the diagonal represents instances in which mitochondrial markers are a more sensitive indicator of population structure; space above the diagonal represents cases in which nuclear markers are more sensitive (see Appendix).

Phylogroups

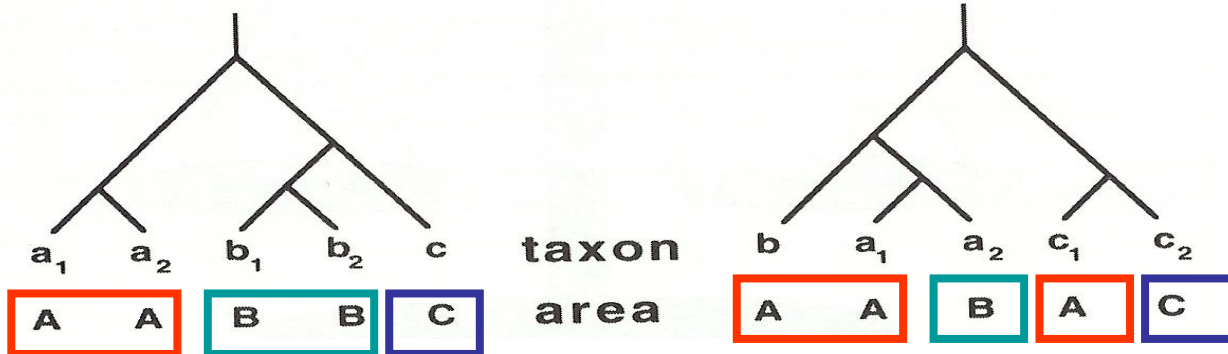
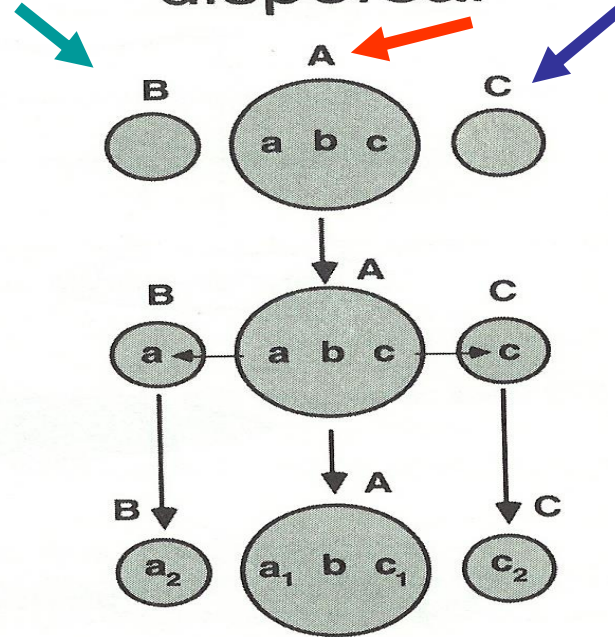


Vicariance vs dispersal

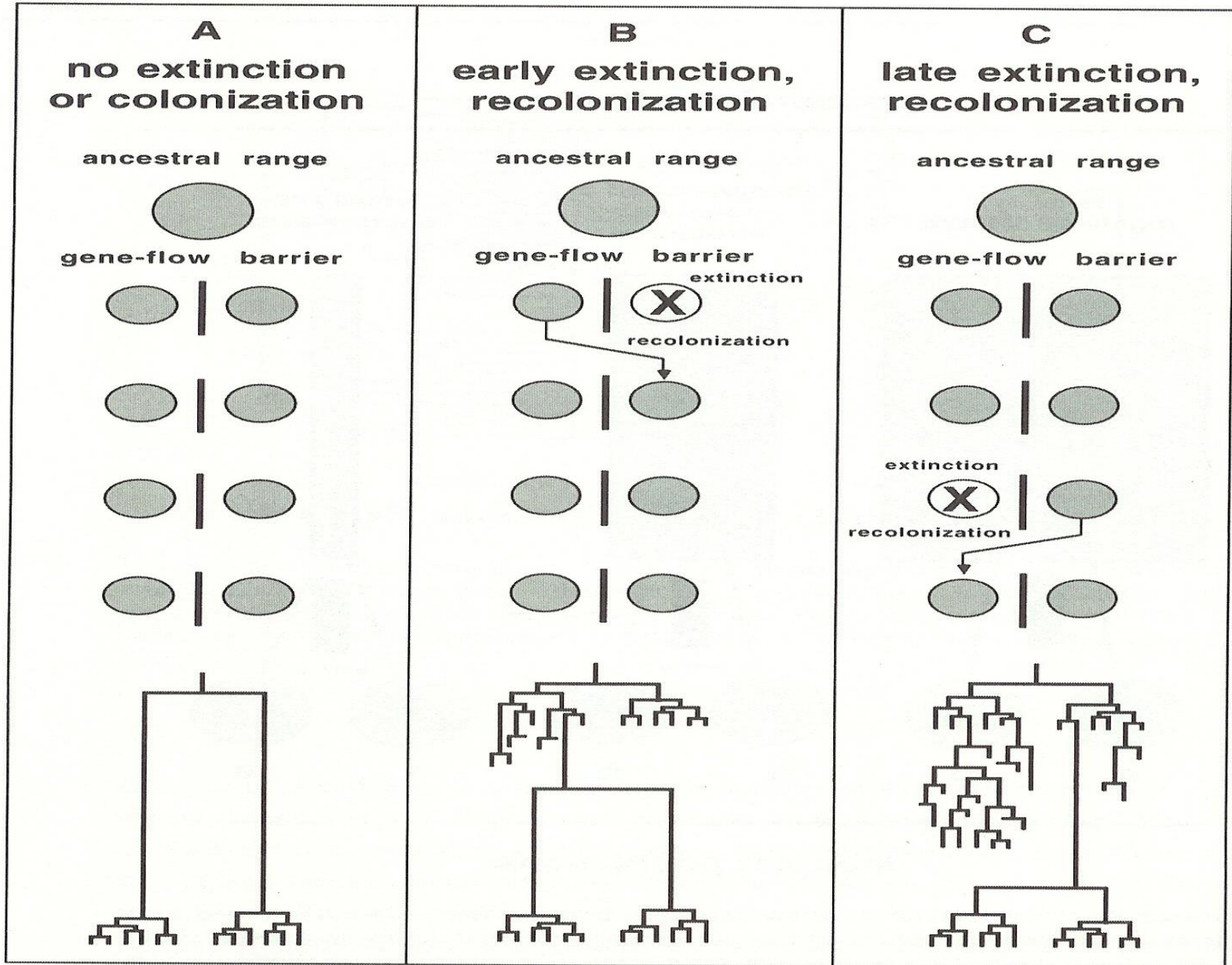
vicariance



dispersal



Genes e Geografia



Genes e Geografia

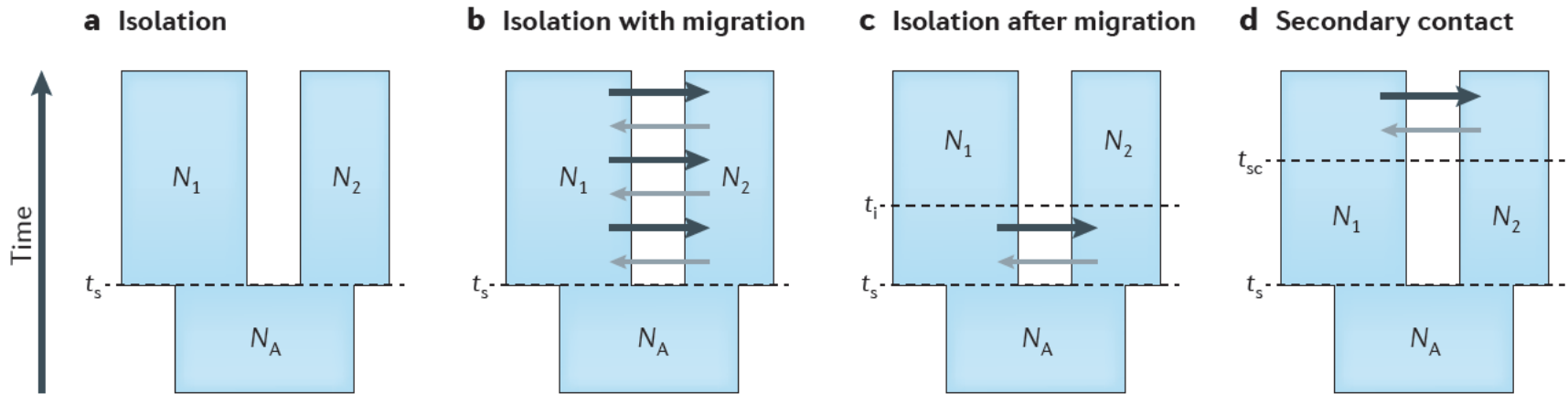


Figure 1 | **Alternative modes of divergence.** All models assume that an ancestral population of size N_A splits into two populations at time of split (t_s). The two present-day populations have effective sizes N_1 and N_2 , respectively. Panel **a** shows the model in which migration rate is zero in both directions, which corresponds to an allopatric divergence scenario. Panels **b–d** represent alternative models in which populations have been exchanging migrants. Gene flow occurs at constant rates since the split from the ancestral population (**b**). Migration rates are assumed to be constant through time, but gene flow can be asymmetric: that is, one migration rate for each direction. Panel **c** shows a scenario in which populations begin diverging in the presence of gene flow but experience a cessation of gene flow after time since isolation (t_i). If the lack of current gene flow in this model is due to reproductive isolation then this represents a history in which divergence occurred to the point of speciation in the presence of gene flow. In panel **d**, we consider the alternative migration history in which populations were isolated and diverged for a period of time in the absence of gene flow, followed by secondary contact at time of secondary contact (t_{sc}) and the introgression of alleles from the other population by gene flow.

Sousa *et al.* 2013 NRG

Neanderthal

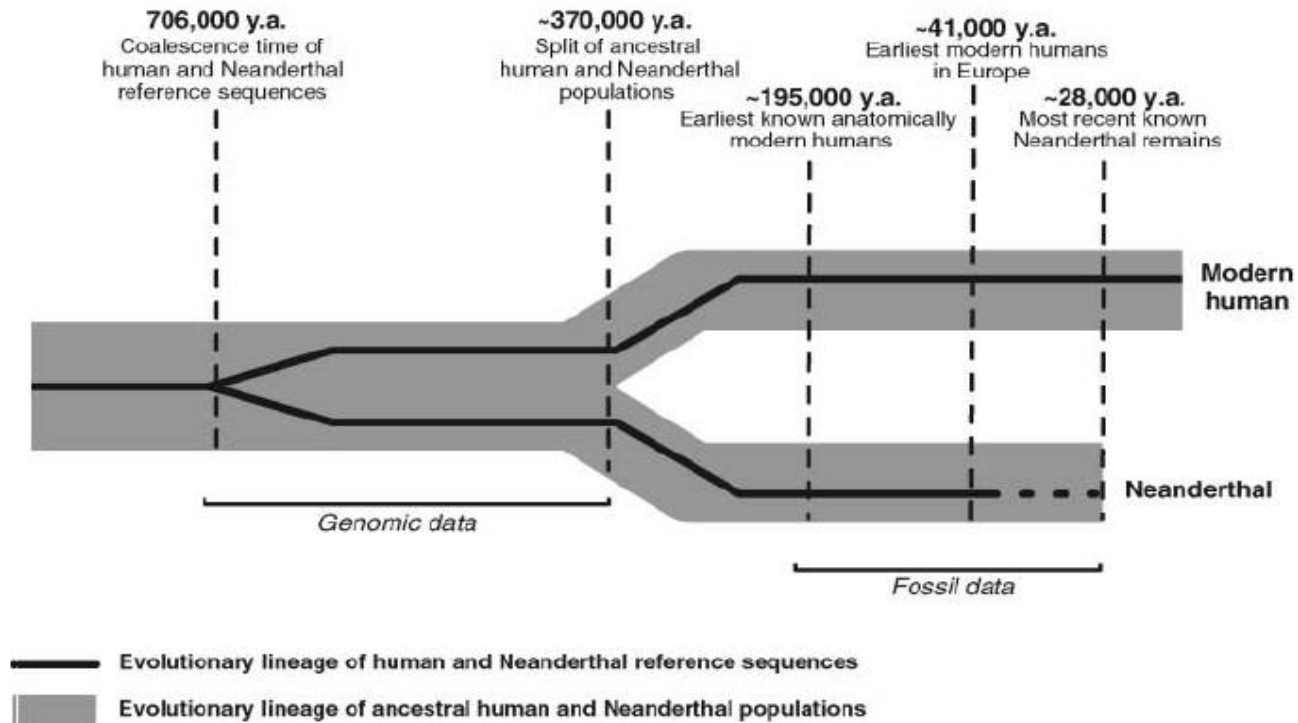
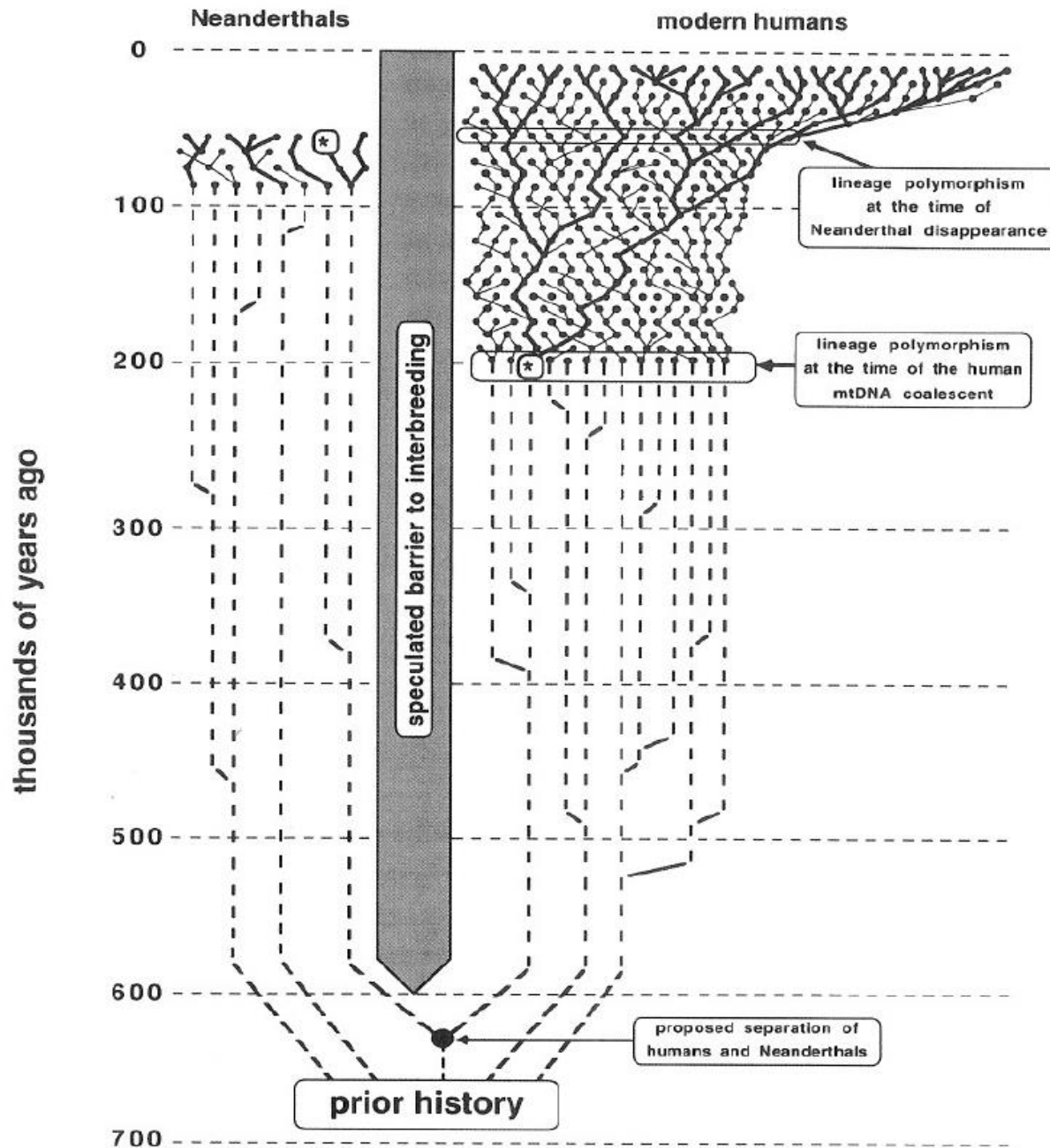



Fig. 6. Divergence estimates for human and Neanderthal genomic sequences and ancestral human and Neanderthal populations, shown relative to dates of critical events in modern human and Neanderthal evolution (2, 22, 25). The branch lengths are schematic and not to scale. y.a., years ago.



Gender dispersal and gene flow

female dispersal and gene flow

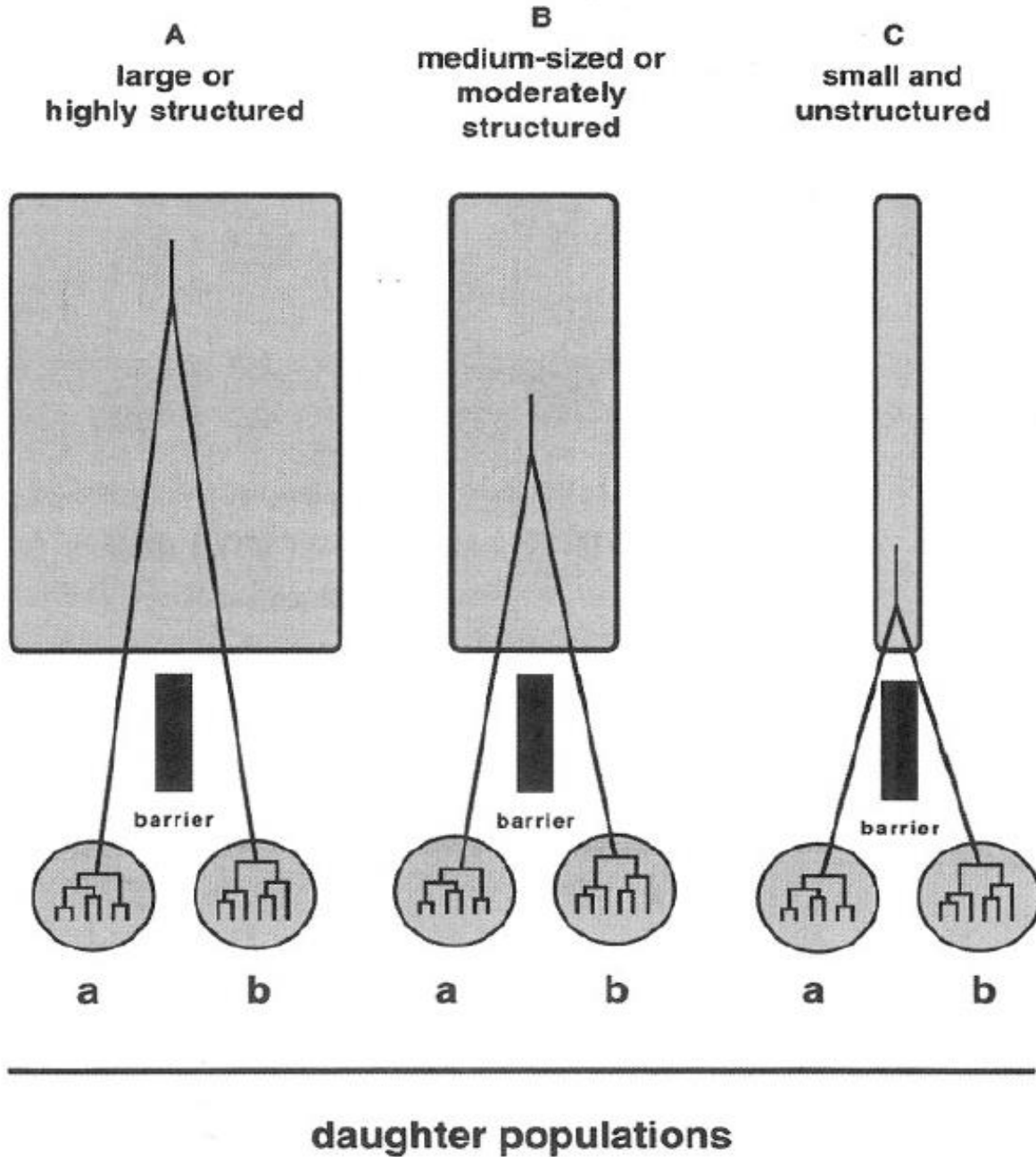
low  high

male dispersal and gene flow

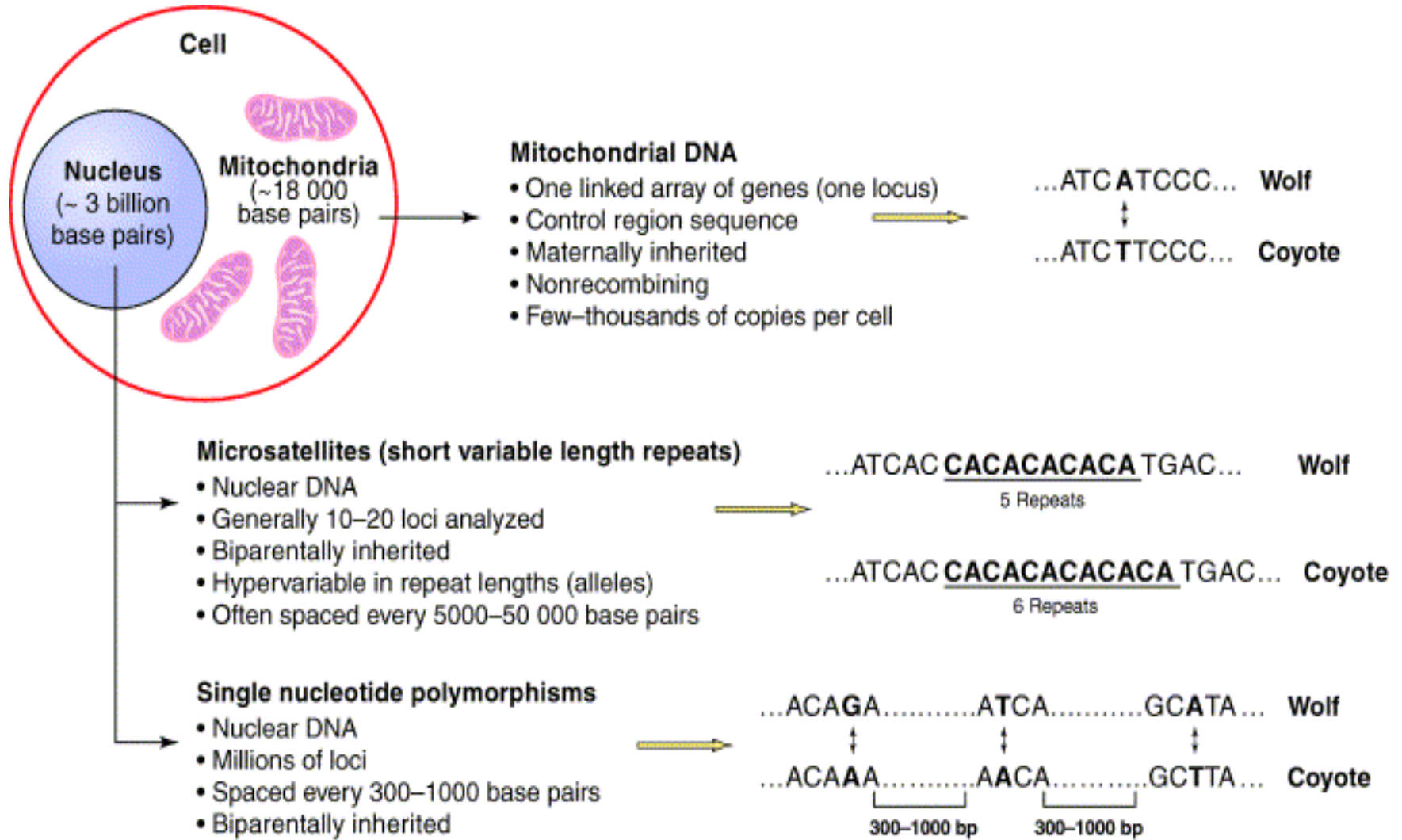
low  high

<p>geographic structure in...</p> <p>mtDNA-- <input checked="" type="radio"/> yes</p> <p>autosomal genes-- yes</p> <p>Y-linked genes-- yes</p> <p>demographic autonomy-- <input checked="" type="radio"/> yes</p>	<p>geographic structure in...</p> <p>mtDNA-- <input type="radio"/> no</p> <p>autosomal genes-- no</p> <p>Y-linked genes-- * * *</p> <p>demographic autonomy-- * * *</p>
<p>geographic structure in...</p> <p>mtDNA (in females)-- <input checked="" type="radio"/> yes</p> <p>autosomal genes-- no</p> <p>Y-linked genes-- no</p> <p>demographic autonomy-- <input checked="" type="radio"/> yes</p>	<p>geographic structure in...</p> <p>mtDNA-- <input type="radio"/> no</p> <p>autosomal genes-- no</p> <p>Y-linked genes-- no</p> <p>demographic autonomy-- <input type="radio"/> no</p>

ancestral populations



Markers



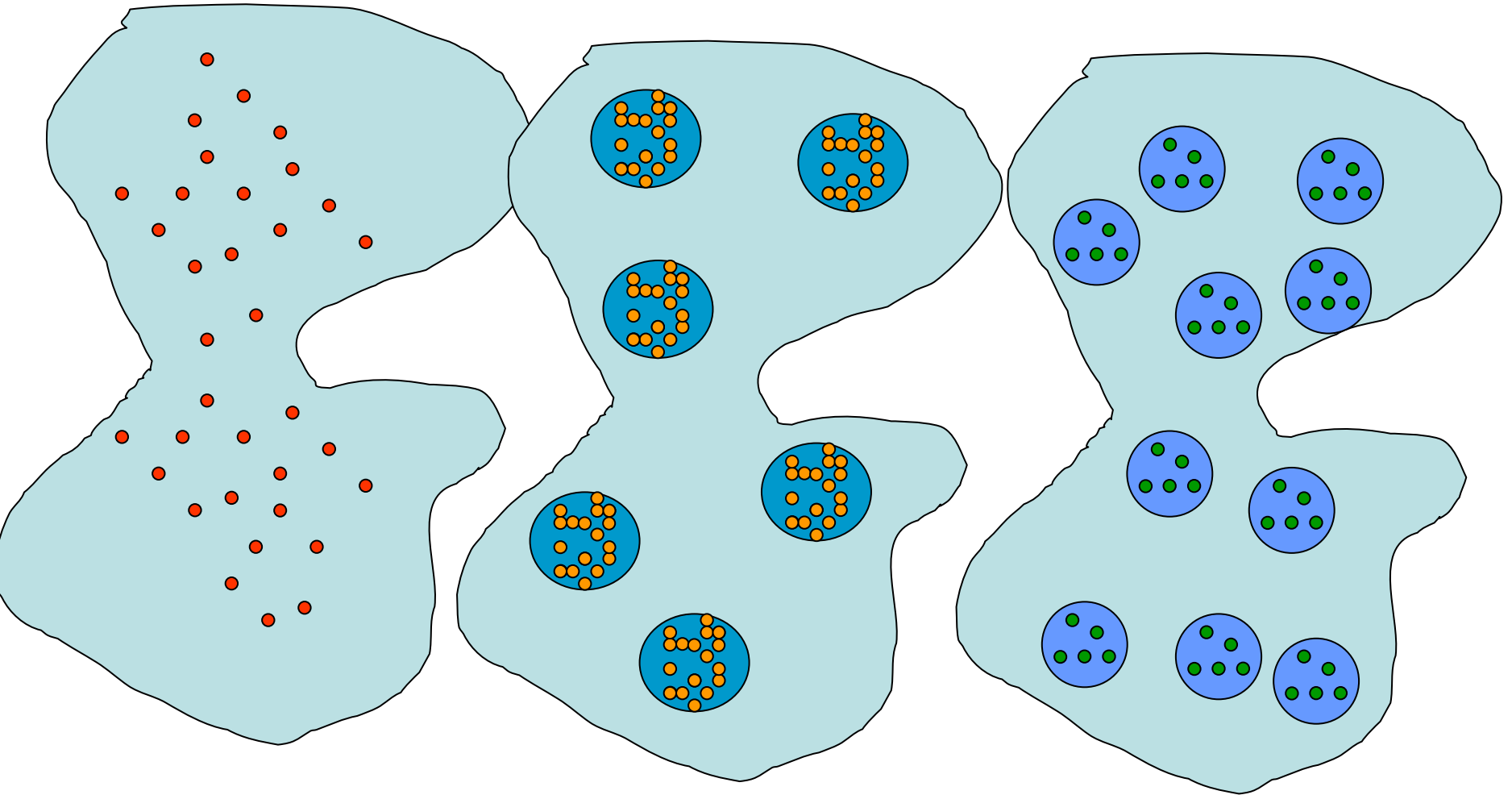
DNA Mitochondrial Problemas?

- 1- numts
- 2- Clonal – ausencia de recombinação
- 3- neutral
- 4- clock

Três tipos de abordagem

- 1- Phylogenetics / cladistic
- 2- Summary statistics and Co
- 3- Coalescence

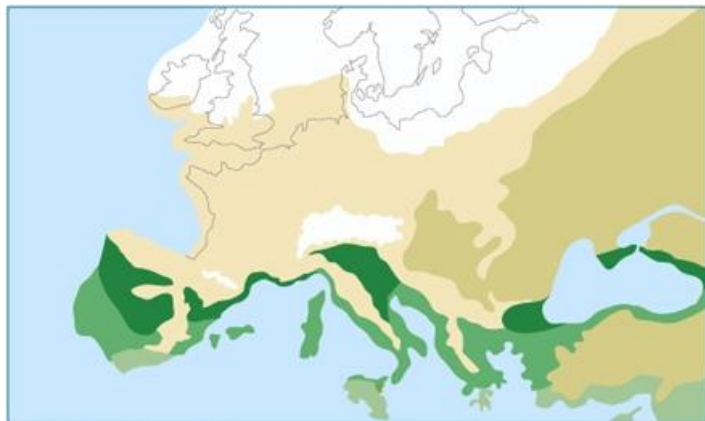
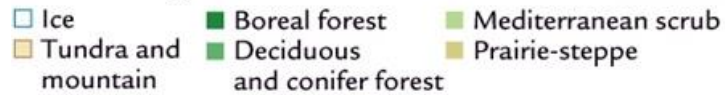
Amostragem/Pergunta



Último máximo glacial

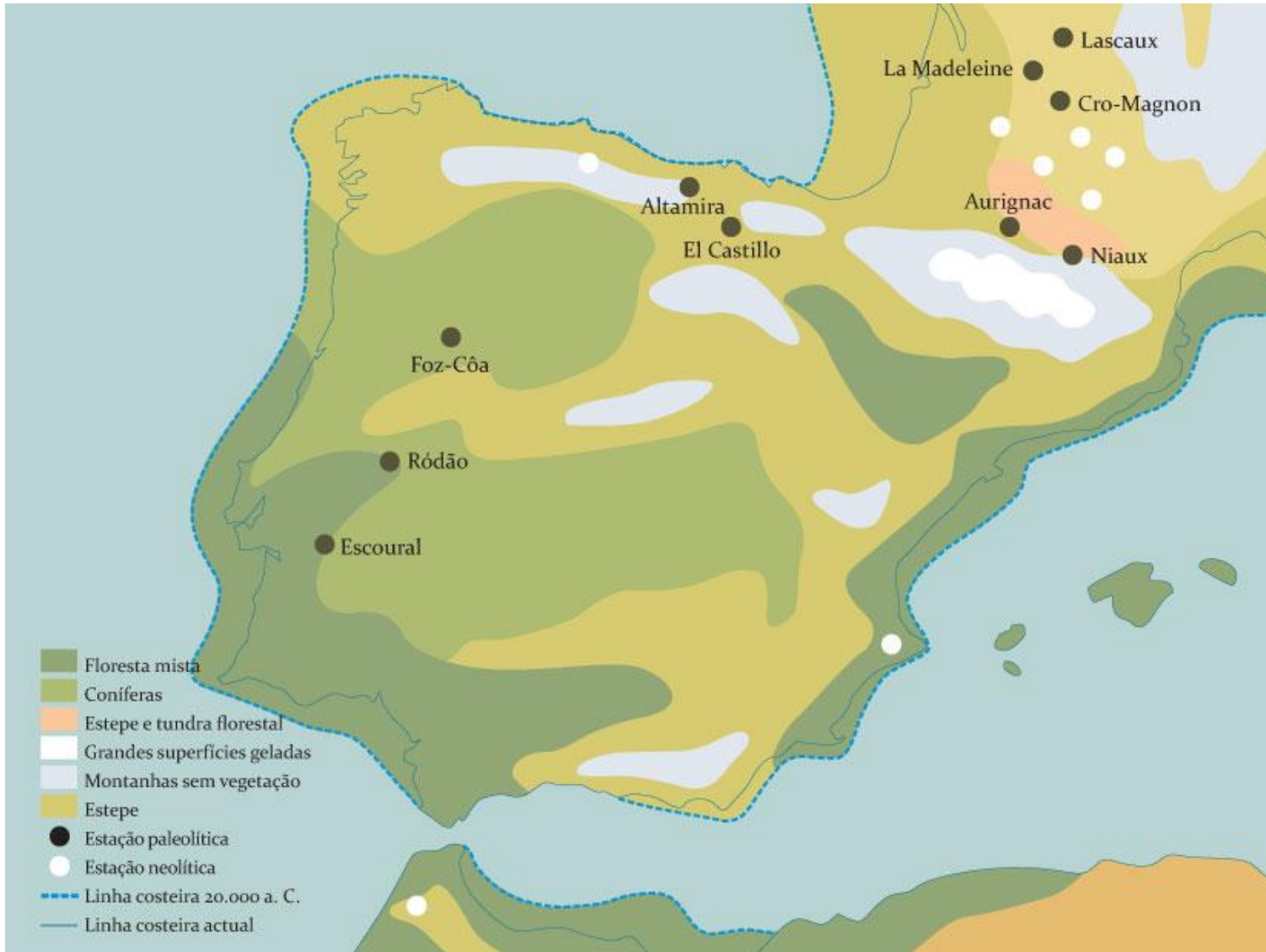


A Modern vegetation

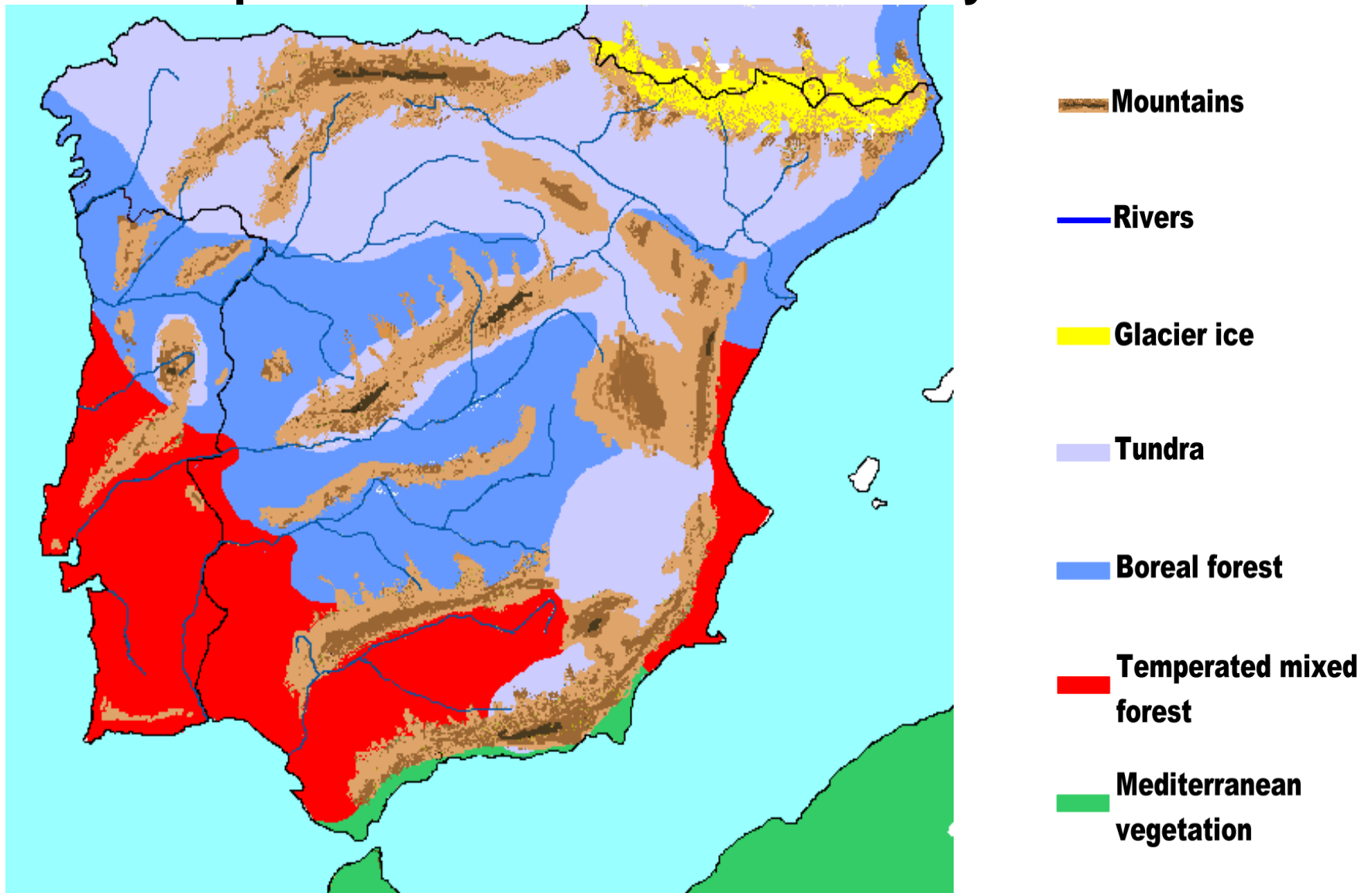


B Glacial vegetation

Biogeografia



Iberian Peninsula: superimposition of ice age vegetation pattern to mountain and river systems



The persistence of Pliocene populations through the Pleistocene climatic cycles: evidence from the phylogeography of an Iberian lizard

O. S. Paulo^{1,4*}, C. Dias¹, M. W. Bruford², W. C. Jordan³ and R. A. Nichols⁴

¹*Centro de Biologia Ambiental/Departamento de Zoologia e Antropologia, Faculdade de Ciências da Universidade de Lisboa, P-1700 Lisboa, Portugal*

²*Cardiff School of Biosciences, Cardiff University, Cathays Park, Cardiff CF1 3TL, UK*

³*Institute of Zoology, Zoological Society of London, London NW1 4RY, UK*

⁴*School of Biological Sciences, Queen Mary and Westfield College, University of London, London E1 4NS, UK*

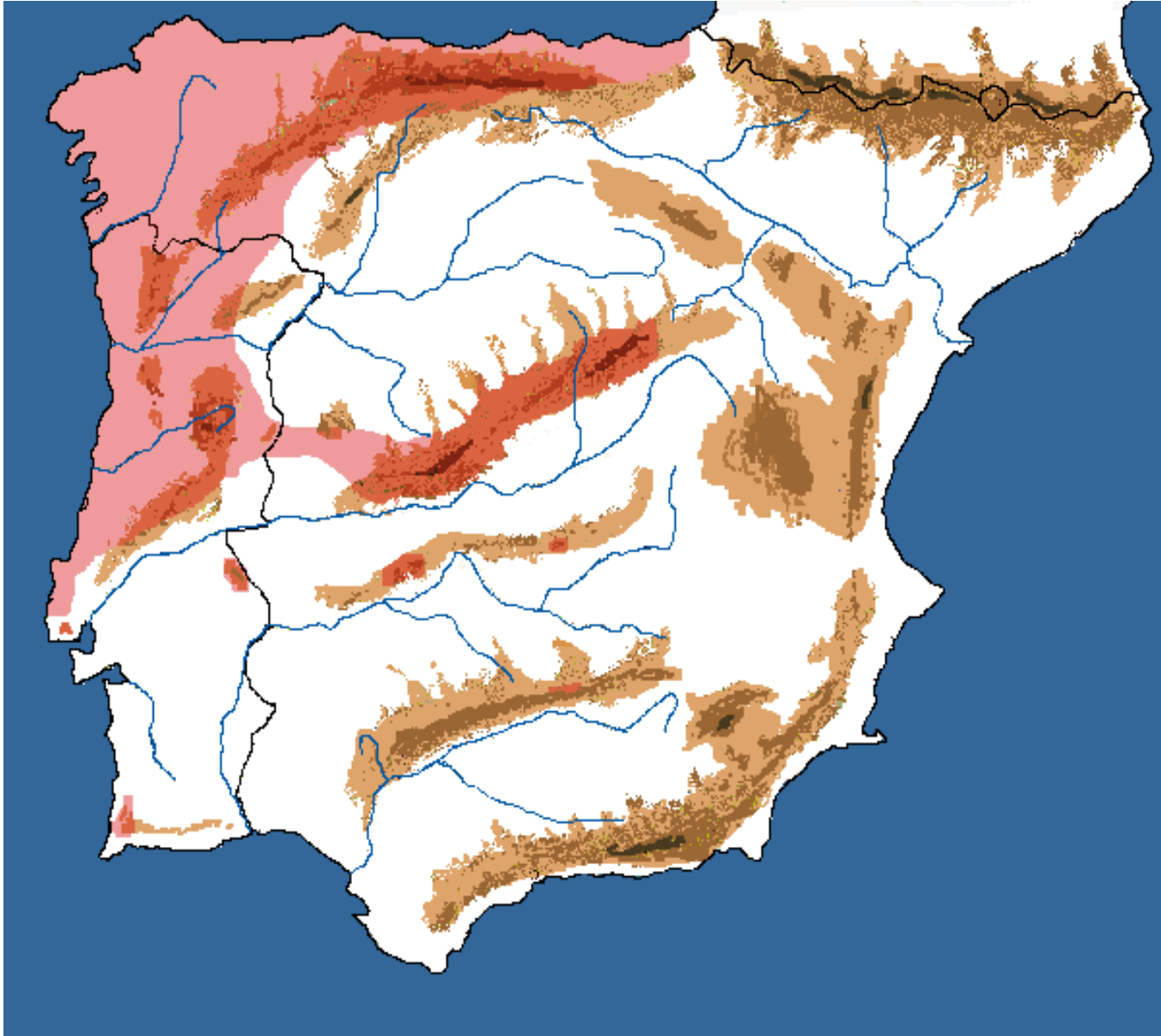
Ancient climatic fluctuations have caused changes in the demography and distribution of many species. The genetic differentiation between populations of the same species and of sister species is often attributed largely to the more recent Pleistocene fluctuations. Recent interpretations, which implicate earlier episodes, have proved controversial. We address the timing of genetic divergence in the Iberian lizard *Lacerta schreiberi* by studying the phylogeography of the cytochrome *b* sequence. The species has a remarkable morphological uniformity, yet our evidence suggests that earlier events in the Pliocene initiated the main divergence between populations. This interpretation implies that the different populations survived through the Pleistocene in separate localities. This conclusion is robust to different molecular clock calibrations. The persistence of earlier differentiation through the Pleistocene has wide implications for our understanding of Pleistocene refugia in this species and, by extension, to the biogeography of the whole region.

Keywords: phylogeography; range expansion; mitochondrial DNA; Pleistocene glaciations; glacial refugia

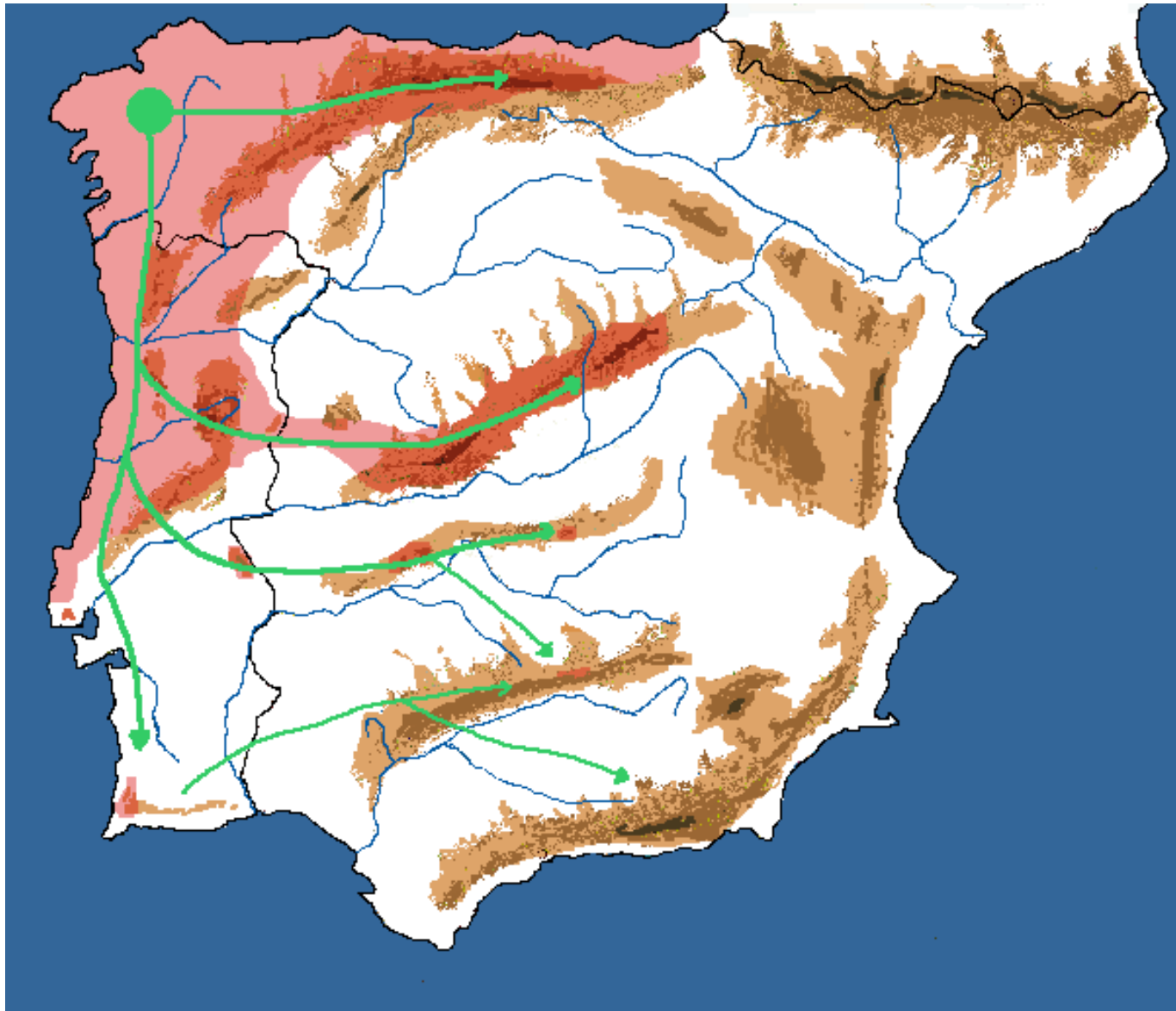


Filogeografia

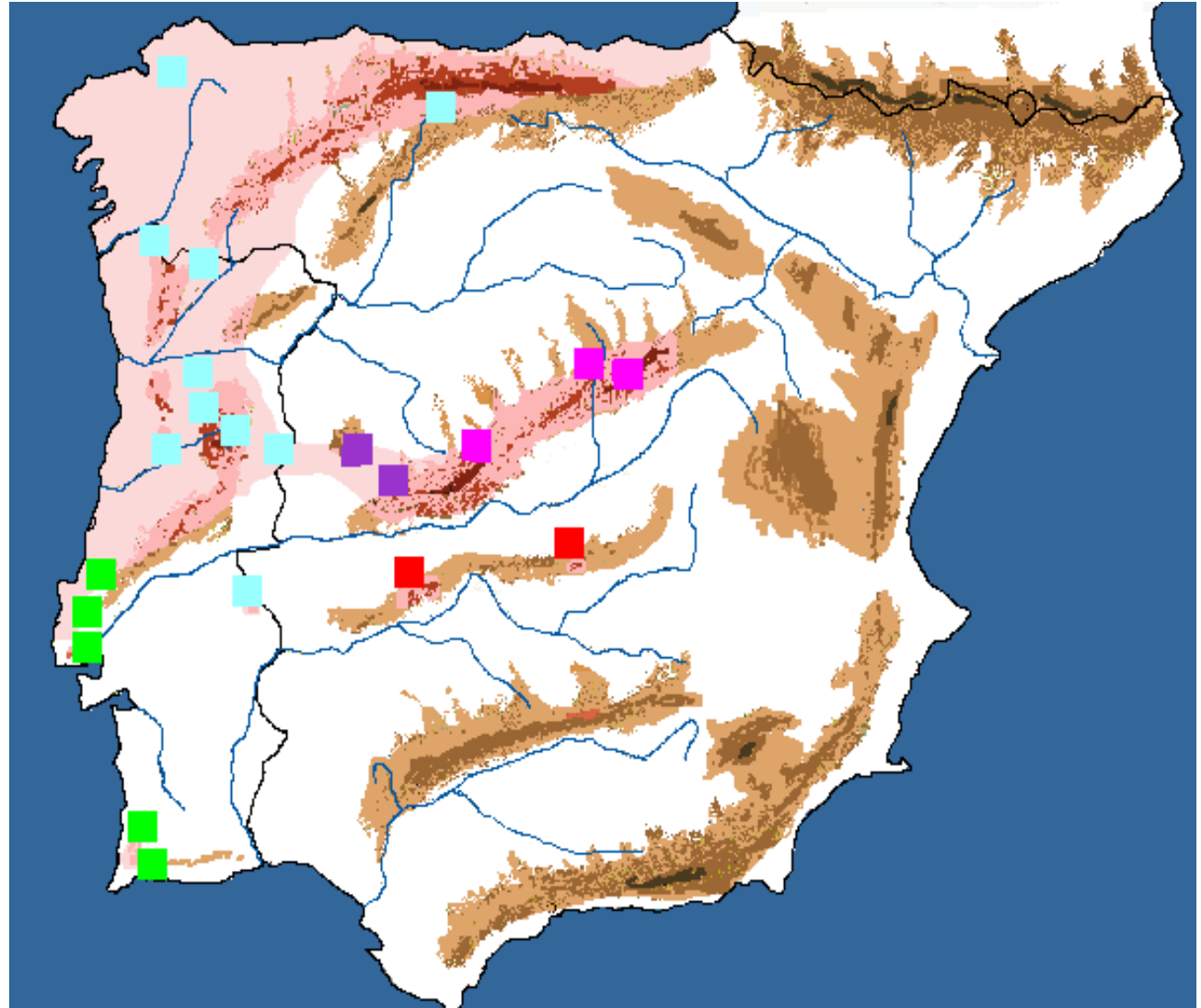
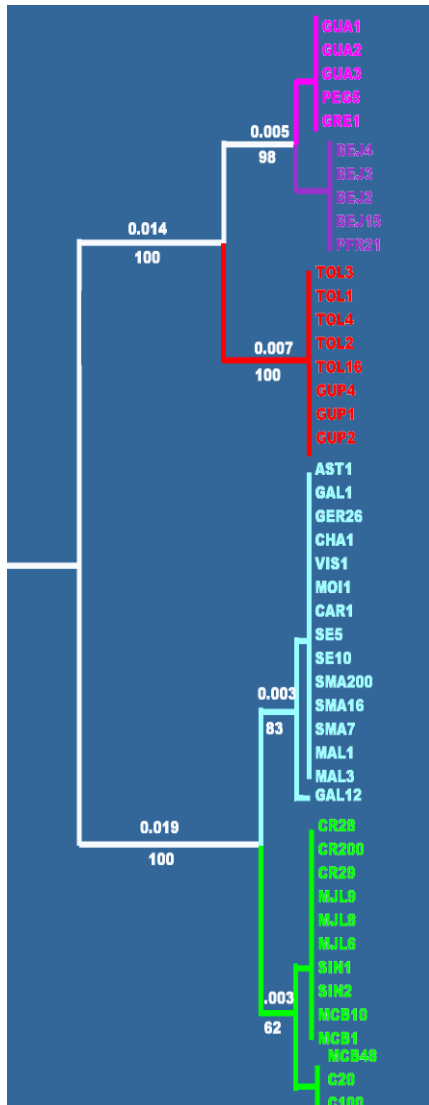
Distribution of *Lacerta schreiberi* (pink area)



Former hypothesis about the origin and postglacial expansions routes for *Lacerta schreiberi*

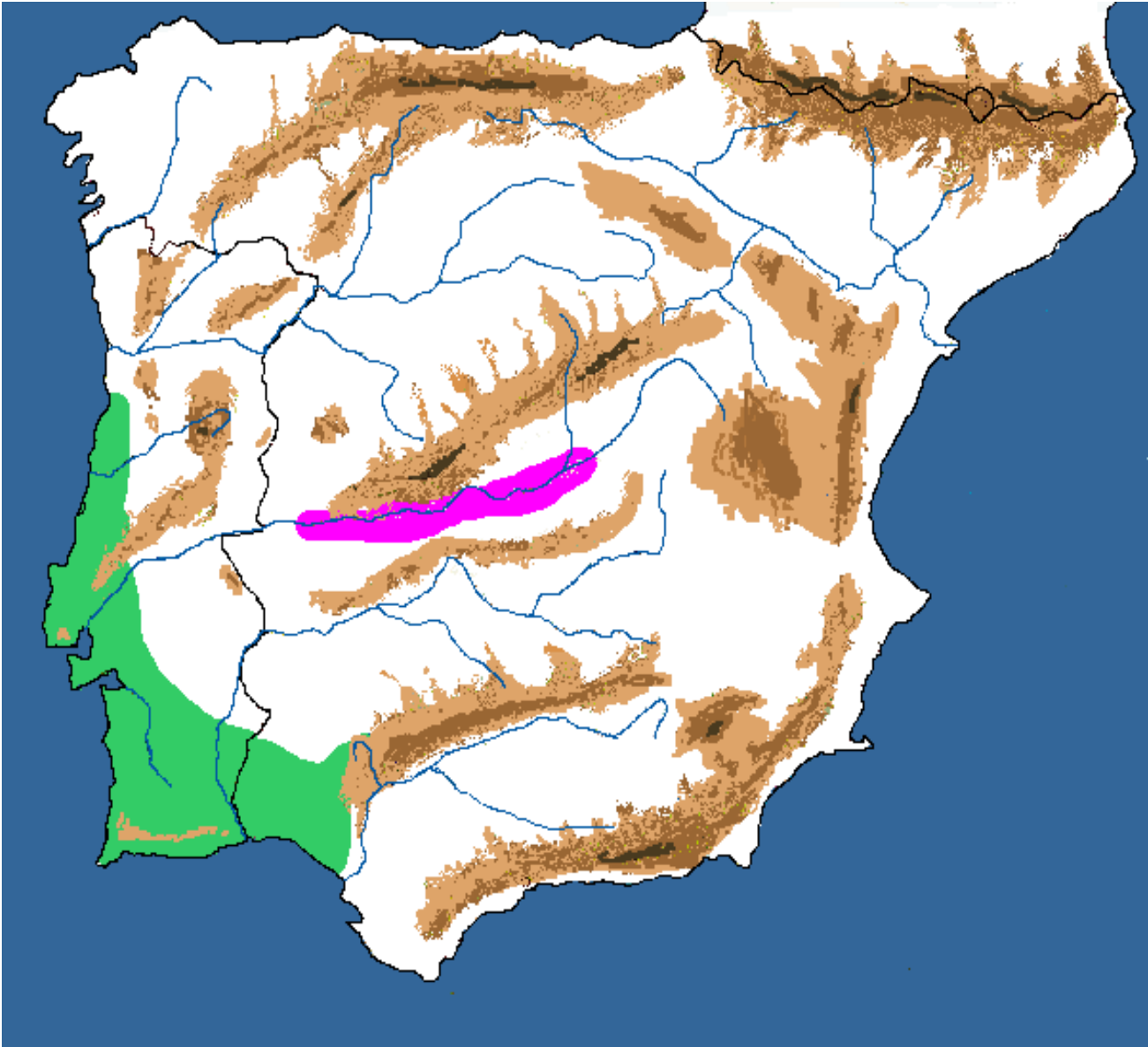
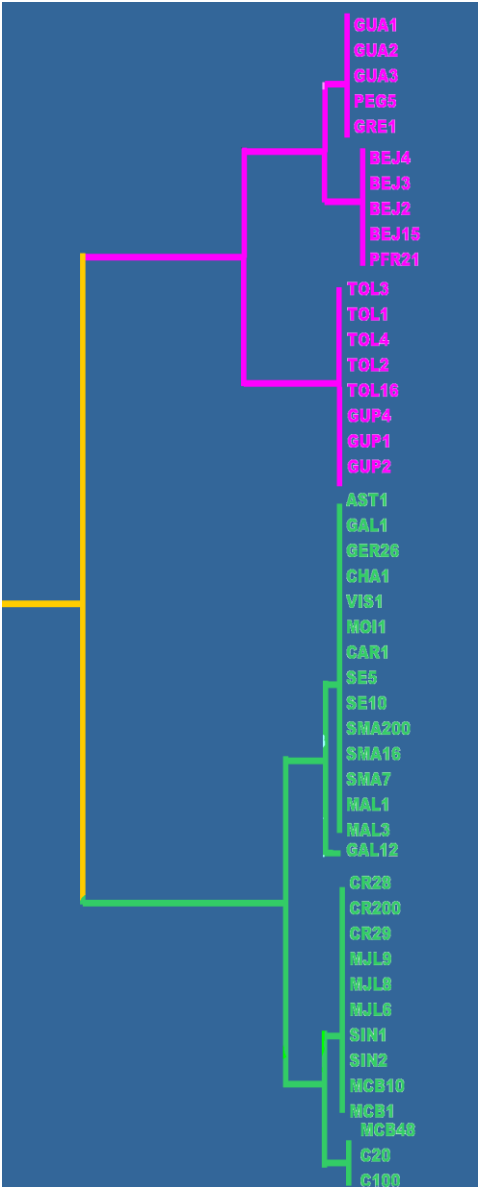


Phylogenetic tree of 663bp of cytochrome *b* of *L. schreiberi* samples and spatial pattern.



Origin of the two main clades of *Lacerta schreiberi*

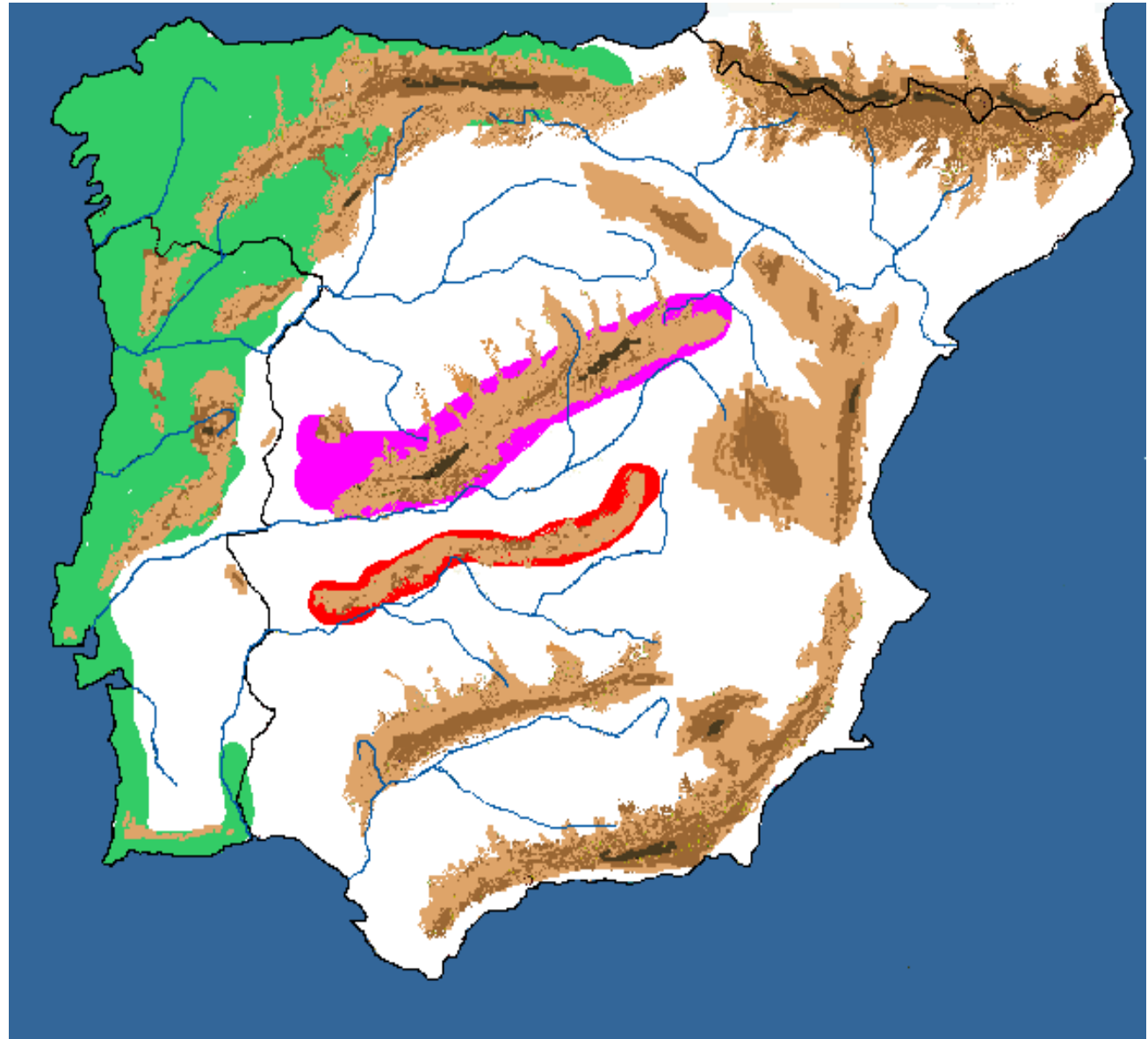
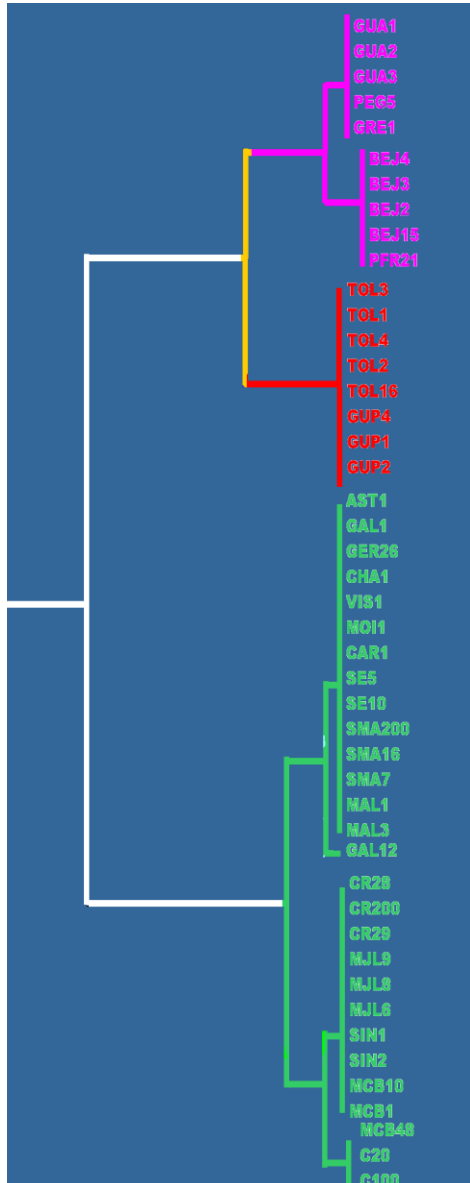
2.75 - 2.34 - 1.64 MYA



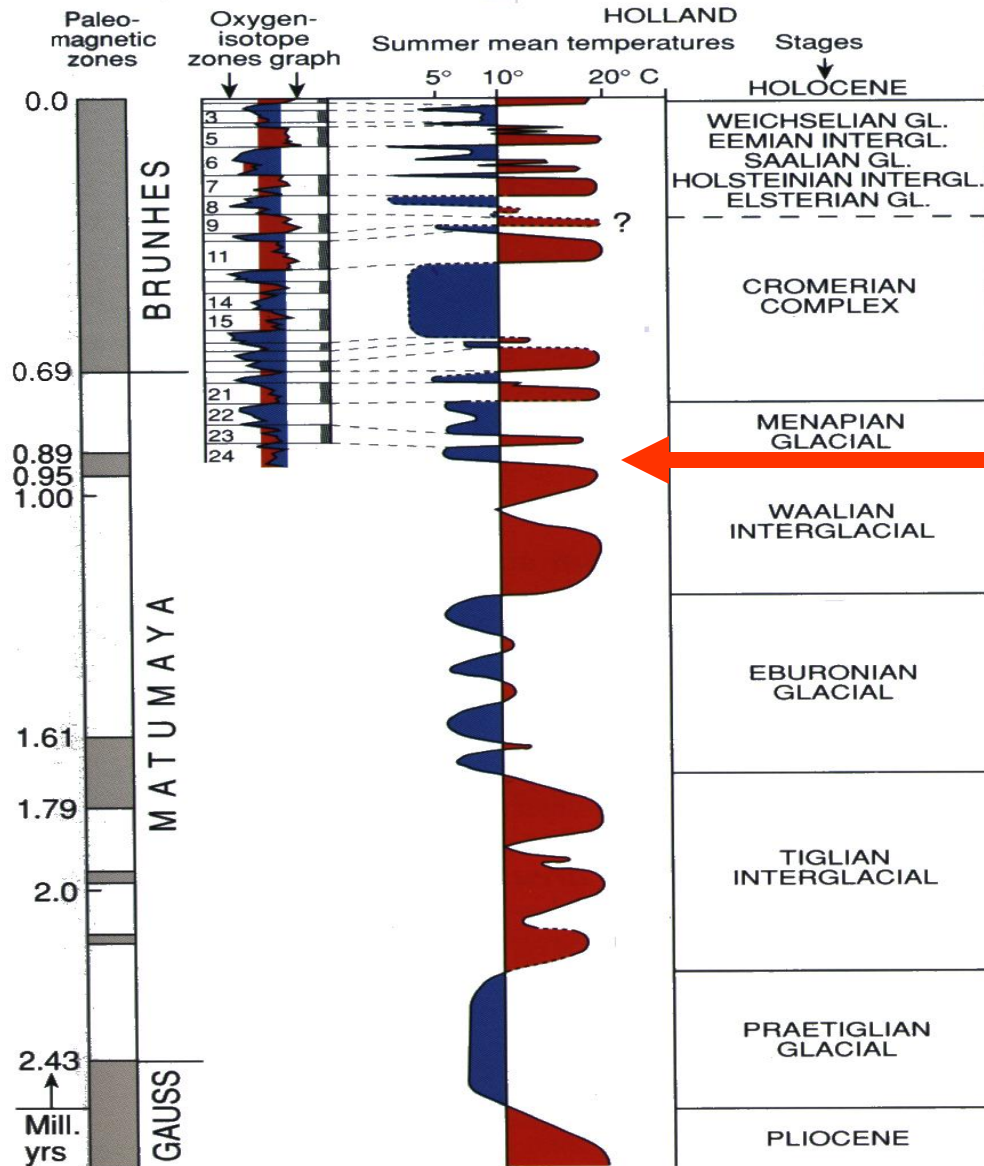
Filogeografia

Origin of the two inland subclades of *Lacerta schreiberi*

0.97 - 0.82 - 0.57 MYA



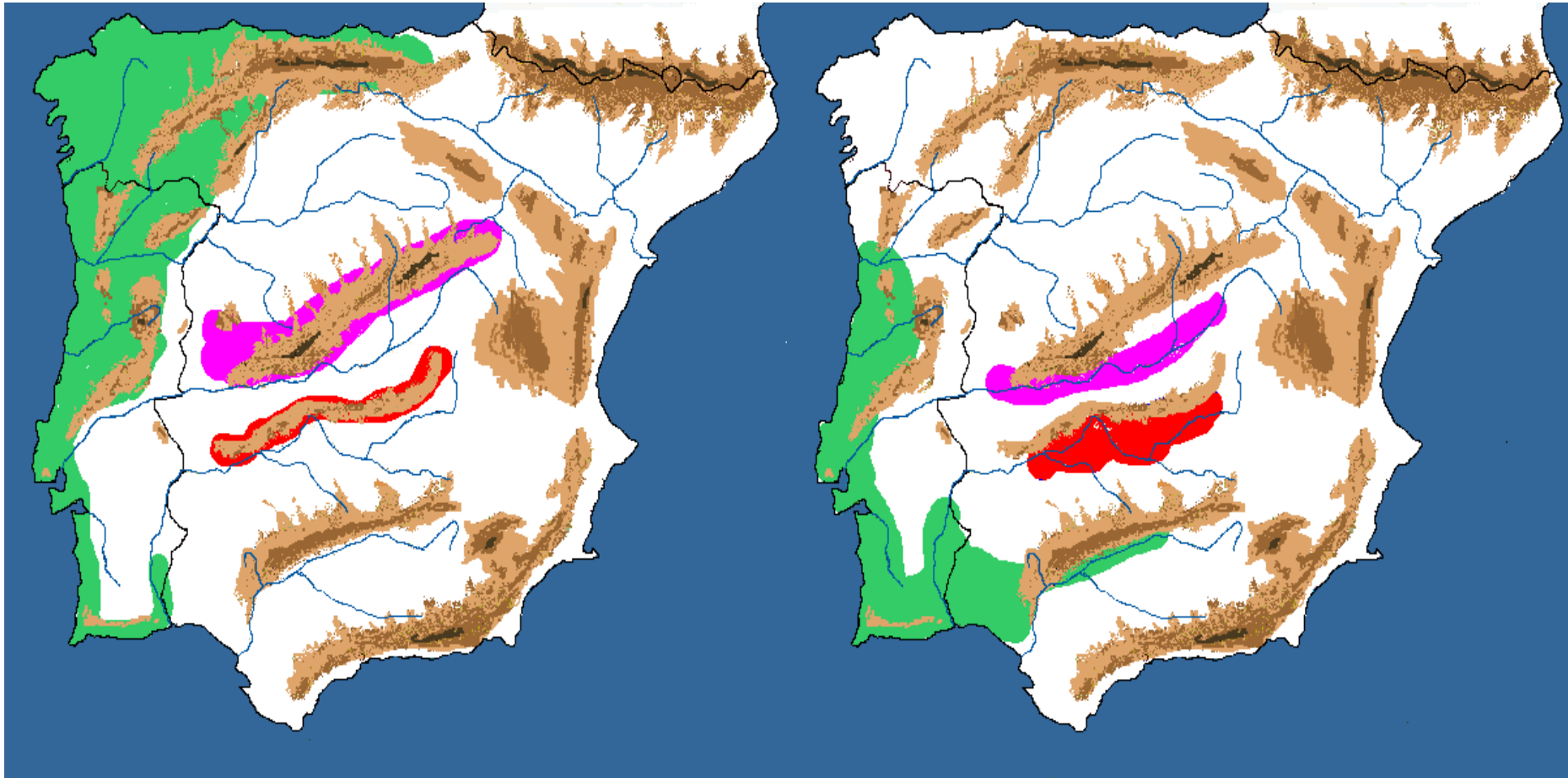
Pleistocene climatic conditions (data from Holland)



Hypothetical distribution of *Lacerta schreiberi* during the glacial and interglacial phases

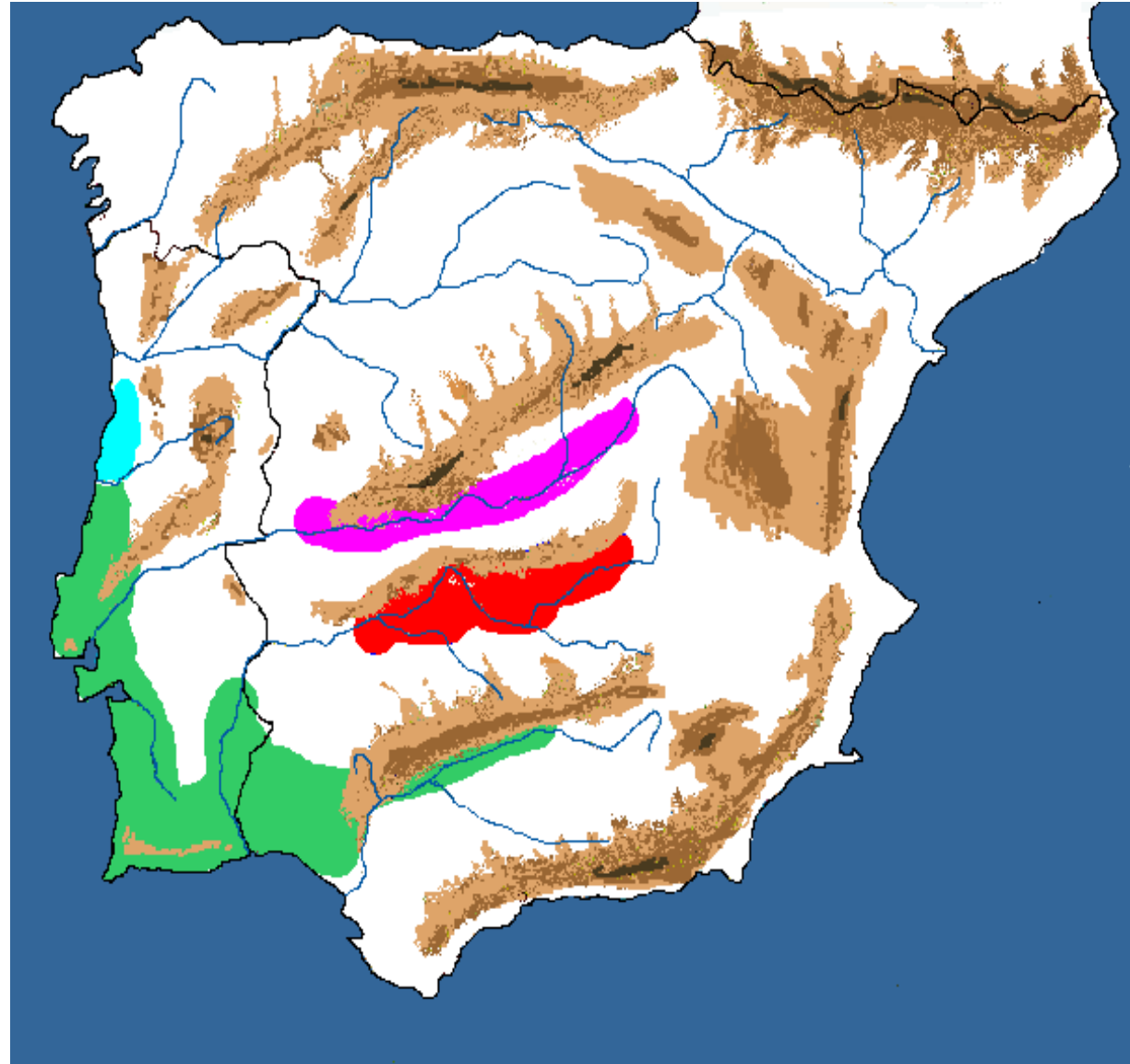
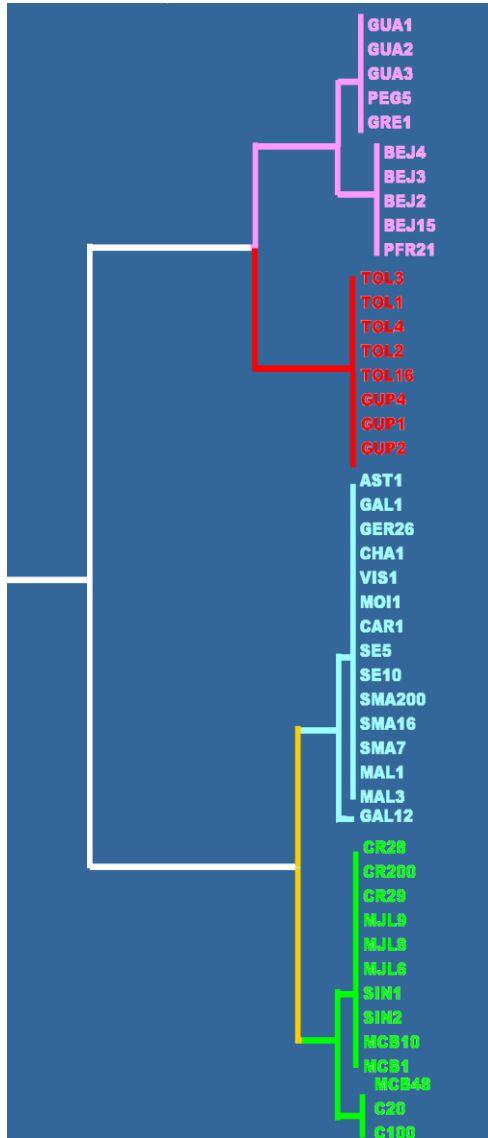
Interglacial

Glacial

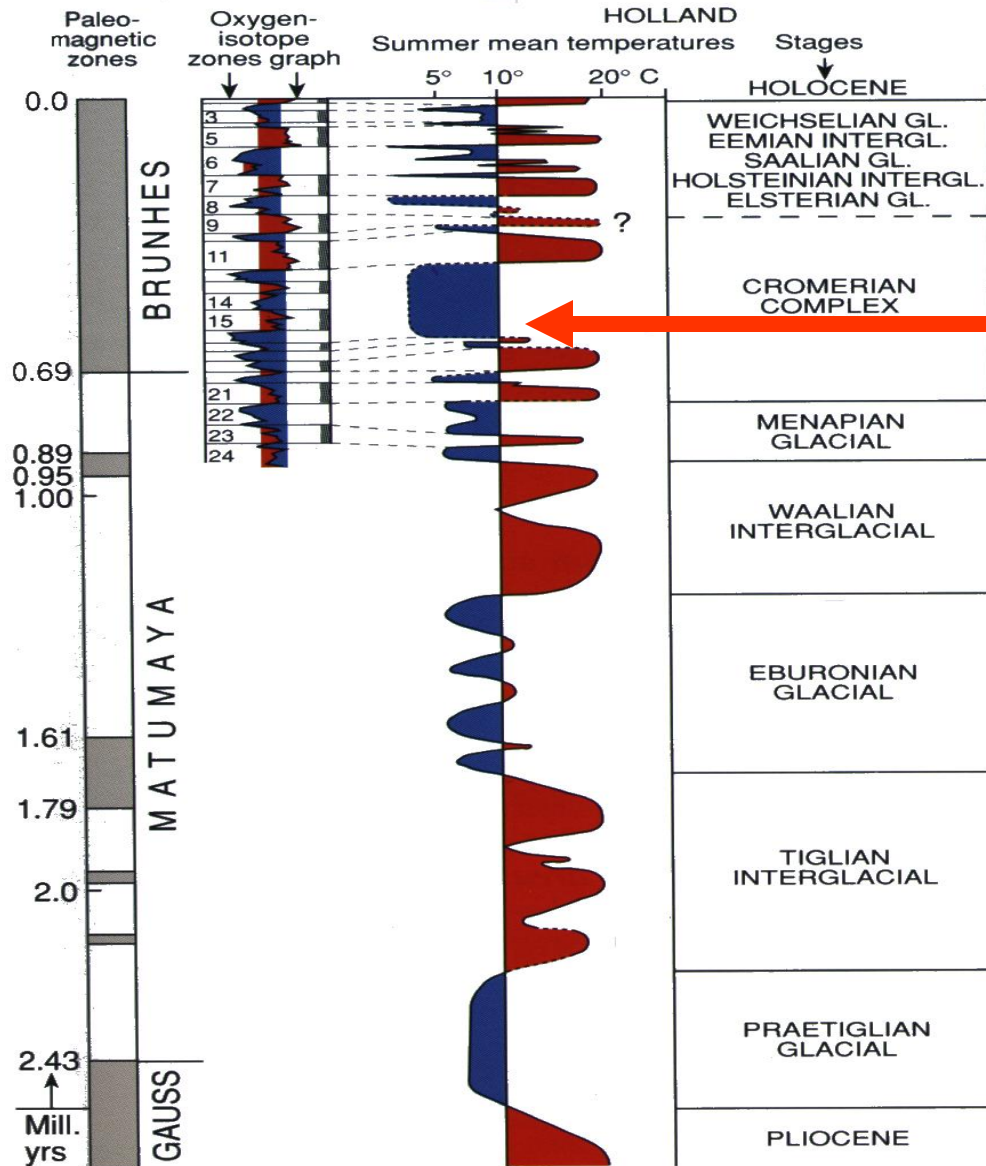


Origin of the two coastal subclades of *Lacerta schreiberi*

0.62 - 0.53 - 0.37 MYA

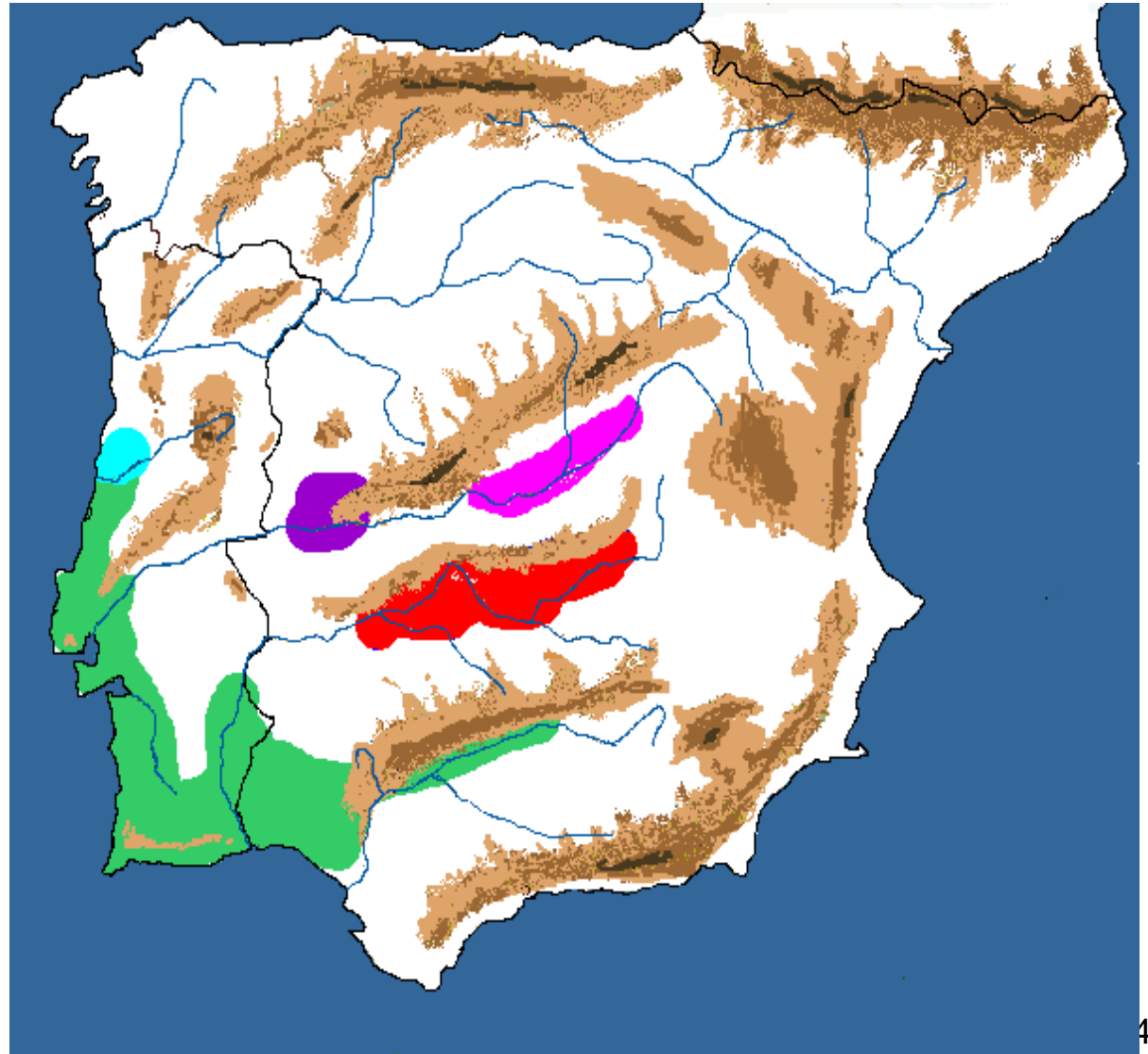
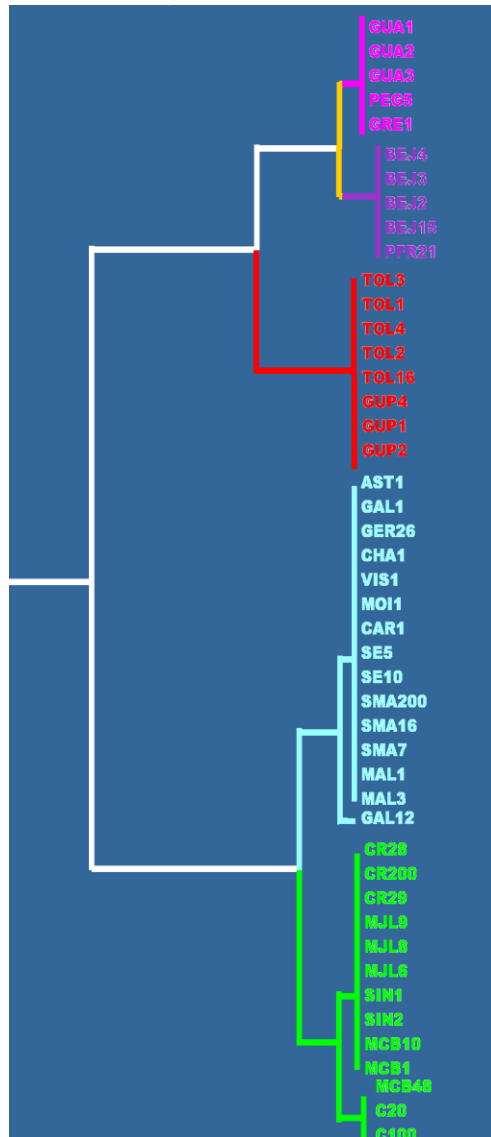


Pleistocene climatic conditions (data from Holland)



Split of the North inland subclade of *L. schreiberi*

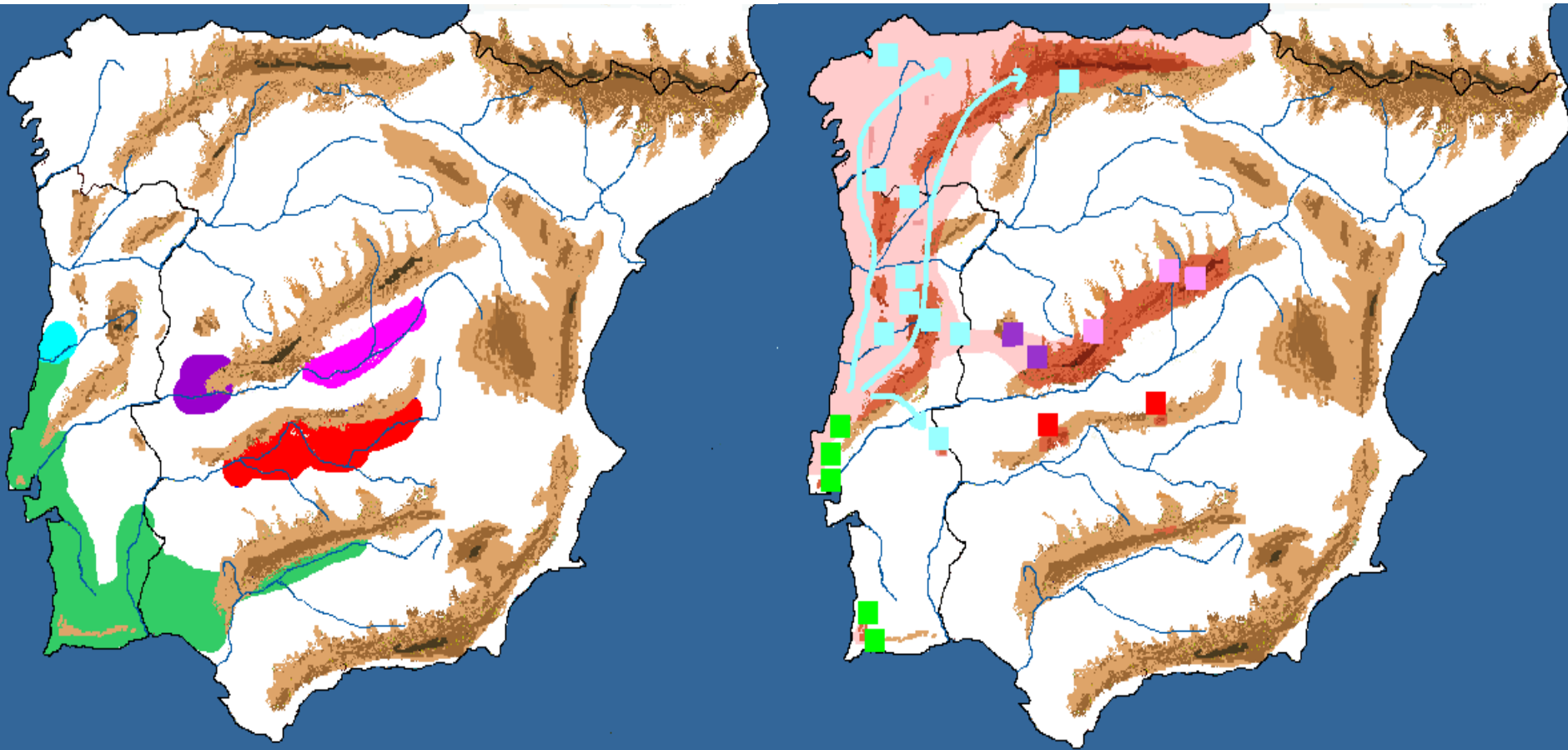
0.39 - 0.28 - 0.19 MYA



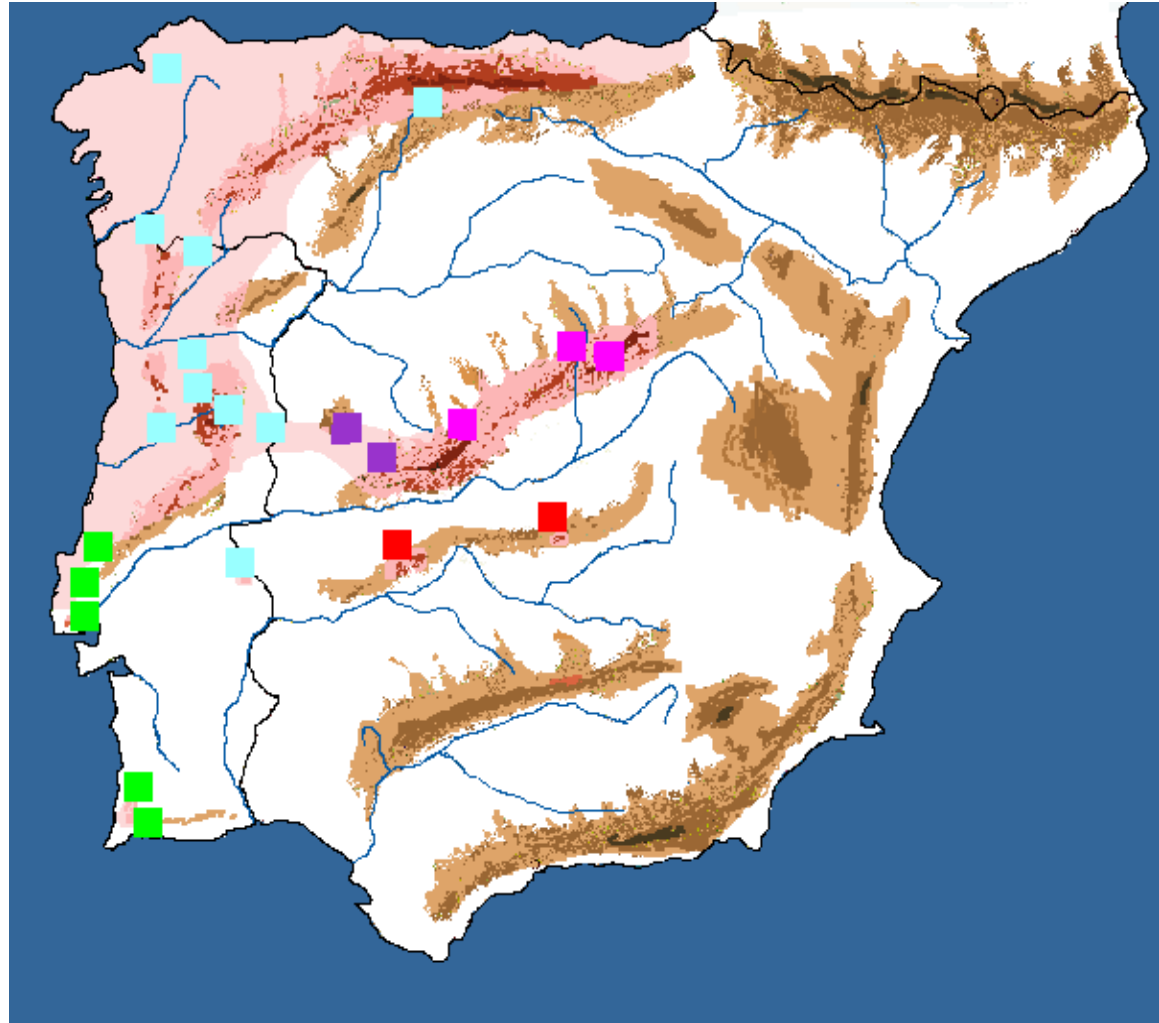
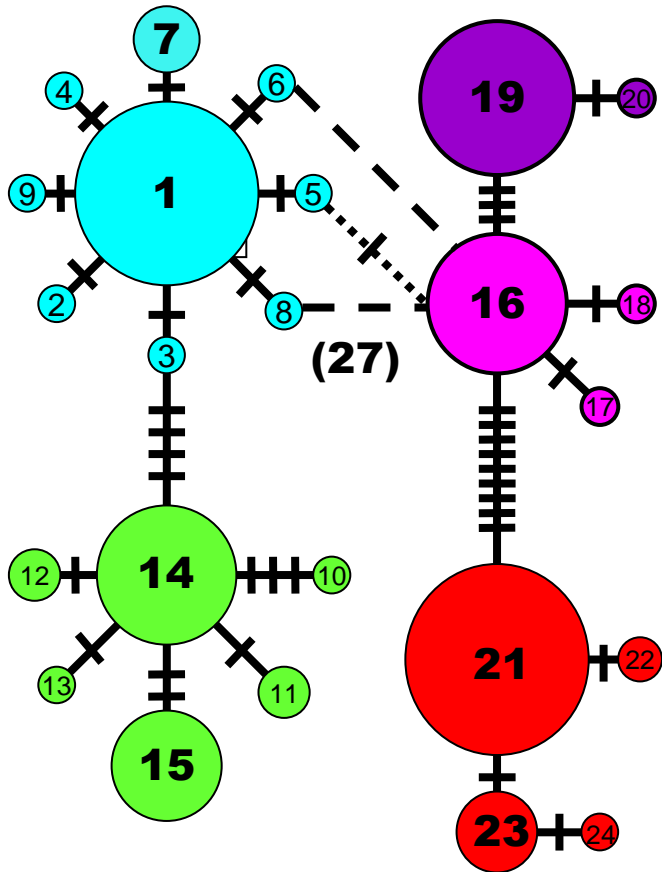
Filogeografia

Expansion of the coastal subclades of *Lacerta schreiberi*

<0.2 MYA



Minimum spanning network of the 24 haplotypes found in the populations of *L. schreiberi*



Using nested clade analysis to assess the history of colonization and the persistence of populations of an Iberian Lizard

OCTÁVIO S. PAULO,* WILLIAM C. JORDAN,† MICHAEL W. BRUFORD‡ and RICHARD A. NICHOLS§

*Centro de Biologia Ambiental/Departamento de Zoologia e Antropologia, Faculdade de Ciências da Universidade de Lisboa, P- 1749–016 Lisboa, Portugal, †Institute of Zoology, Zoological Society of London, London NW1 4RY, UK, ‡Cardiff School of Biosciences, Cardiff University, Cathays Park, Cardiff CF10 3TL, UK, §School of Biological Sciences, Queen Mary and Westfield College, University of London, London E1 4NS, UK

Abstract

The distribution of the lizard *Lacerta schreiberi* is likely to have been severely affected by the climatic cycles that have influenced the Iberian Peninsula. Information about the species ecology and Iberian physiogeography was used to generate specific hypotheses about episodes of colonization and subsequent population persistence. These hypotheses generated predictions about the distribution of genetic variation, which were tested using nested clade analysis (NCA) supplemented by analysis of molecular variance (amova). Two predictions were confirmed by NCA; that is those that specified multiple and allopatric refugia. However, the remaining three predictions were not corroborated by the analyses. Firstly, a simple analysis of the distribution of genetic variability failed to detect an expected difference in the pattern of colonization between the inland mountain system and the coastal region. Moreover, while NCA did detect the expected genetic pattern in southern coastal populations, it was explained in terms of long-distance migration, which seems implausible because of the extent of unsuitable habitat. A more likely cause of the pattern is population fragmentation and a reduction in population size caused during the Holocene. Finally, NCA also failed to detect a northwestern population expansion, which is supported by other evidence. We conclude that NCA has a limited ability to detect range expansion led by individuals with more ancestral (interior) haplotypes.

Keywords: Iberian Peninsula, *Lacerta schreiberi*, phylogeography, Pleistocene, Pliocene, range expansion

RESEARCH ARTICLE

Open Access

Phylogeography and demographic history of *Lacerta lepida* in the Iberian Peninsula: multiple refugia, range expansions and secondary contact zones

Andreia Miraldo^{1,3*}, Godfrey M Hewitt¹, Octavio S Paulo² and Brent C Emerson^{1,4}



Octávio S. Paulo – FEM

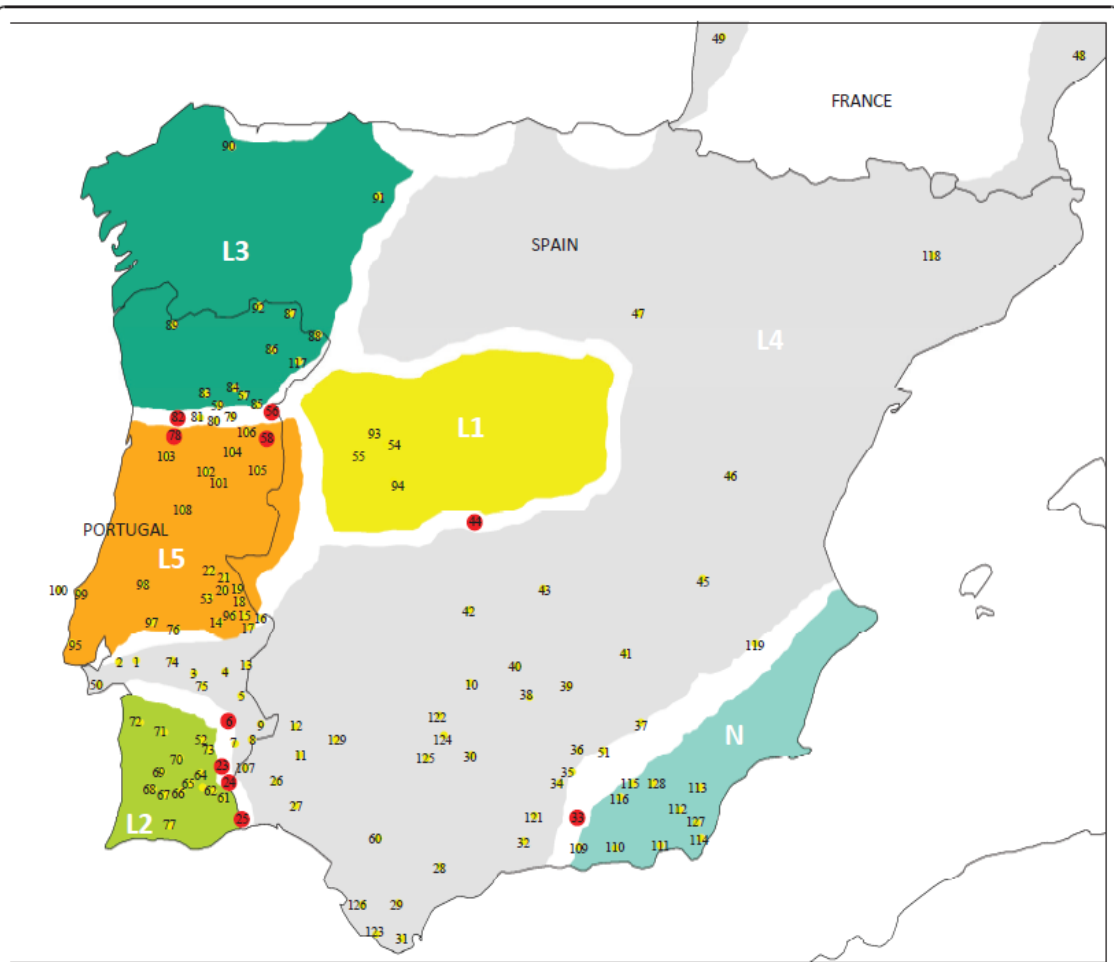
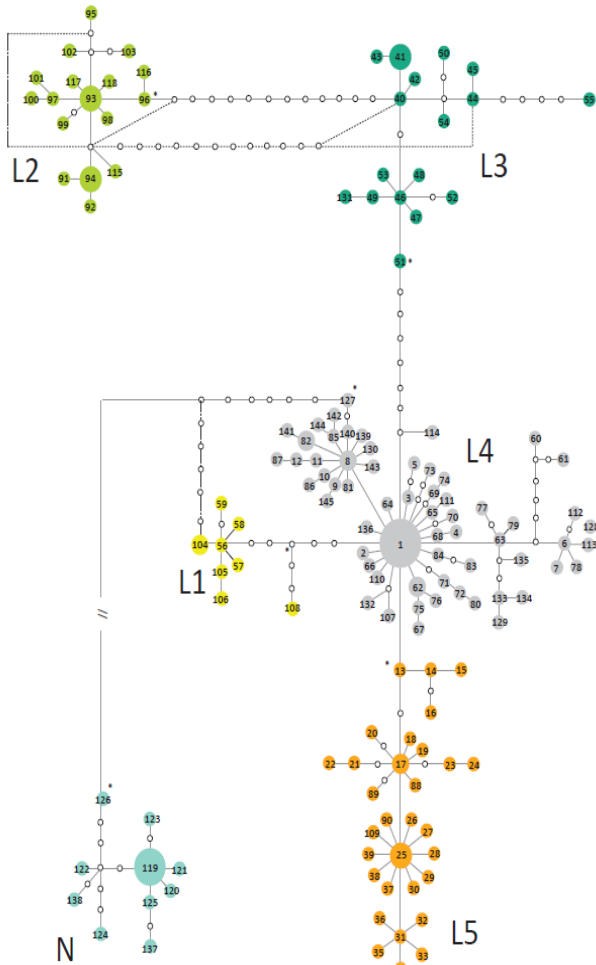


Figure 4 Distribution of *Lacerta lepida* mitochondrial phylogroups based on 627 bp of the cytochrome b gene. Colours are the same as in Figures 2 and 3. Filled red circles represent populations where divergent haplotypes from two or more phylogroups were detected in sympatry. Numbers correspond to sampling localities as in Figure 1 and Additional file 1, Table S1.

Figure 3 Statistical Parsimony network of *Lacerta lepida* cytochrome b haplotypes from 312 samples. Open circles with no numbers represent unsampled or extinct haplotypes. L1, L2, L3, L4, L5 and N represent different mitochondrial phylogroups. The ancestral haplotype within each phylogroup is marked with an asterisk. Phylogroup N connects to the main network through 65 mutations, represented by an interrupted line.

Coalesce, mutate or recombine

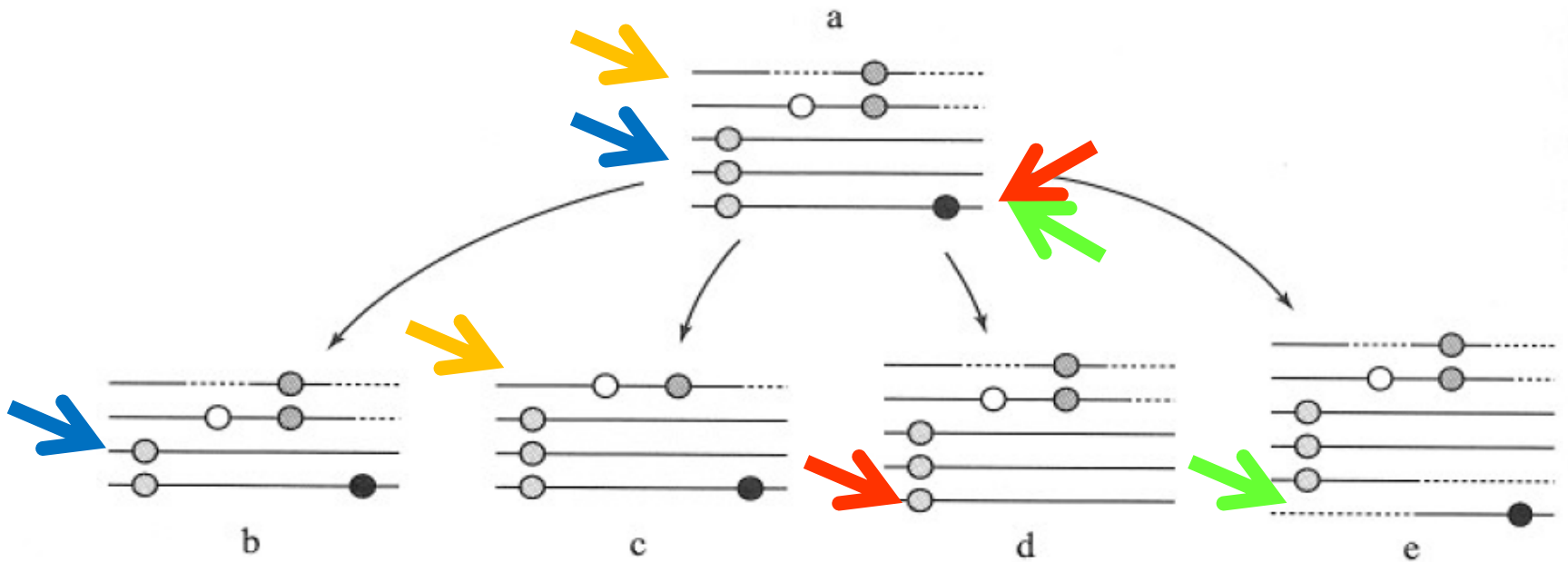


Figure 5.21 The present configuration (top of figure) can be created by four classes of events. Two identical sequences could coalesce, two sequences identical in common ancestral material could coalesce, a sequence can mutate and lastly a sequence could have arisen by recombination. This is here illustrated by taking one representative of the four