

Spathe or leaf-like bract



Spadix

Amorphophallus titanum

Largest unbranched inflorescence in the world

Monecious and protogynous

Carrion flower (fly/beetle pollinated)

Indigenous
to the forests

of Sumatra



Questions

- How is genetic variation distributed in time and space?
- What factors account for the spatial and temporal distribution of genetic lineages?

Phylogeography: Outline

1. What is phylogeography?
2. The benefits of phylogeography
 - i.e. understanding genetic structure
3. How do we do it?
 - From historical to current approaches

Phylogeography: Introduction

What is phylogeography?

“A field of study concerned with the principles and processes governing the geographic distribution of genealogical lineages, especially those within and among closely related species.”

Avise 2000

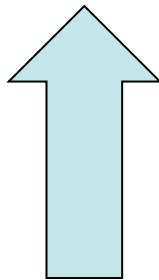
Phylogeography: Introduction

Population genetics

- Within populations
- Shallow timescale

Phylogenetics

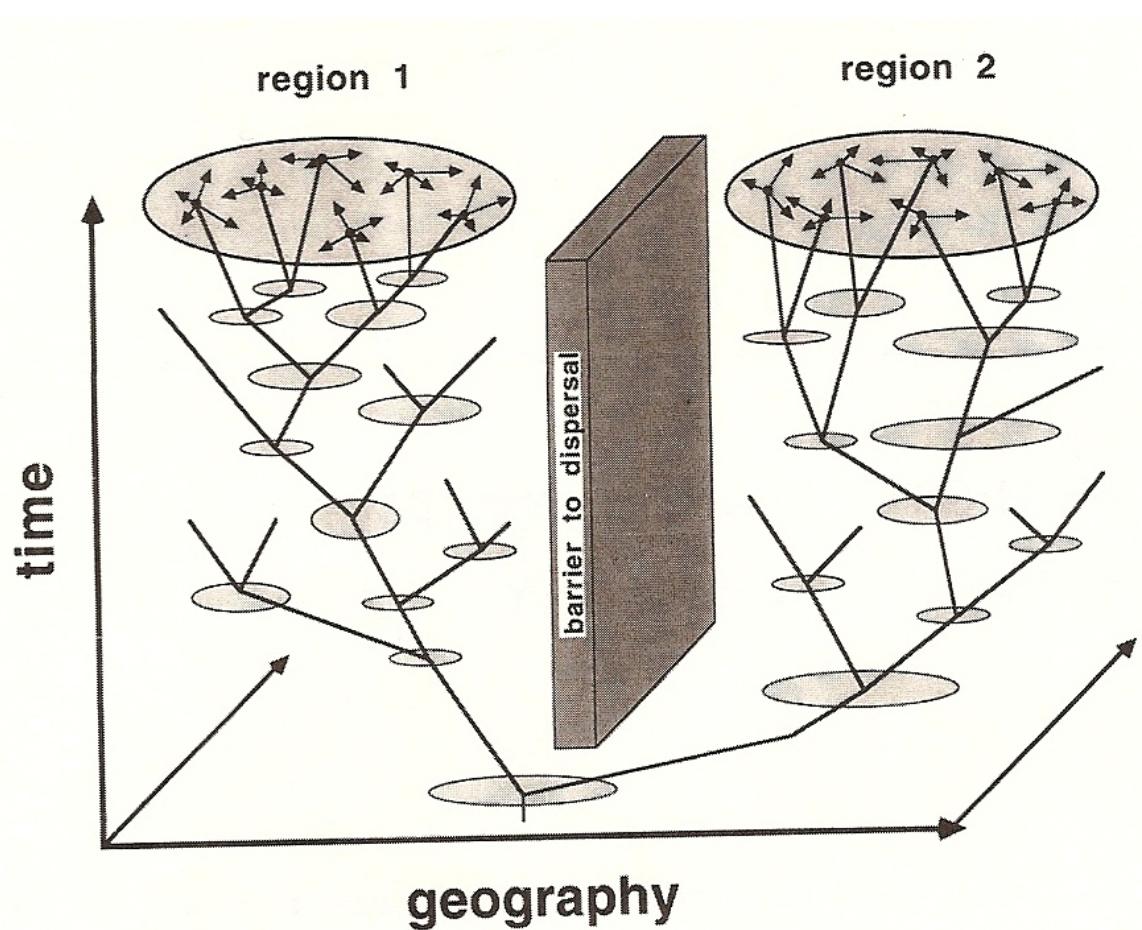
- Between species
- Deep timescale



Phylogeography

- Between populations and species
 - Medium timescale
 - Plus geography

Phylogeography: Introduction



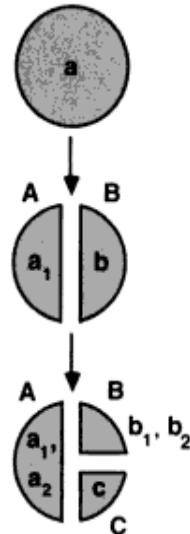
Gene
genealogies of
interest are
mapped in
space and time.

Phylogeography: Introduction

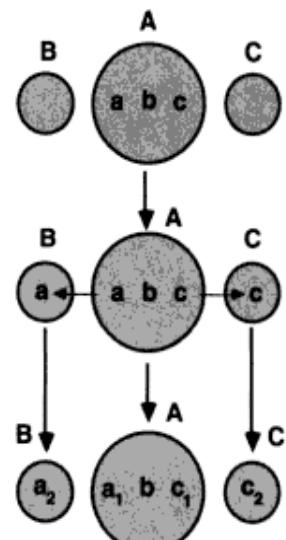
Goal: To understand the factors contributing to the formation of population (or species-level) genetic structure.

Can evaluate alternative historical scenarios that account for current spatial patterns.

vicariance



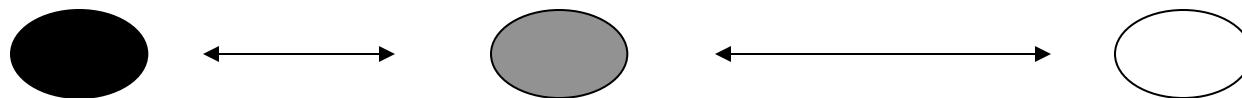
dispersal



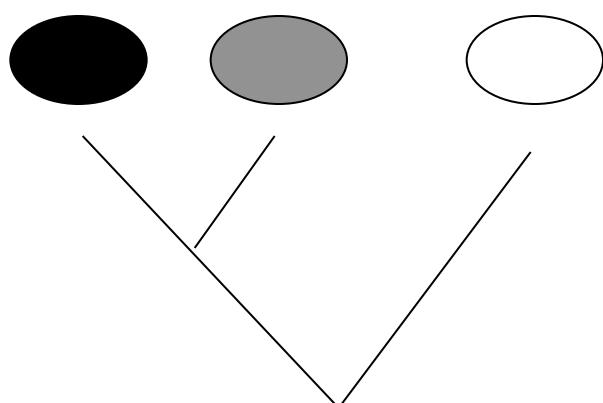
Determinants of Genetic Structure

Historical relationships and contemporary gene flow influence spatial patterns of genetic variation.

Similarity due to gene flow



Similarity due to historical relationships



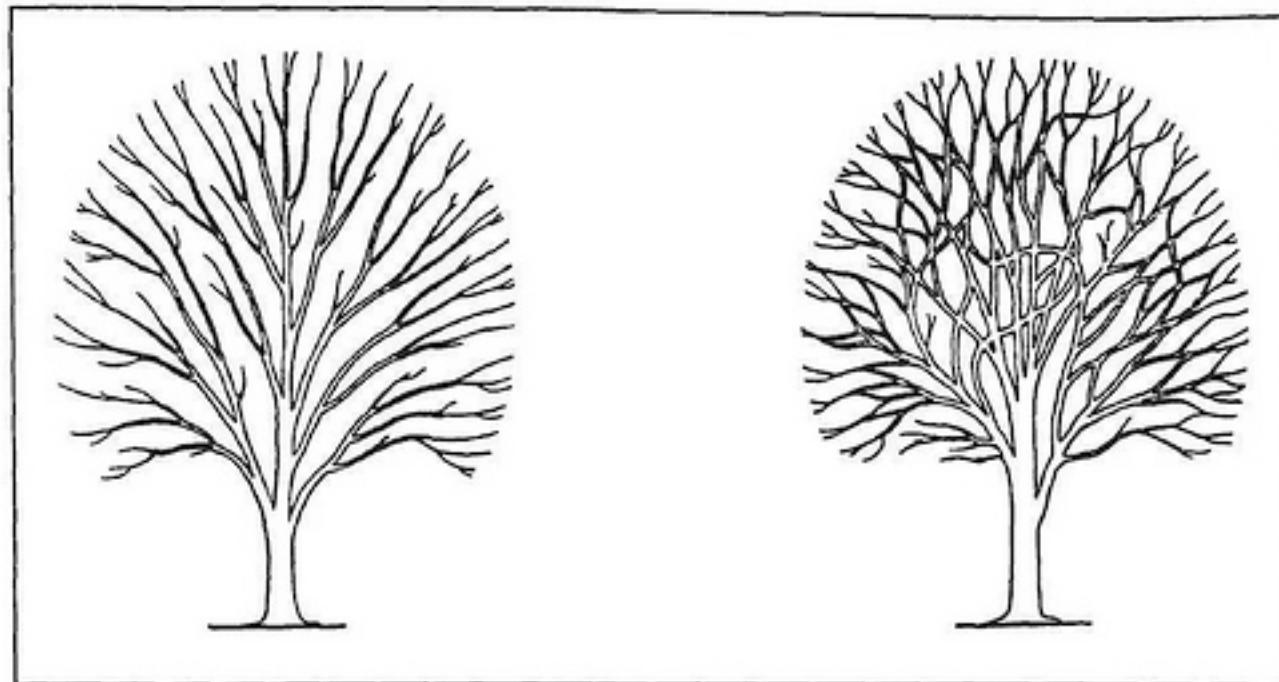
Understanding Genetic Structure

- Traditionally:
 - **Within species**, *genetic exchange* has been emphasized as the cause of similarity.
 - *Caveat*: when estimating Nm , we assume an equilibrium between genetic drift and gene flow.
 - **Among species**, *historical relationship* has been emphasized as the cause of similarity.
 - *Caveat*: cladistics assumes non-reticulation.

What is Reticulation?

Plant species boundaries are often porous (interspecific hybridization and introgression is common).

Populations form reticulating lineages (i.e. networks)



Understanding Genetic Structure

“**Phylogeographic methods** provide a means of examining the history of genetic exchange among populations, with the potential to distinguish biogeographic patterns of genetic variation caused by gene flow from those caused by common ancestry.” - Schaal 1998

We can tease apart *contemporary* forces of genetic exchange from *historical* relationships

Understanding Genetic Structure

Phylogeography uses principles from **population genetics** and **phylogenetics**.

However...

- 1) Does not assume an equilibrium between drift and gene flow
- 2) Does not assume non reticulation

Relying on the use of **gene genealogies**, from non-recombining segments of DNA (e.g. mtDNA, cpDNA), phylogeographic methods provide historical information that individuals or populations cannot.

Genetic variation

We need significant genetic variation at the appropriate level (i.e. among the populations or taxonomic units under investigation)

Mitochondrial DNA

- great for animals
- low rates of molecular evolution in plants, intramolecular recombination

Chloroplast DNA

- faster rates of molecular evolution but still low (and variable among species)
- no recombination
- chloroplast capture can occur (e.g. in oaks, sunflowers, poplar)

Organelle genomes are essentially a single locus - inferences based on multiple loci give more accurate estimations of the population history of a species

Nuclear DNA

- problems of intra-allelic recombination, heterozygosity and gene families

Phylogeography: the early years

Phylogeography was essentially *descriptive*:

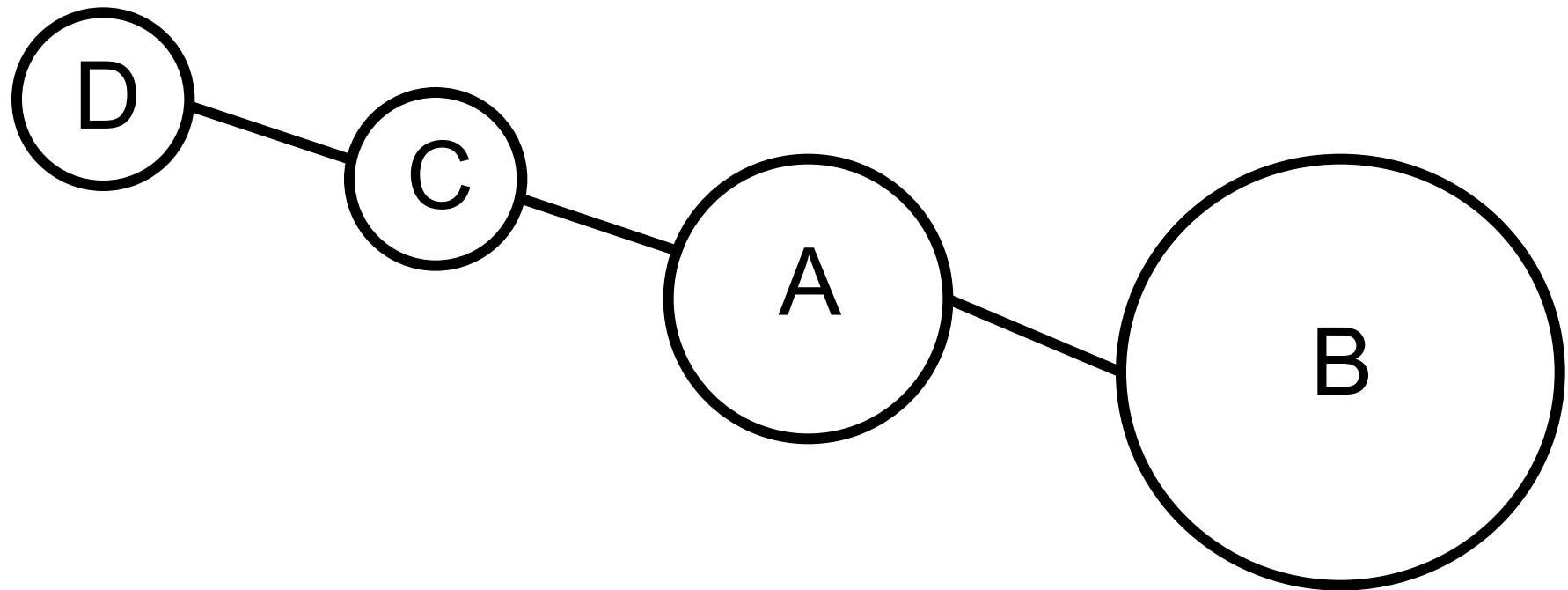
Plot haplotypes on map

Classify pattern of phylogeographic structure

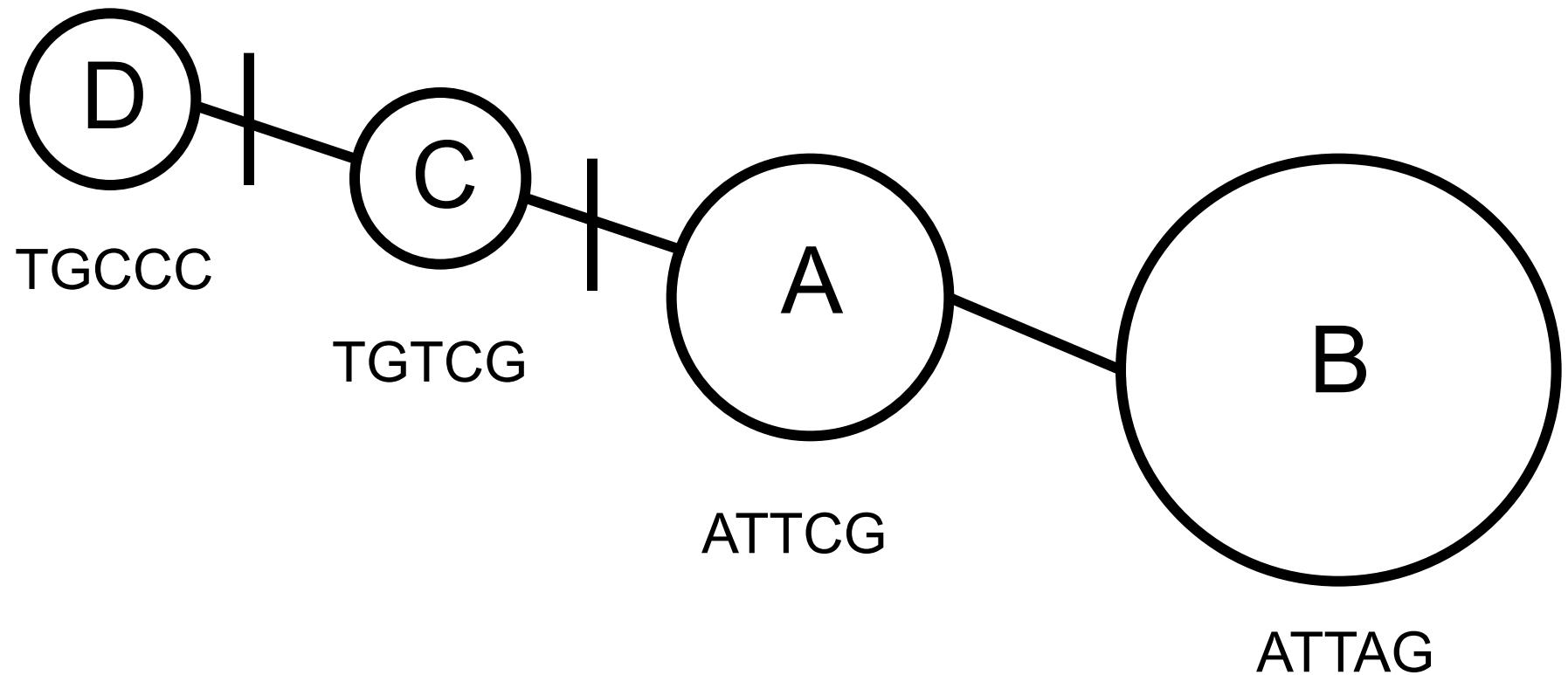
Consider historical explanations

Look for concordance among different species (comparative phylogeography)

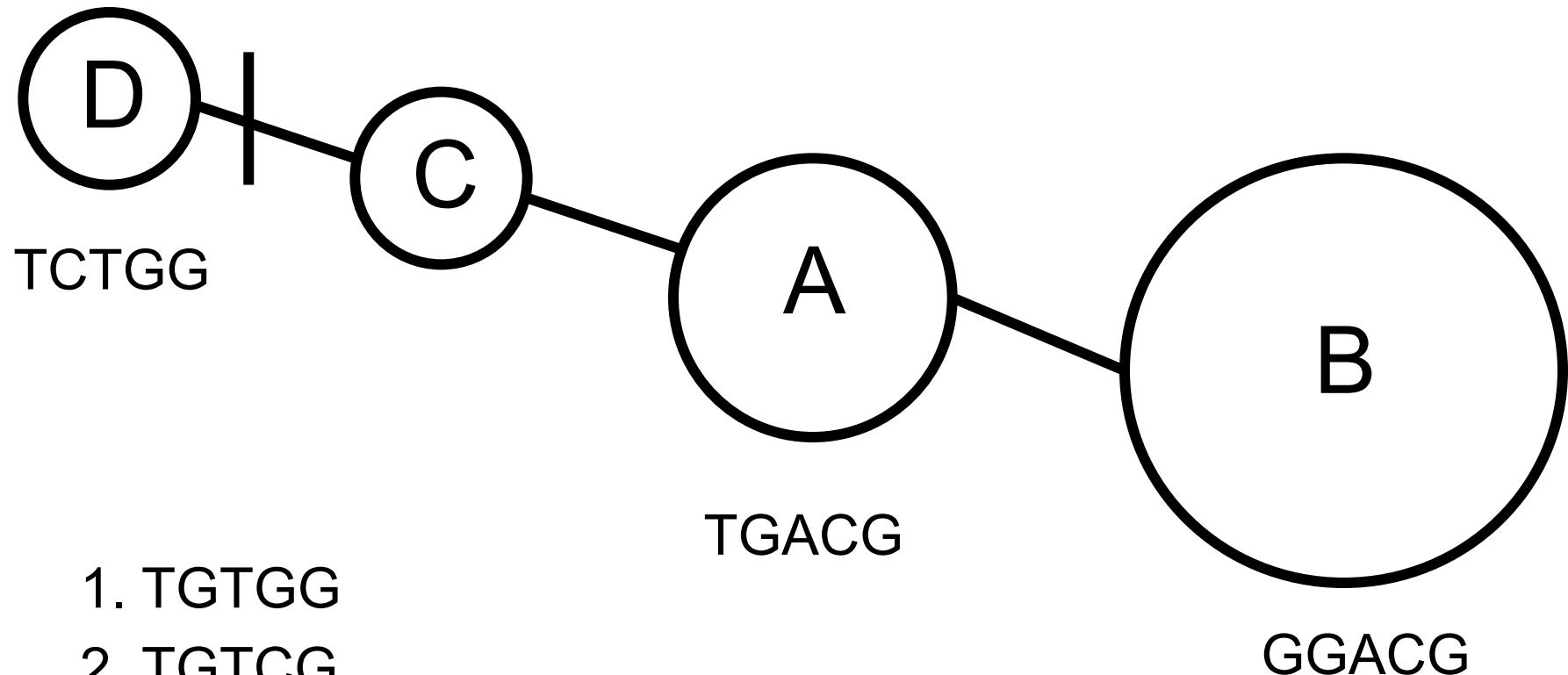
Haplotype network



Haplotype network

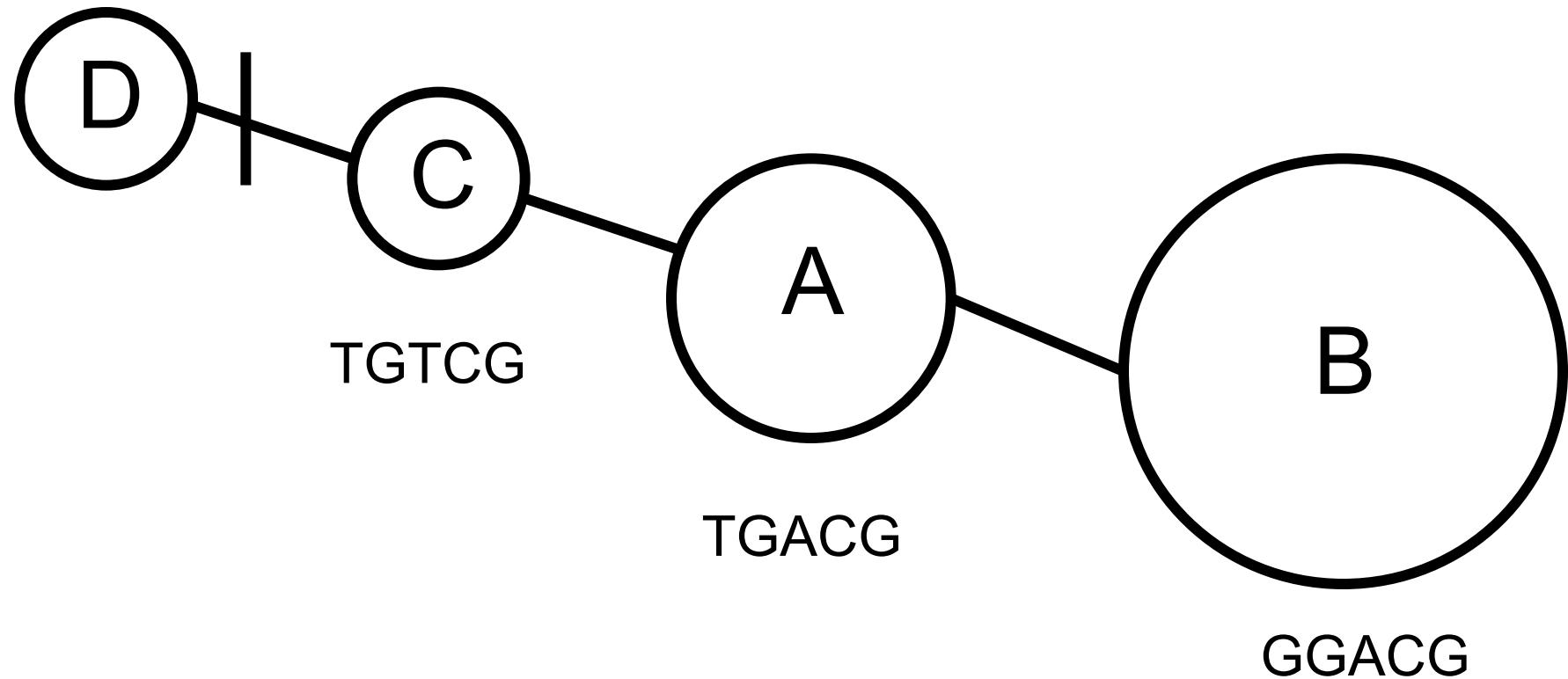


What is the C haplotype?

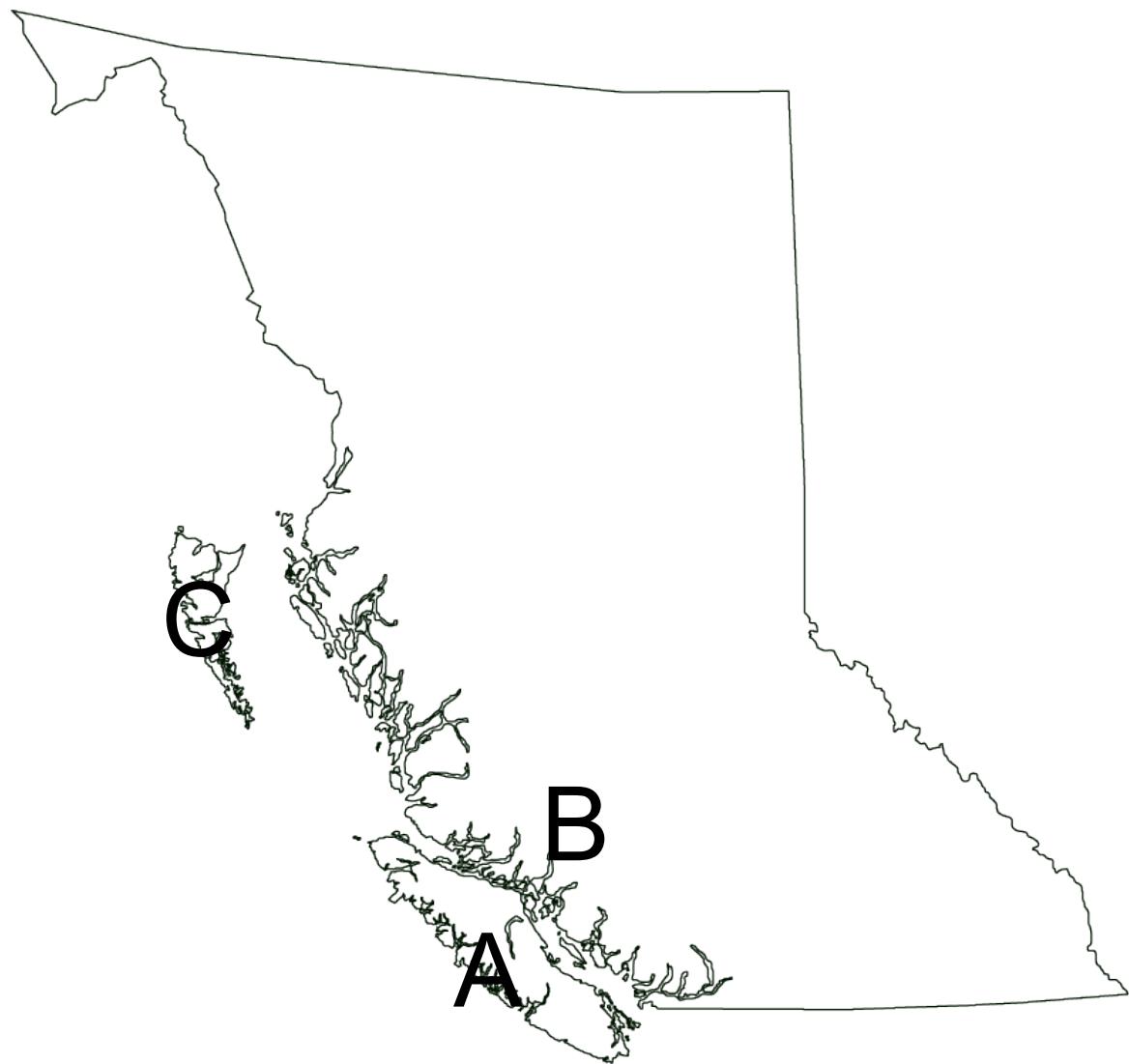
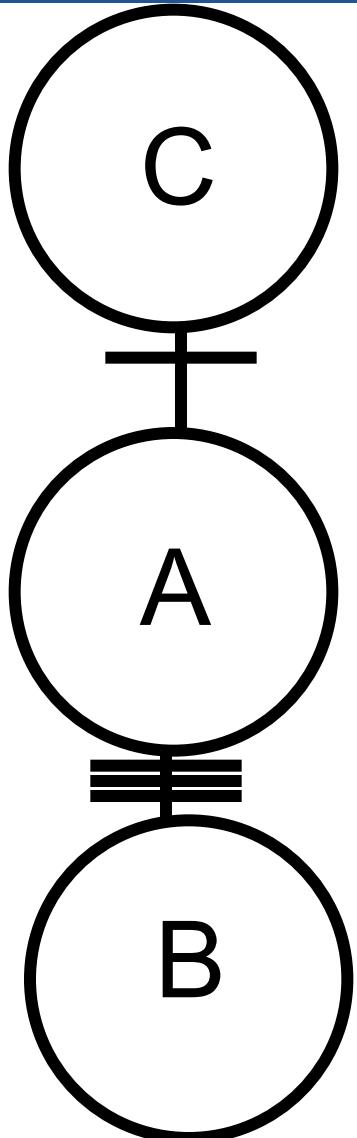


1. TGTGG
2. TGTCG
3. TCACG
4. GGTGG

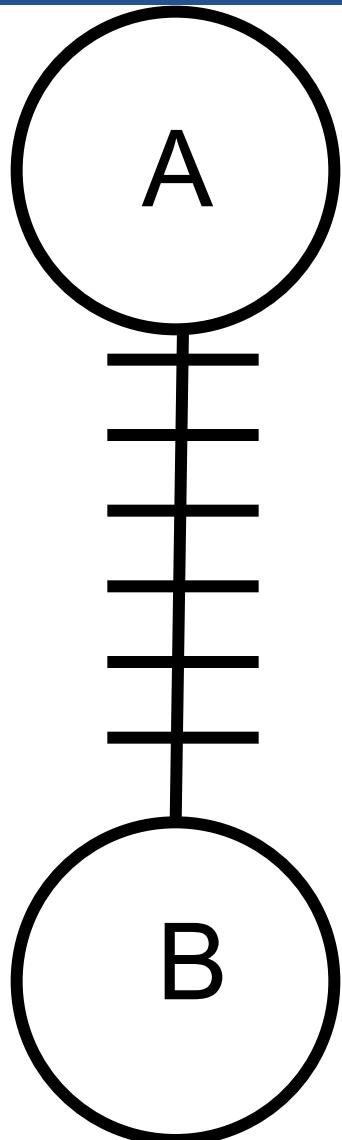
What is the D haplotype?



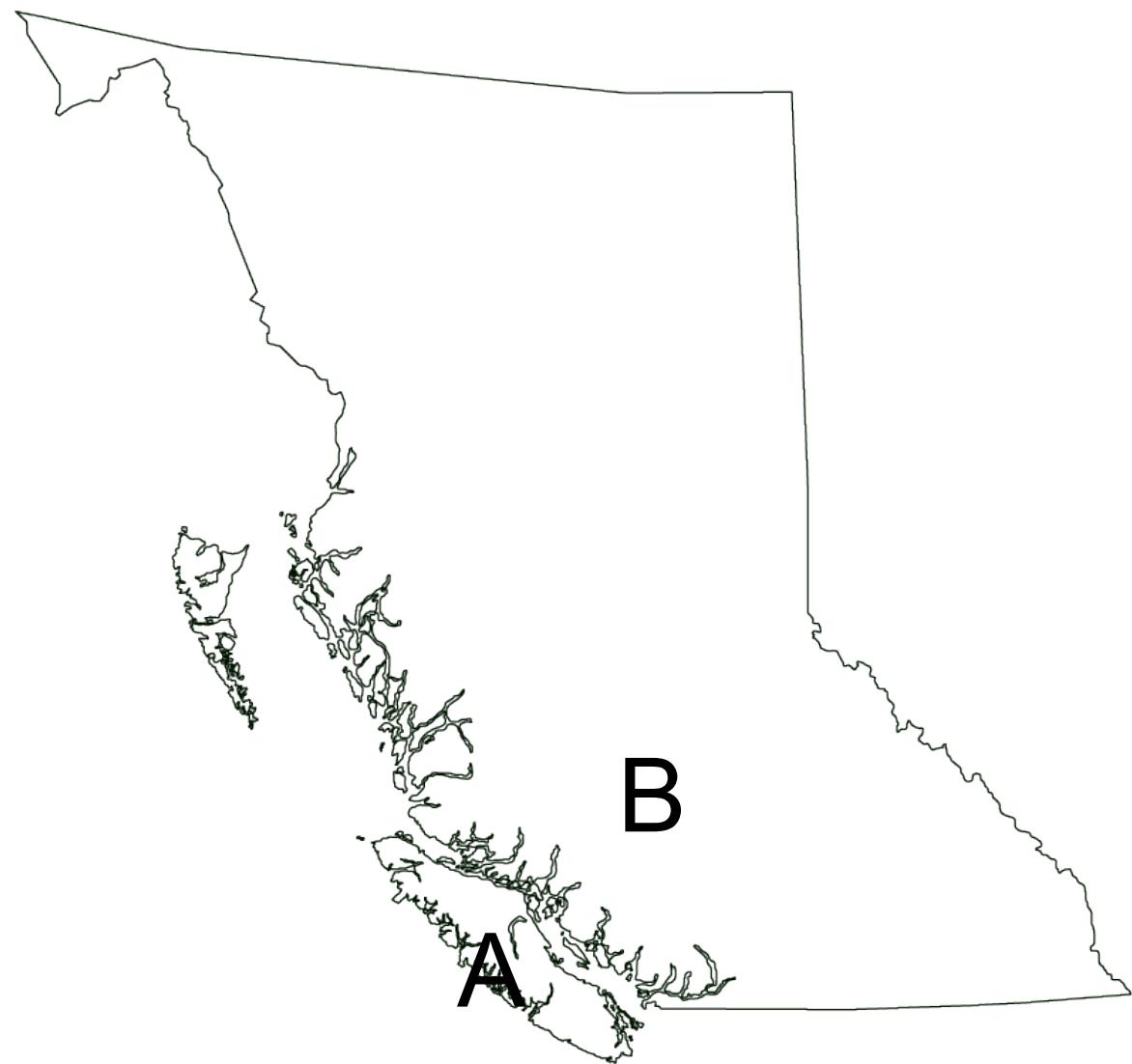
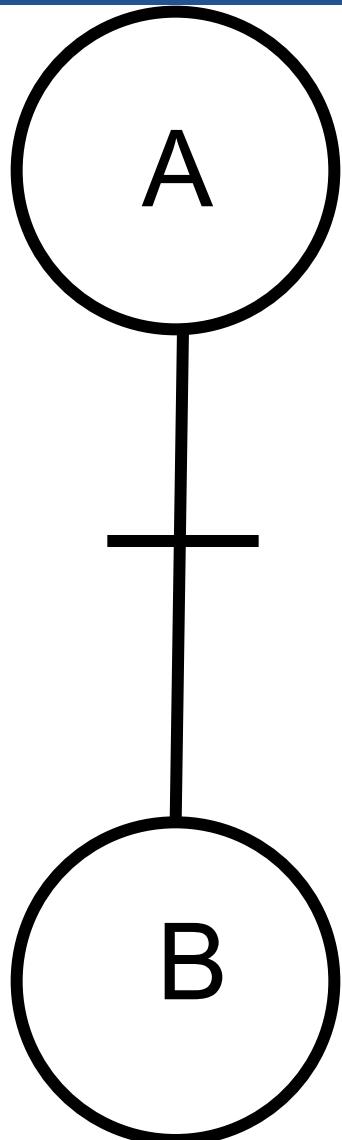
Where did C originate from?



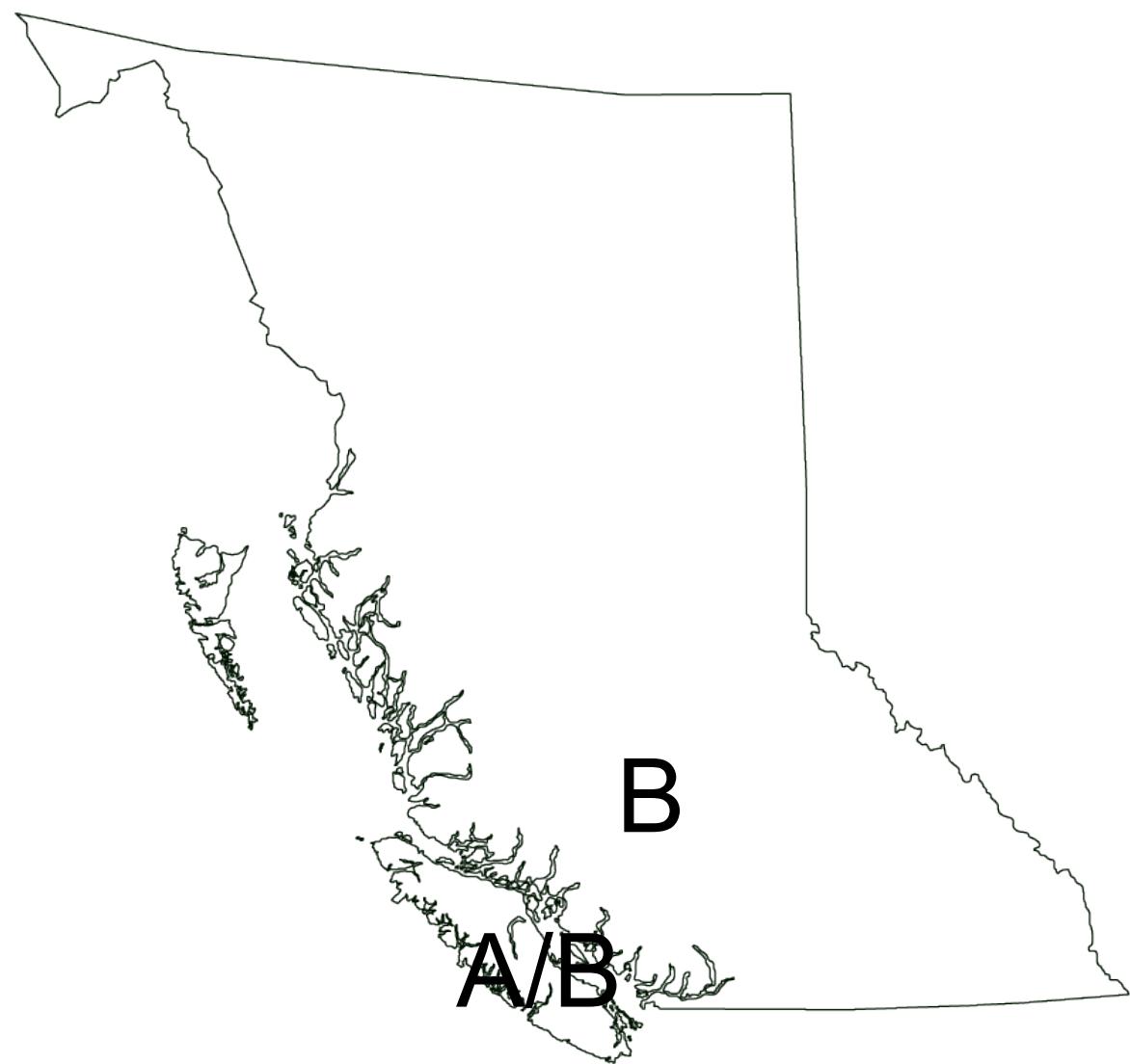
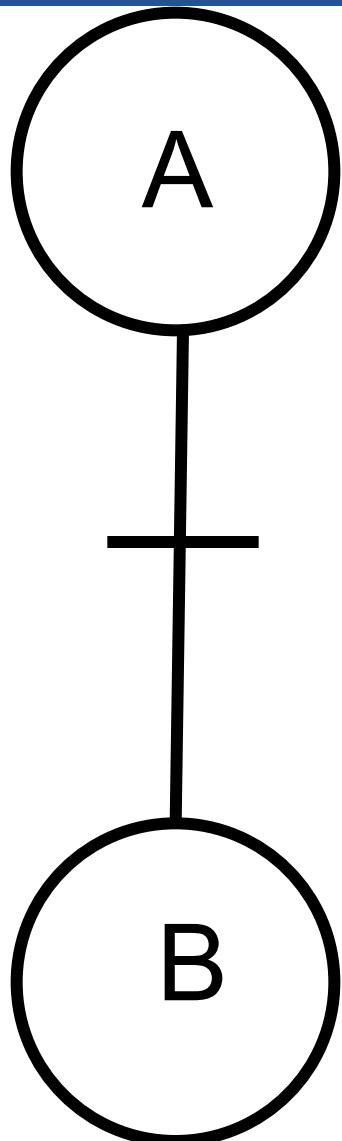
Deep divergence



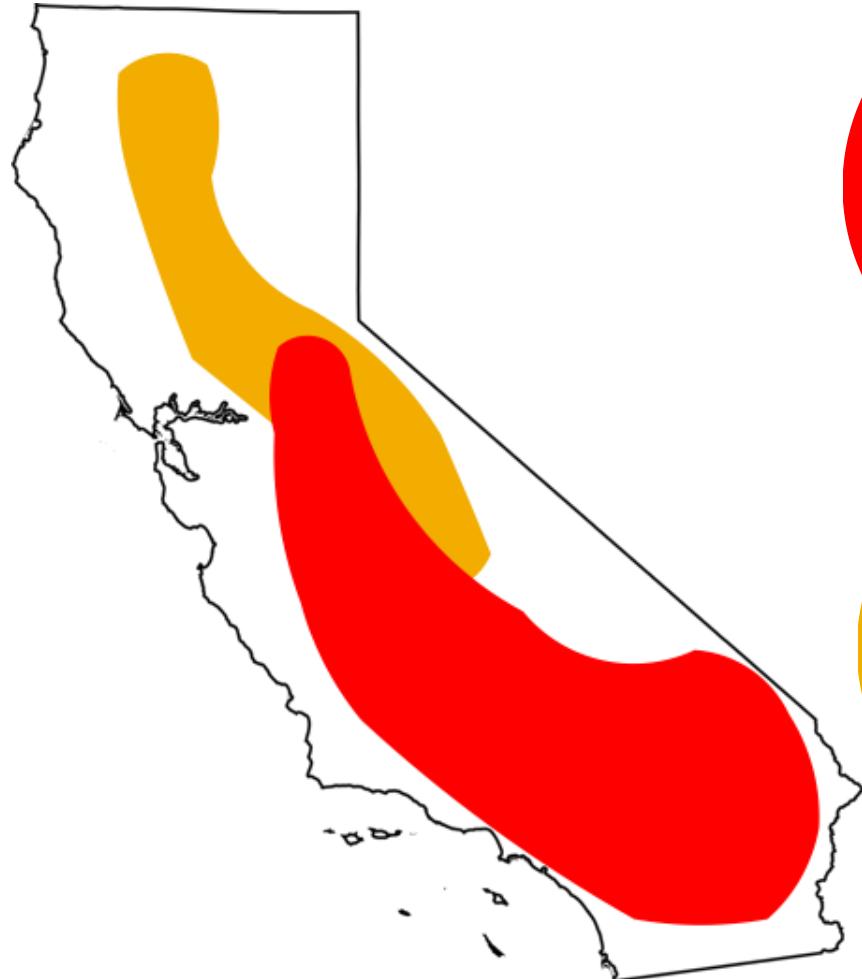
Shallow divergence



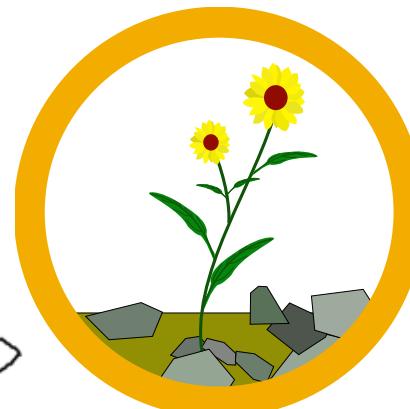
Gene flow



Detecting gene flow



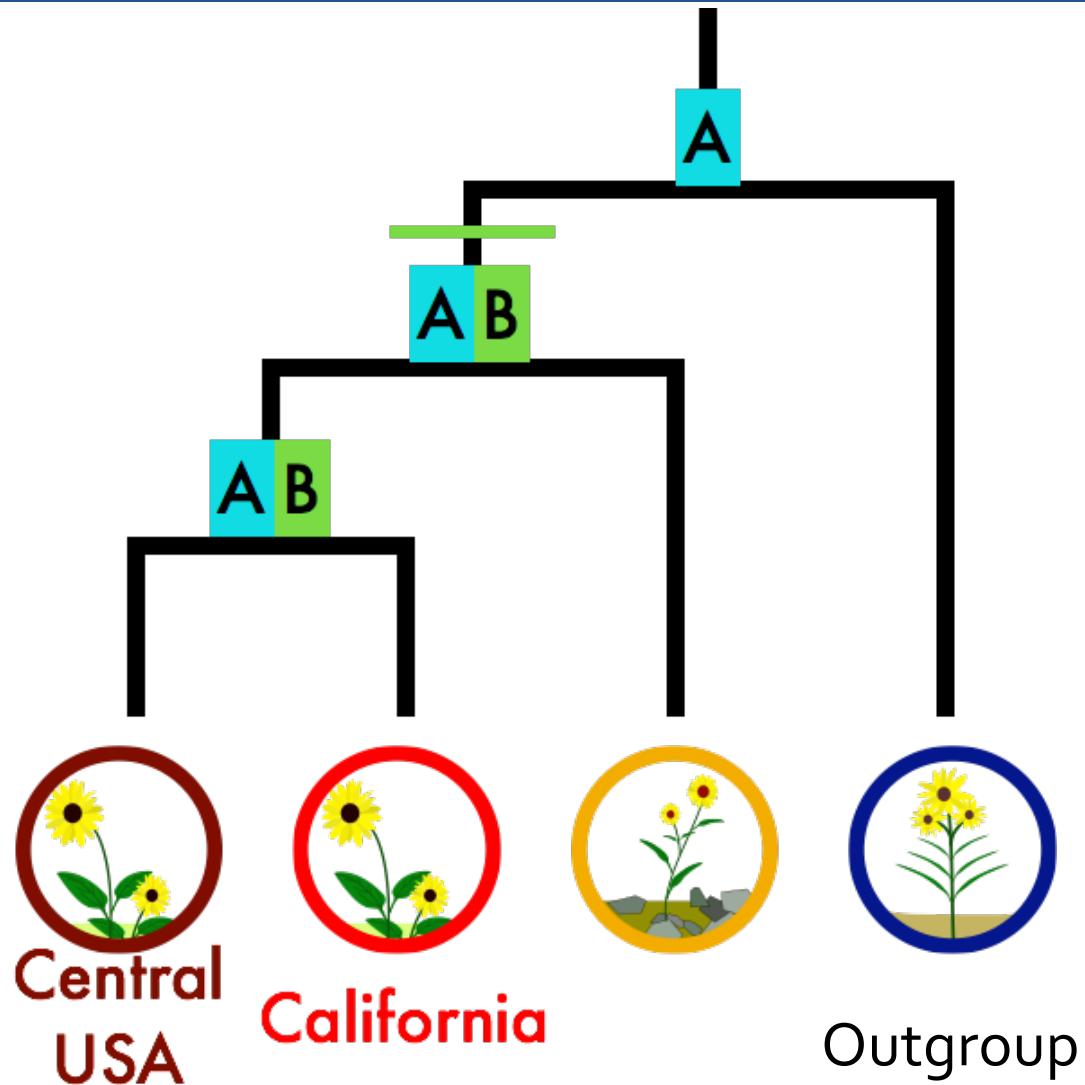
H. annuus



H. bolanderi

also central USA

ABBA-BABA test

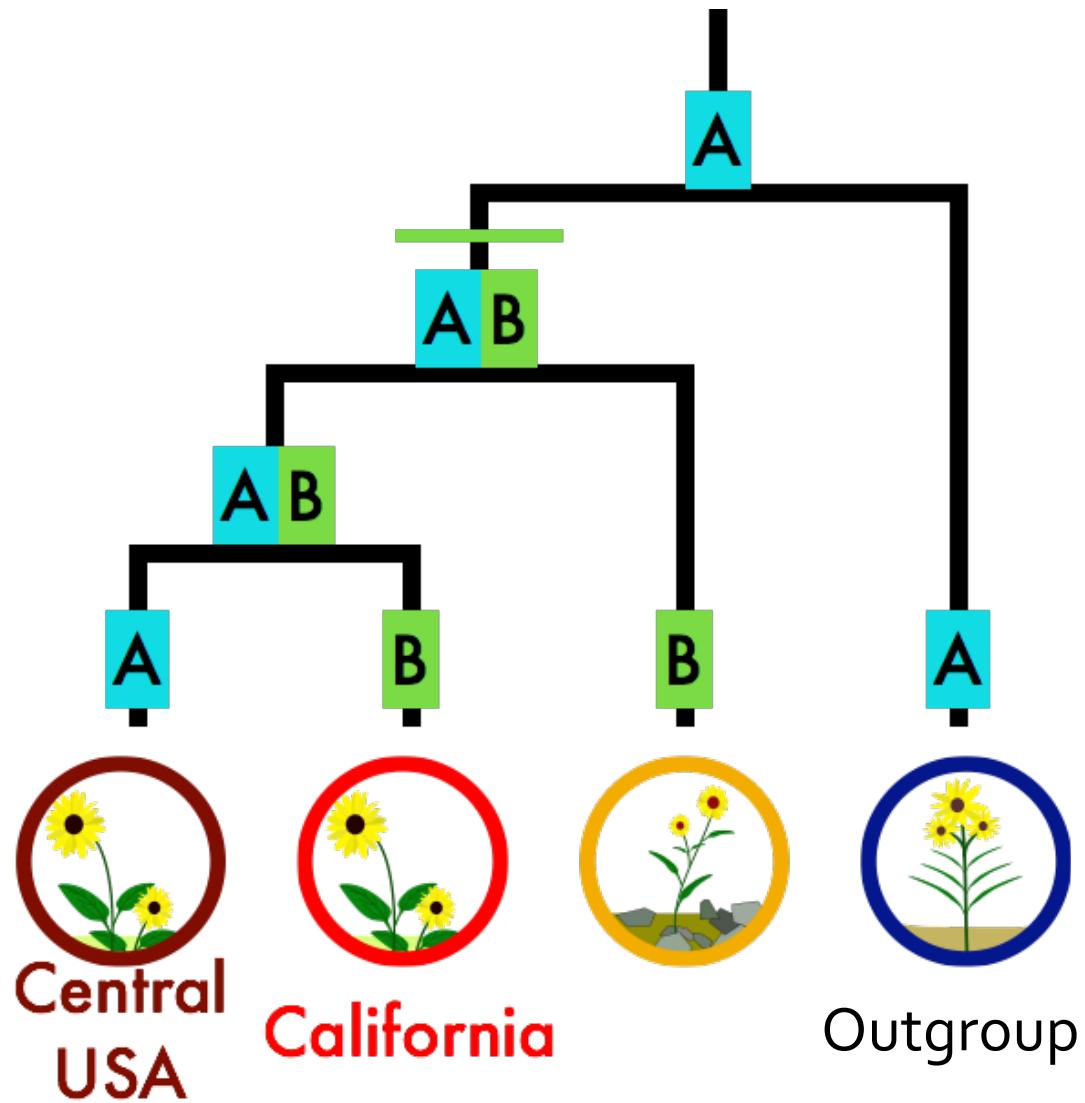


A = Ancestral
B = Derived

H. bolanderi

H. annuus

ABBA-BABA test

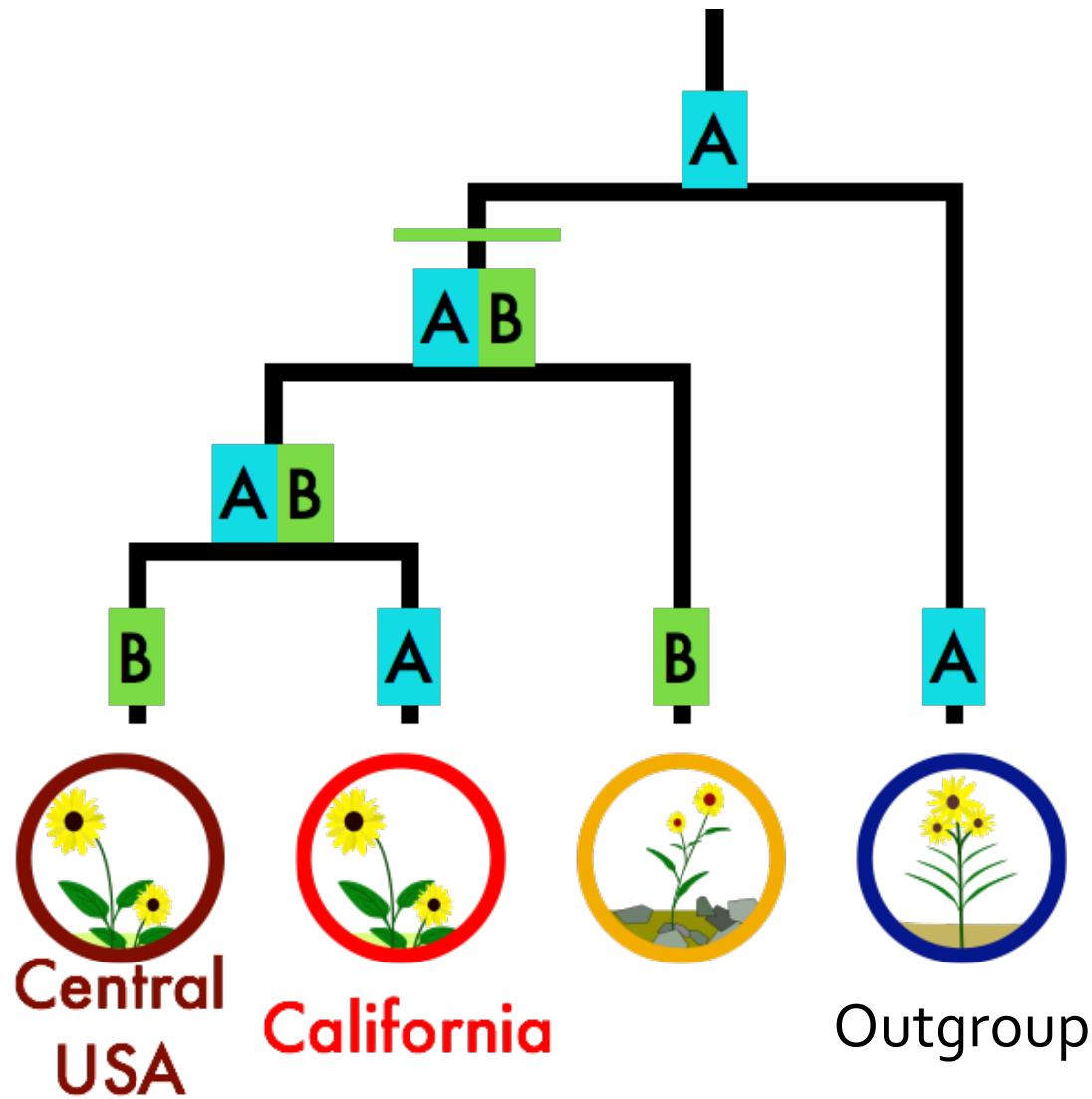


A = Ancestral
B = Derived

H. bolanderi

H. annuus

ABBA-BABA test



A = Ancestral
B = Derived

H. bolanderi

H. annuus

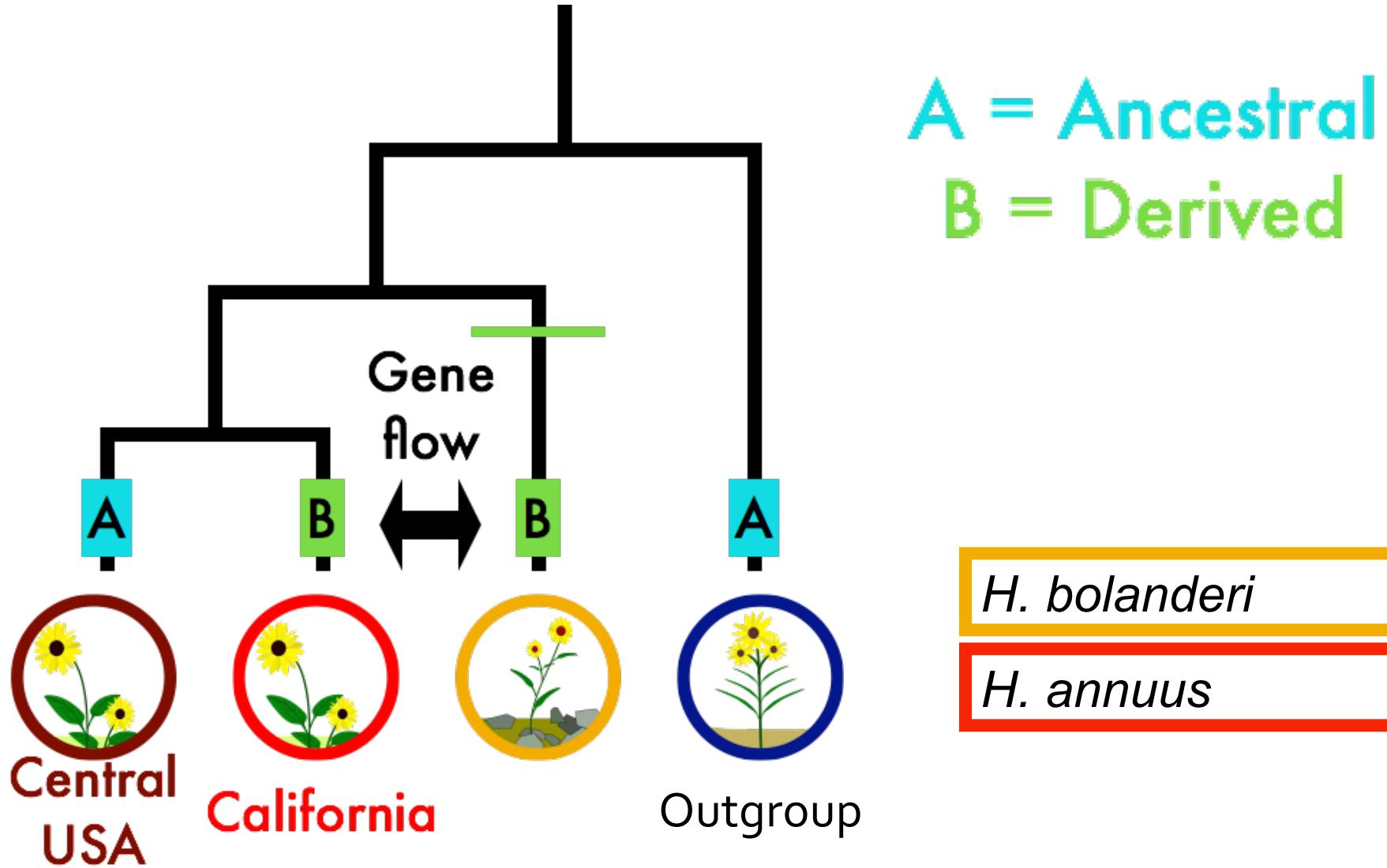
ABBA-BABA test

D-statistic = #ABBA - #BABA

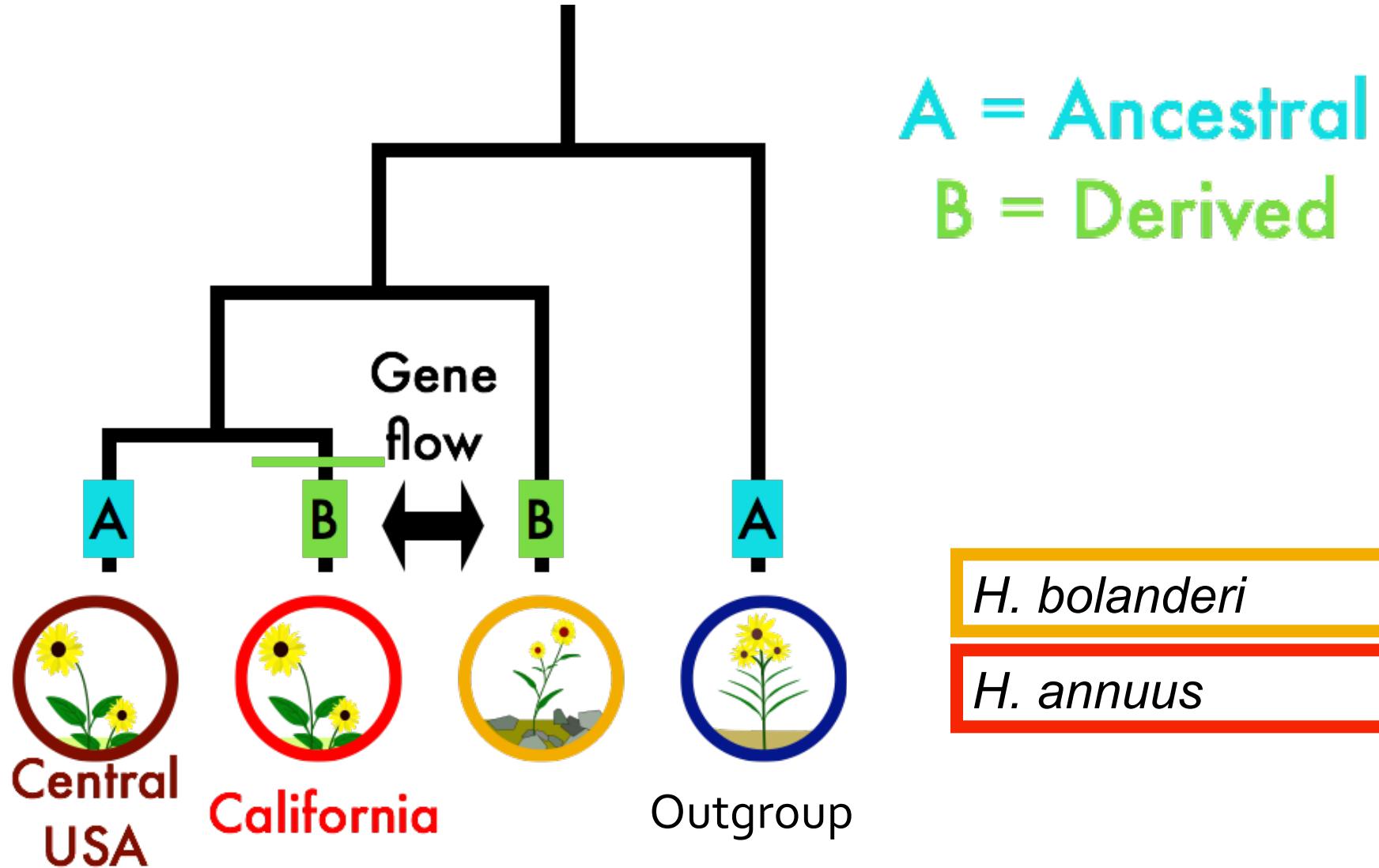
Neutrality: D = 0

Gene flow: D = +

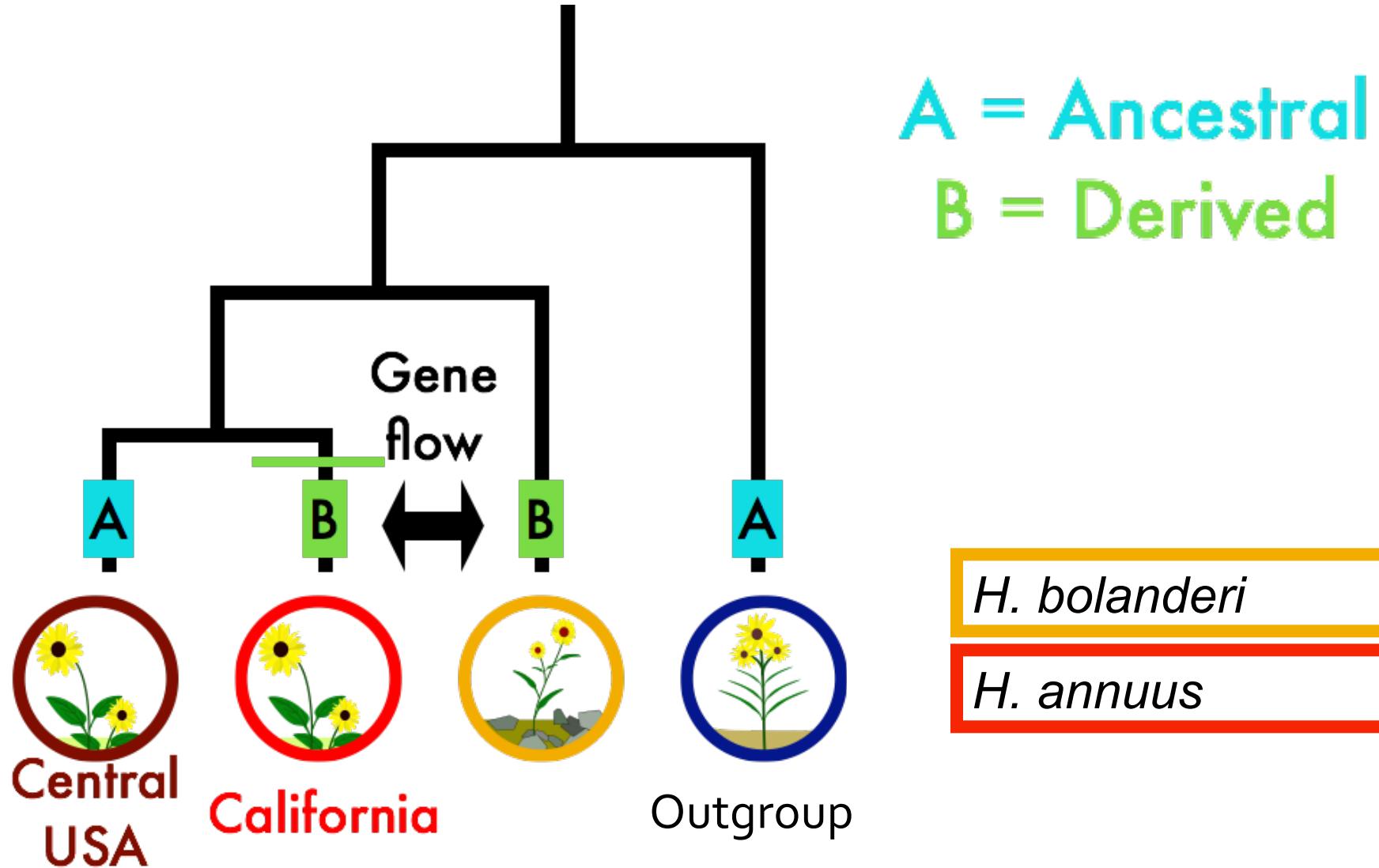
More ABBA under gene flow



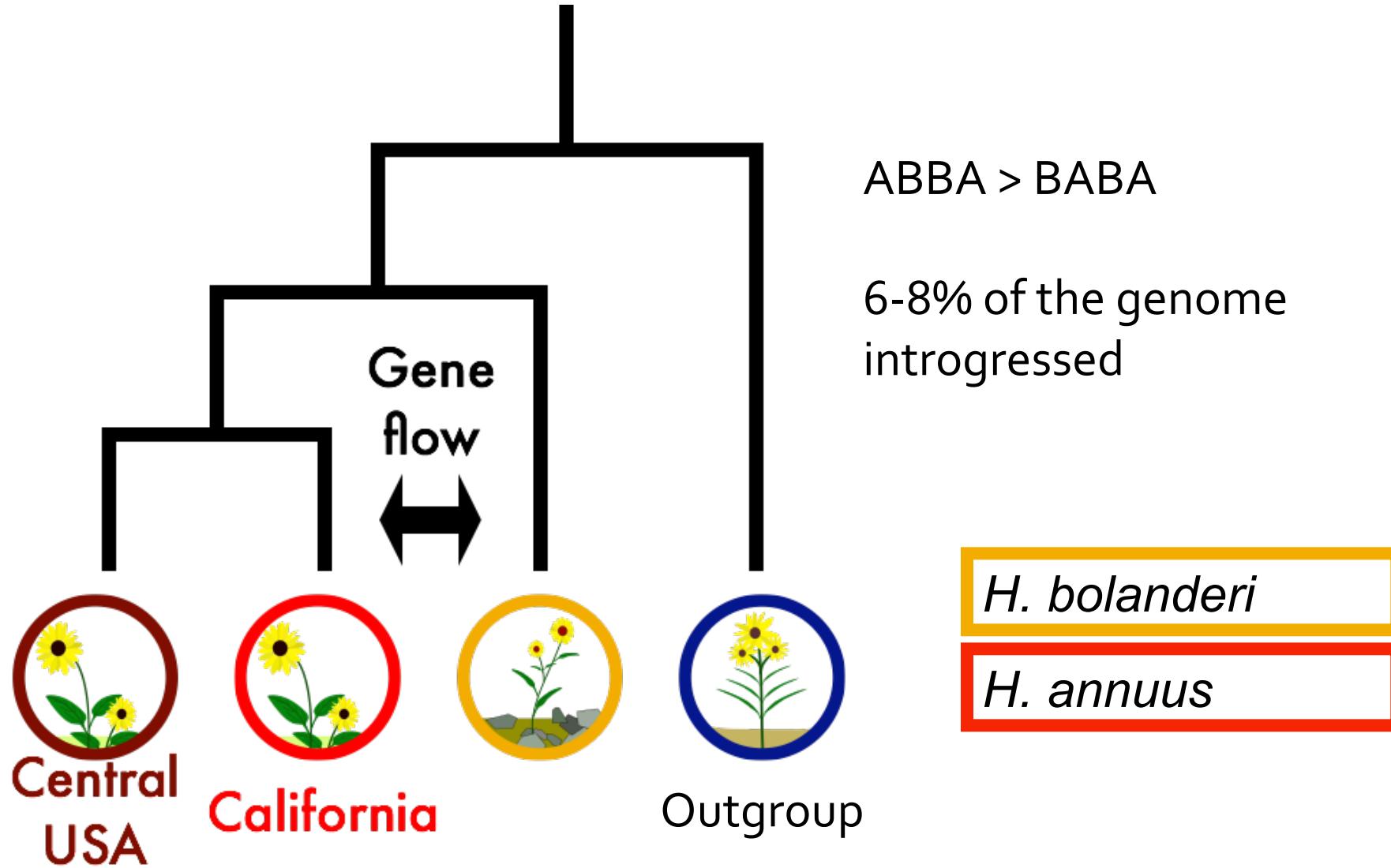
More ABBA under gene flow



More ABBA under gene flow

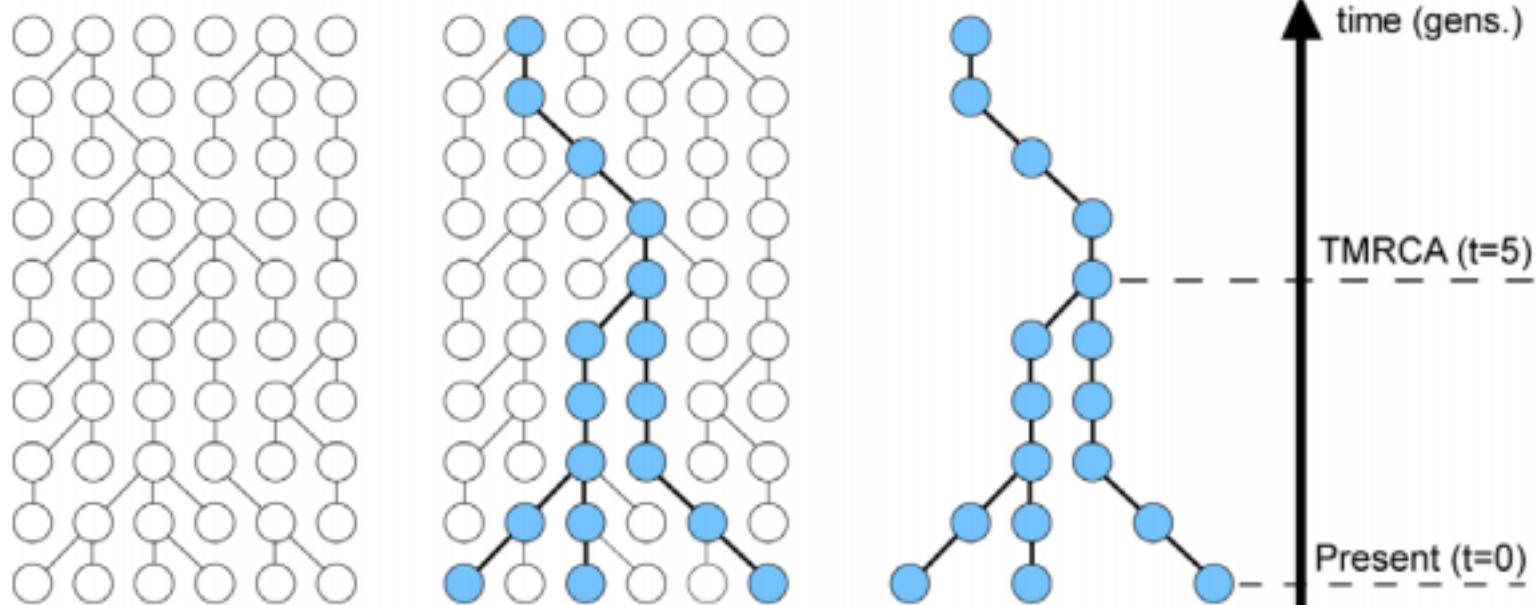


$$D \text{ statistic} = 0.123 \pm 0.033$$



New Approach: Coalescent theory

- A statistical framework for the analysis of genetic polymorphism data
- Is an extension of classical population genetics theory
- It is used to estimate time (number of generations) since lineages coalesced
- Many applications



Coalescent theory

We go backwards in time from the present

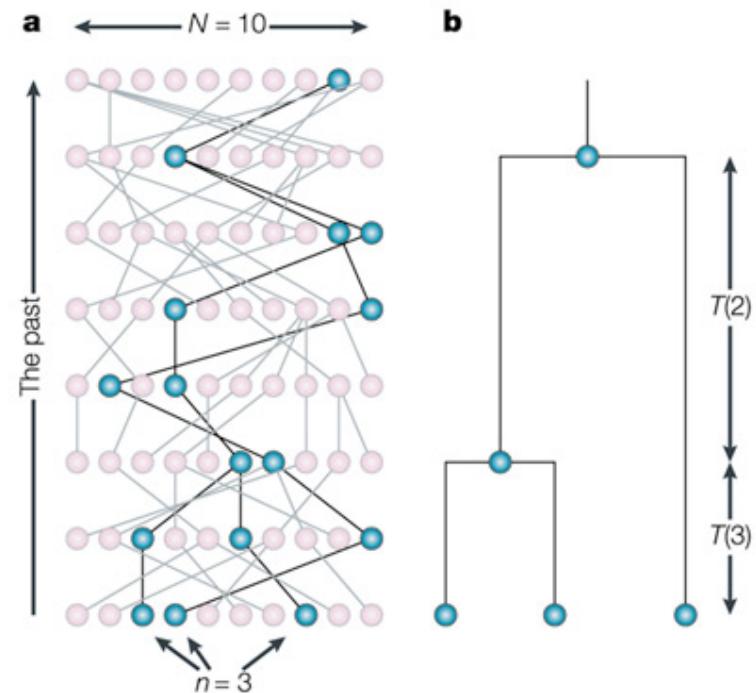
Lineages can randomly “pick” their parents as we go back in time (no selection)

When two lineages pick the same parent they coalesce

All lineages will eventually coalesce to the MRCA of the sample

The rate that lineages coalesce depends on the number of lineages (more lineages the faster the rate) and the size of the population (the more parents to pick from, the slower the rate)

Recombination, gene flow and selection can be incorporated into the model



Nature Reviews | Genetics

N number of alleles in the population
n number of lineages sampled
T time between coalescent events
Rosenberg & Nordborg 2002

Conceptual distinctions

Classical population genetics

- based on allele frequencies
- alleles are either the same or different, no degrees of similarity

Phylogeography

- based on gene genealogies
- both the frequencies of sequence haplotypes and their relationship is considered

Coalescent Theory

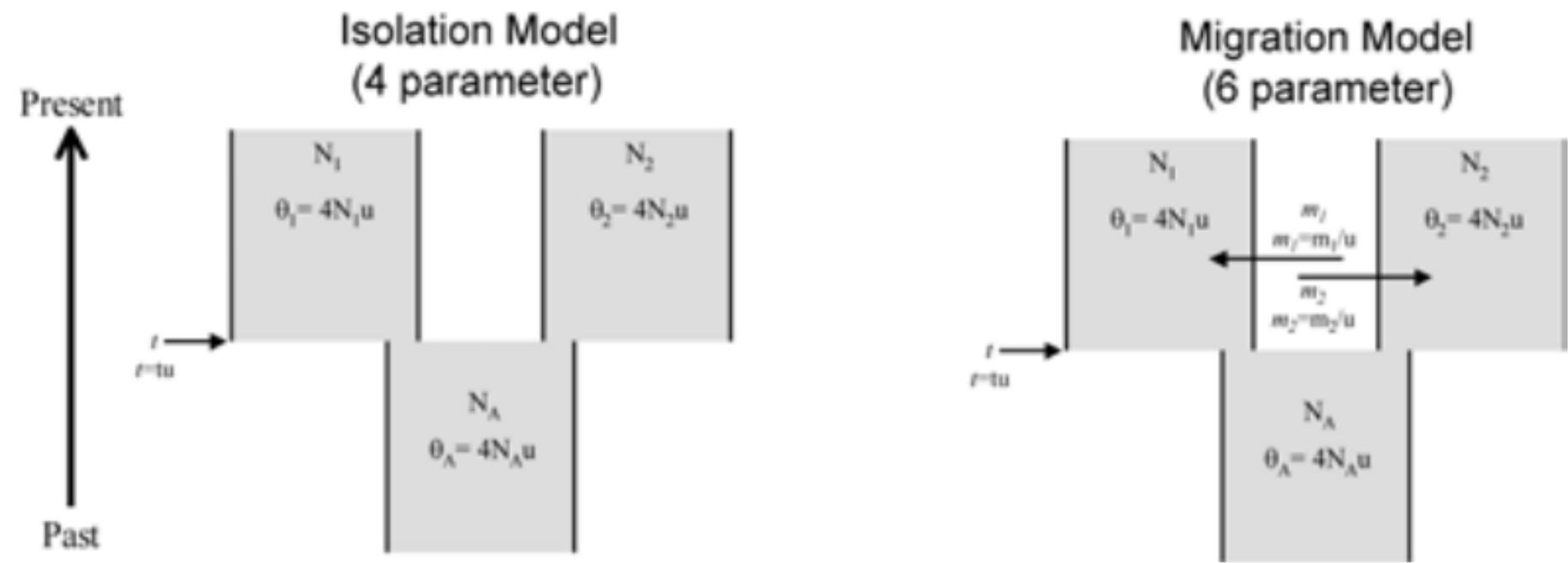
- a mathematical approach to solving population genetic problems
- explicitly considers genealogical processes, even if the data are allele frequencies

Model based methods

1. Collect data

Model based methods

1. Collect data
2. Develop phylogeographic models



Model based methods

1. Collect data
2. Develop phylogeographic models
3. Calculate likelihood support for models based on your data

Model based methods

1. Collect data
2. Develop phylogeographic models
3. Optimize parameters based on your data
4. Calculate likelihood support for models based on your data
5. Pick best supported model

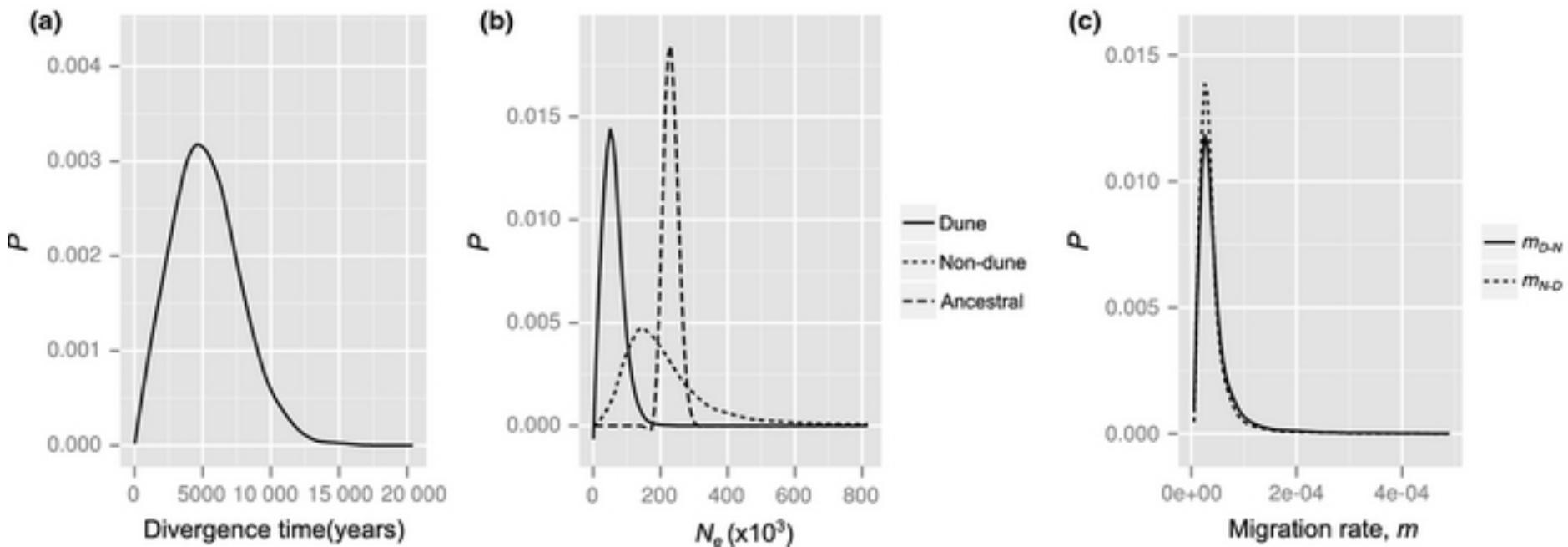
Model based methods

Dune and non-dune sunflower (*Helianthus petiolaris*)



Rose Andrew and Kate Ostevik

Model based methods



Phylogeography: summary

Phylogeography's objective is to understand the processes underlying the spatial and temporal dimensions of genetic variation.

ABBA-BABA test detects gene flow.

Statistical phylogeography, relying on coalescent models, provides a rigorous statistical method to assess demographic hypotheses

Gene genealogies from multiple unlinked nuclear loci are required to provide a better historical record for species

Unanswered questions

How much cryptic diversity exists in nature?

What is the phylogeographic history of ecological communities?