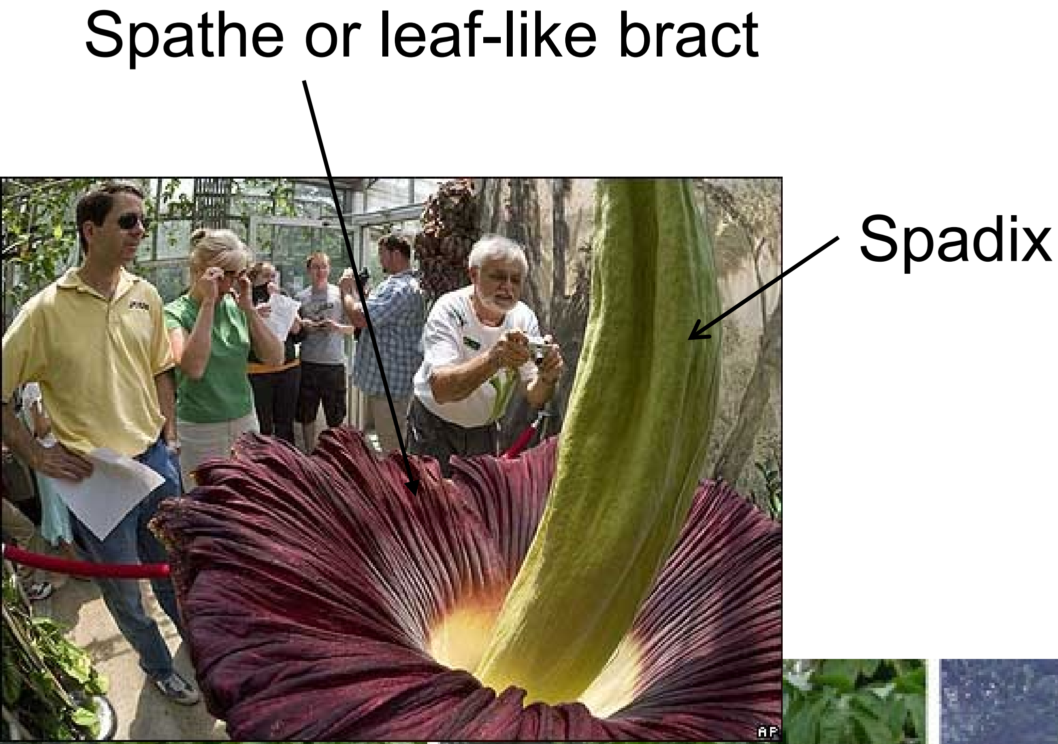
Amorphophallus titanum



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

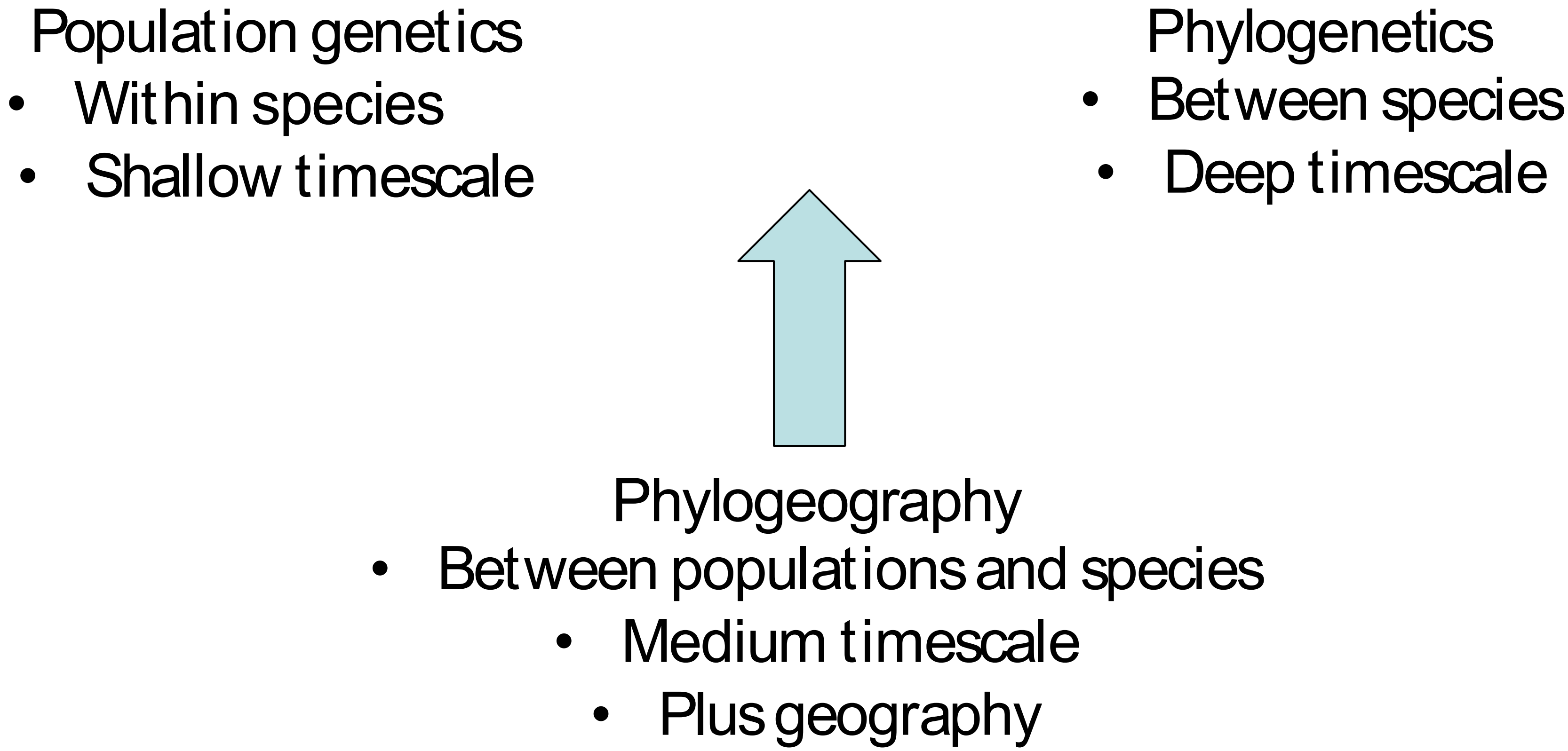
What is phylogeography?

= phylogenetics + biogeography)

“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 species
- Shallow timescale

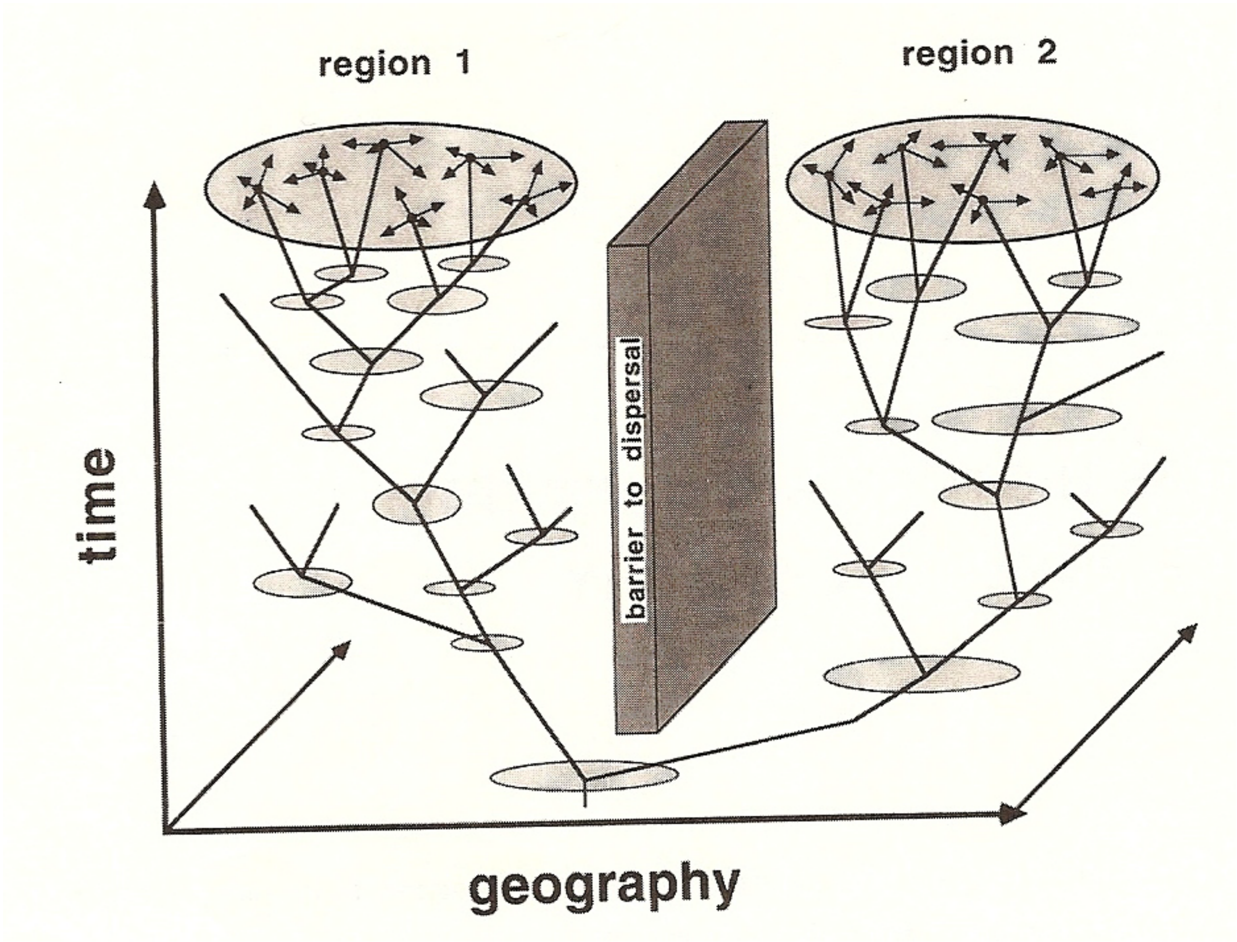
Phylogenetics

- Between species
- Deep timescale

Phylogeography

- Between populations and species
- Medium timescale
- Plus geography

Phylogeography: Introduction



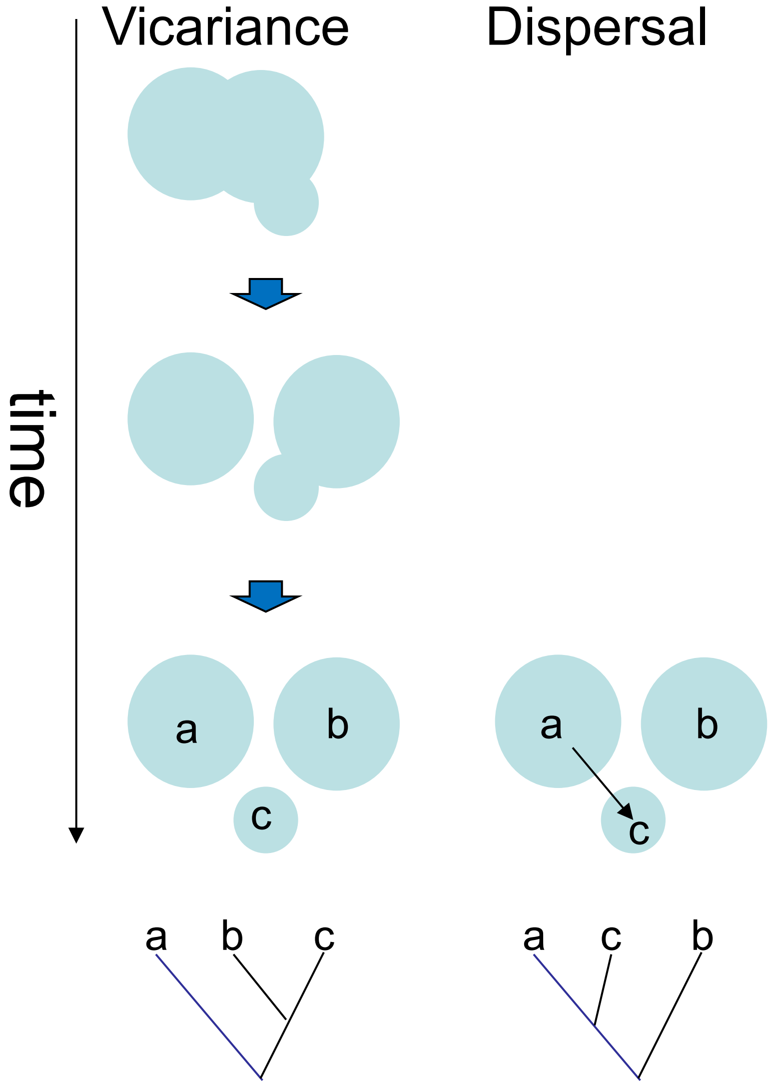
Gene genealogies of interest are mapped in space and time.

Avise 2000

Phylogeography: Introduction

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

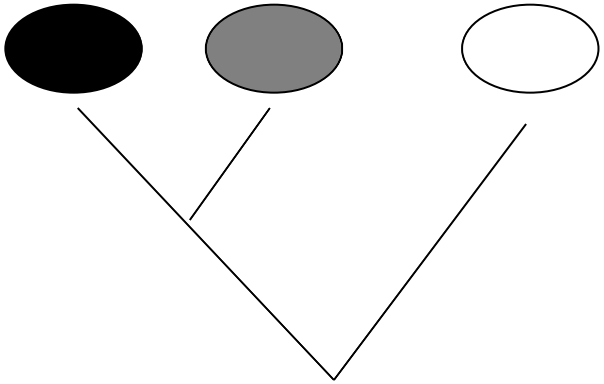
Can evaluate alternative historical scenarios that account for current spatial patterns.



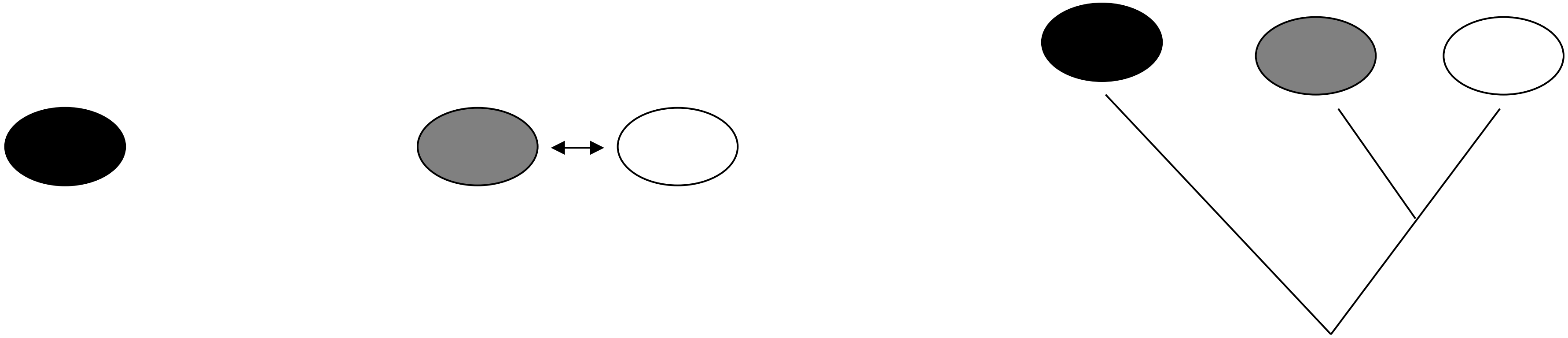
Determinants of Genetic Structure

Common descent and contemporary gene flow influence spatial patterns of genetic variation.

Similarity due to common descent



Similarity due to gene flow



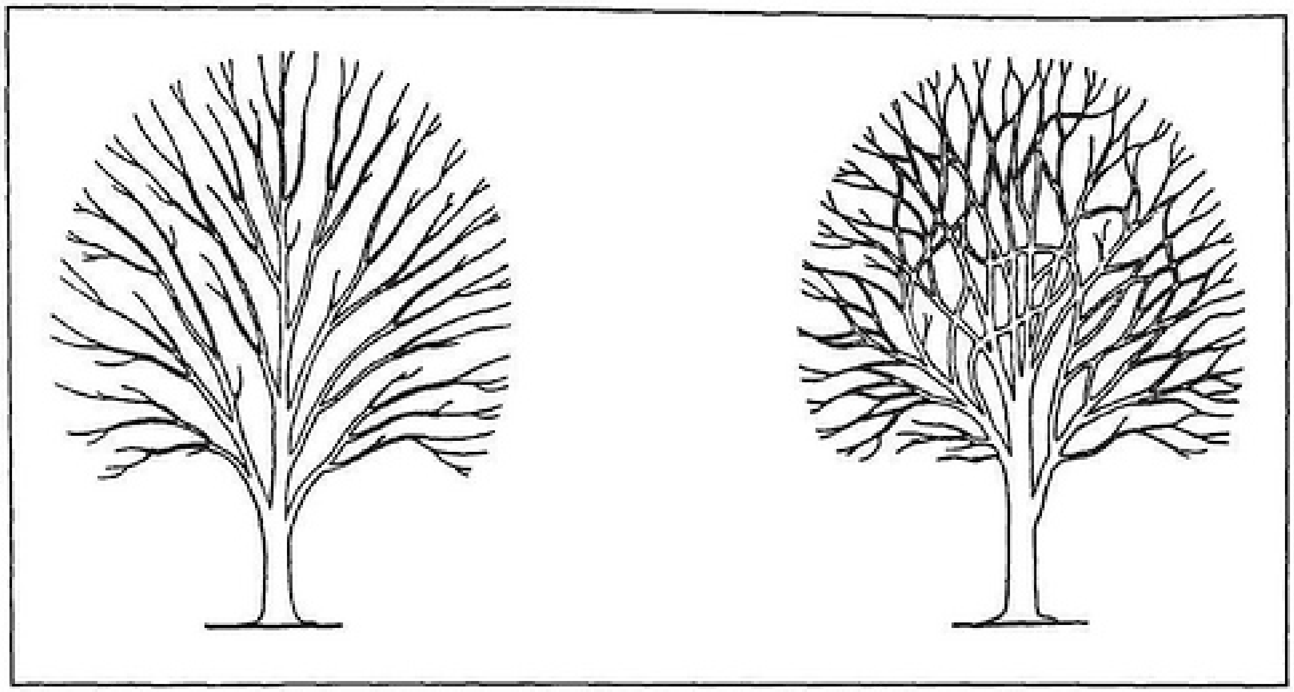
Understanding Genetic Structure

- Traditionally:
  - **Within species**, *gene flow* has been emphasized as the cause of similarity.
    - Assumes drift / gene flow equilibrium
  - **Among species**, *common descent* has been emphasized as the cause of similarity.
    - most phylogenetic methods assume no 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 *gene flow* from *common descent*

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 no reticulation

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

Step 1: the search for 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 (except in inverted repeat)

-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, but offers thousands of loci for study

Phylogeography: the early years

Phylogeography was essentially *descriptive*:

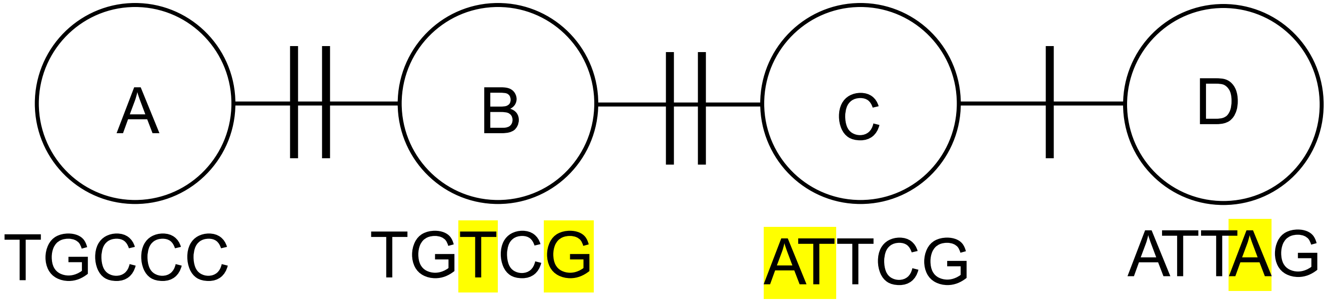
Plot haplotypes (alleles) on map

Classify pattern of phylogeographic structure

Consider historical explanations

Look for concordance among different species (comparative phylogeography)

Haplotype network

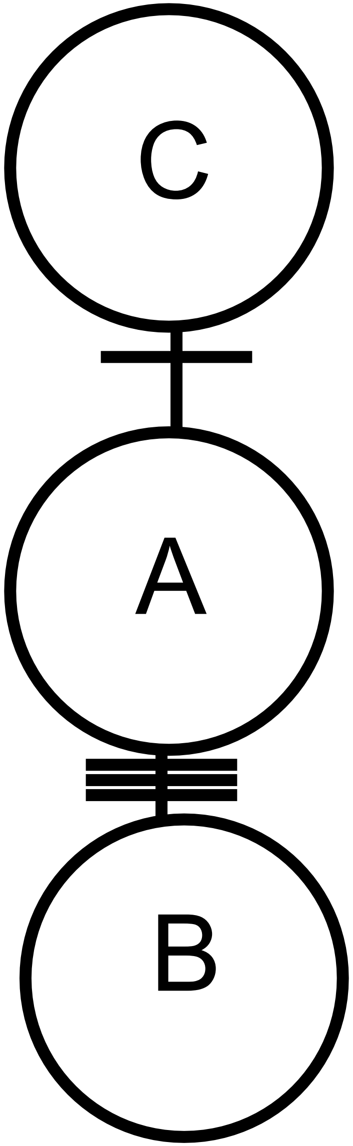


A B C D

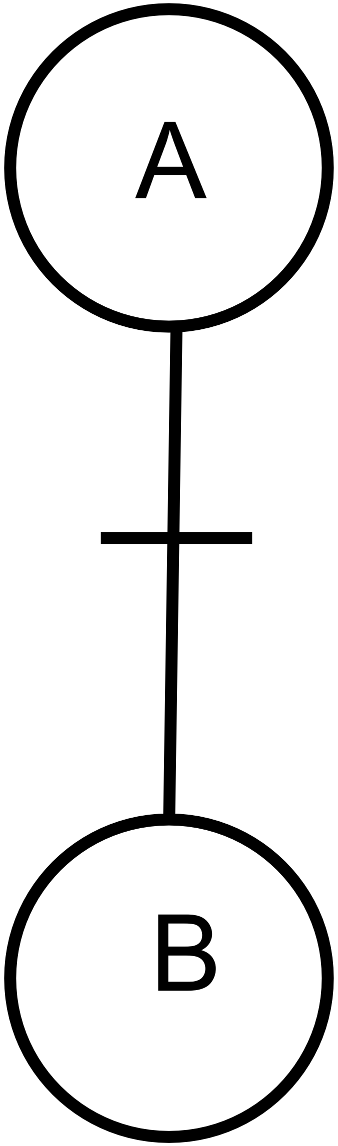
TGCCC TGTCG ATTCG ATTAG

Haplotype = stretch of DNA that is inherited as a unit

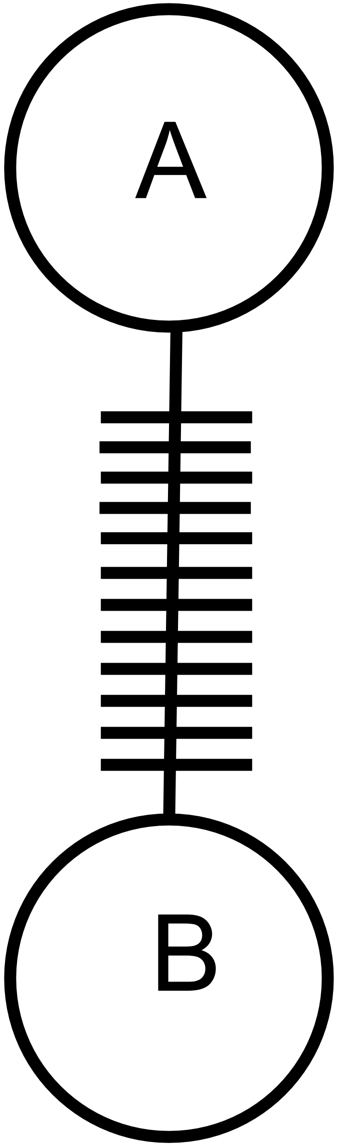
Where did C originate from?



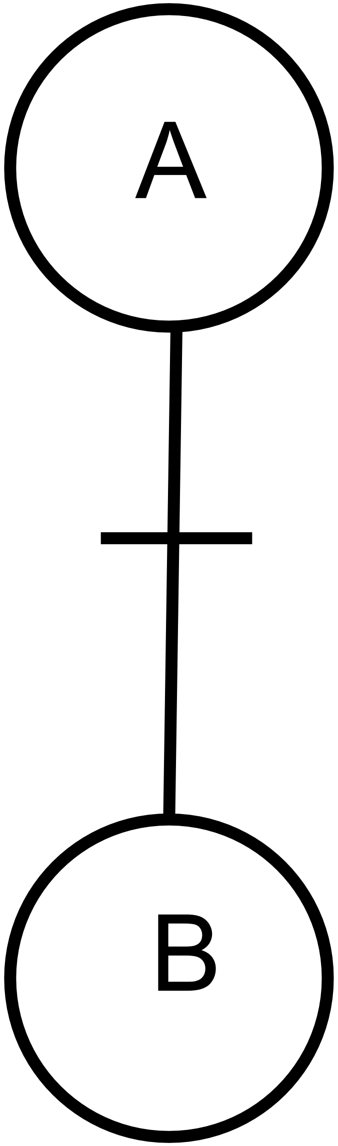
Shallow Divergence

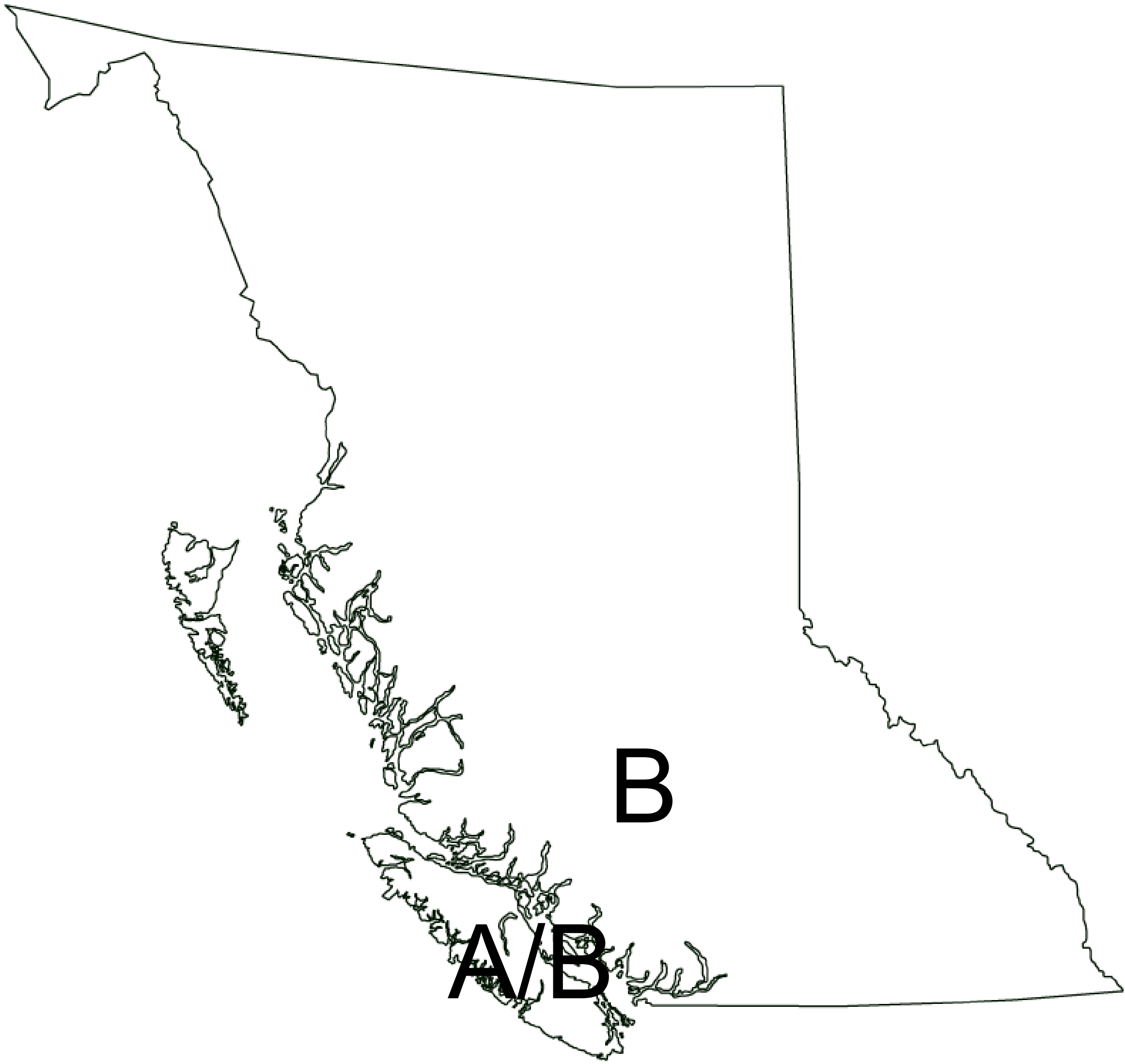


Deep Divergence



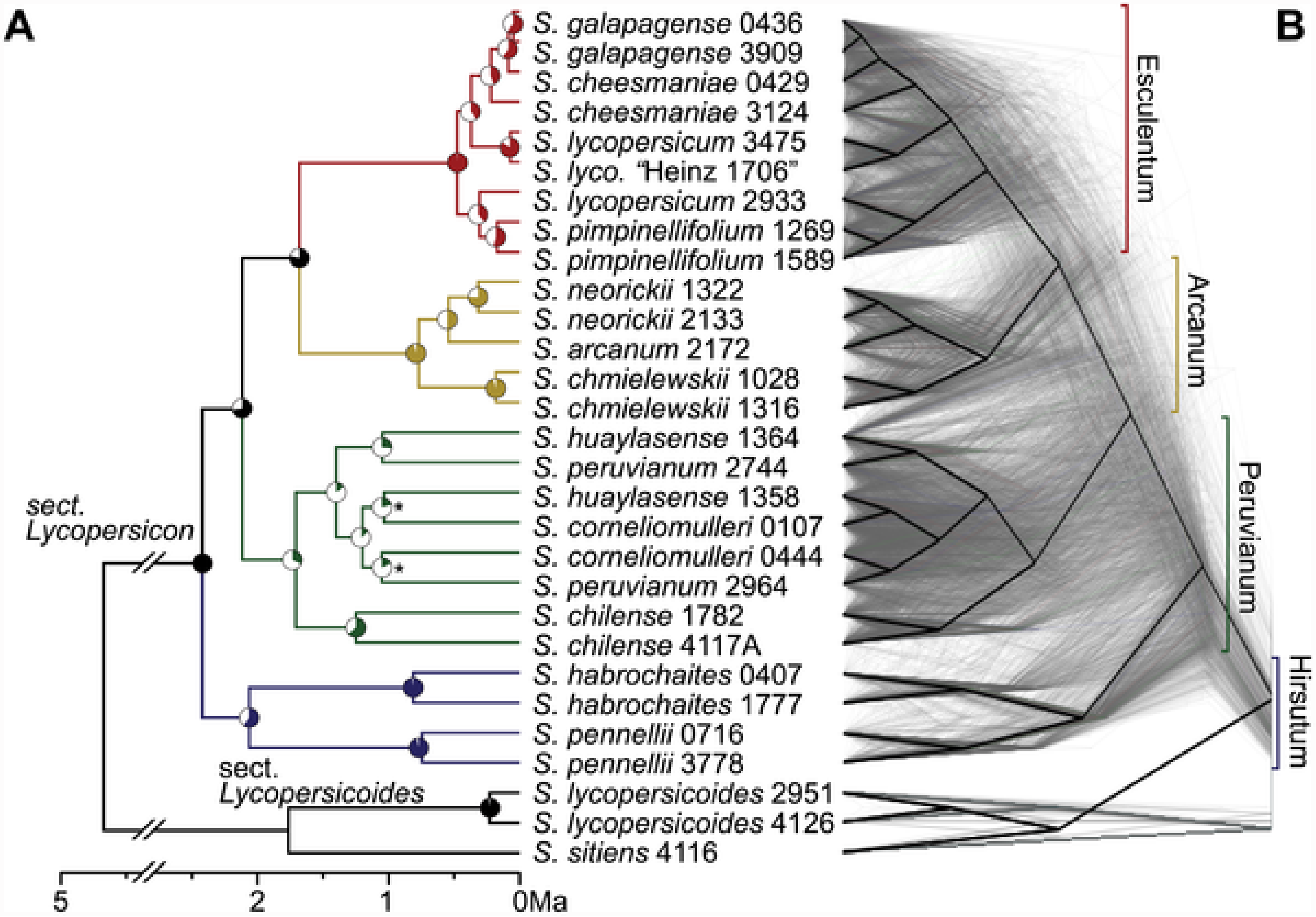
Gene flow





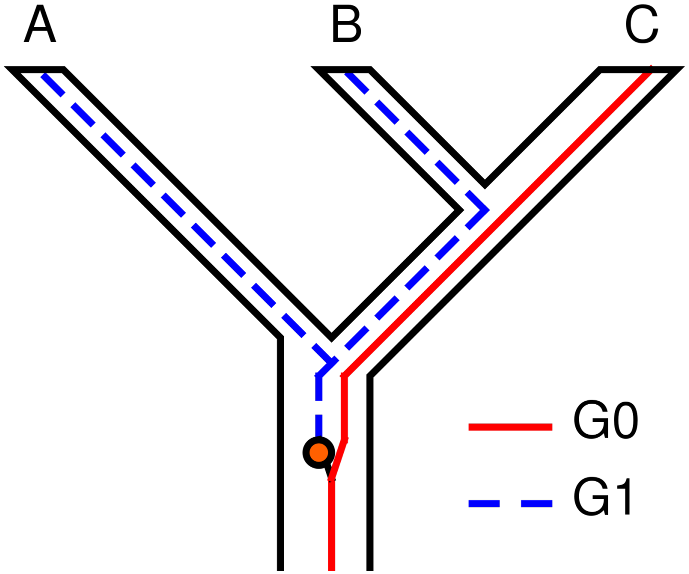
Problem – gene tree heterogeneity

The phylogeny of Solanum sect. Lycopersicon.

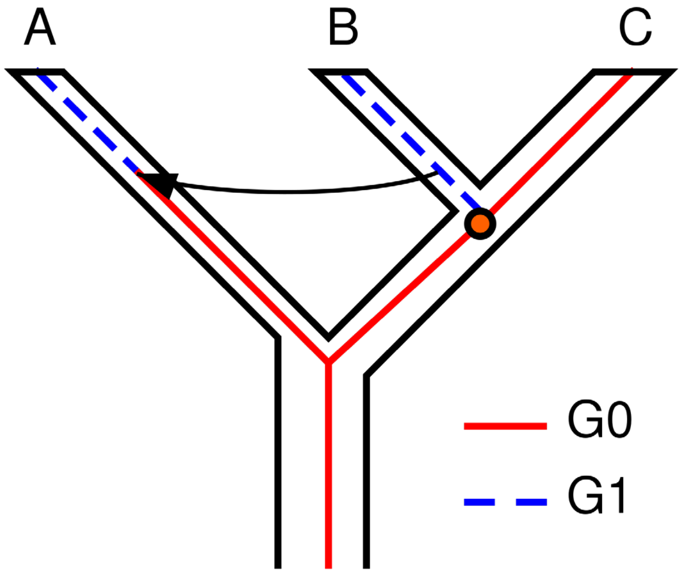


Pease JB, Haak DC, Hahn MW, Moyle LC (2016)

Causes of gene tree heterogeneity

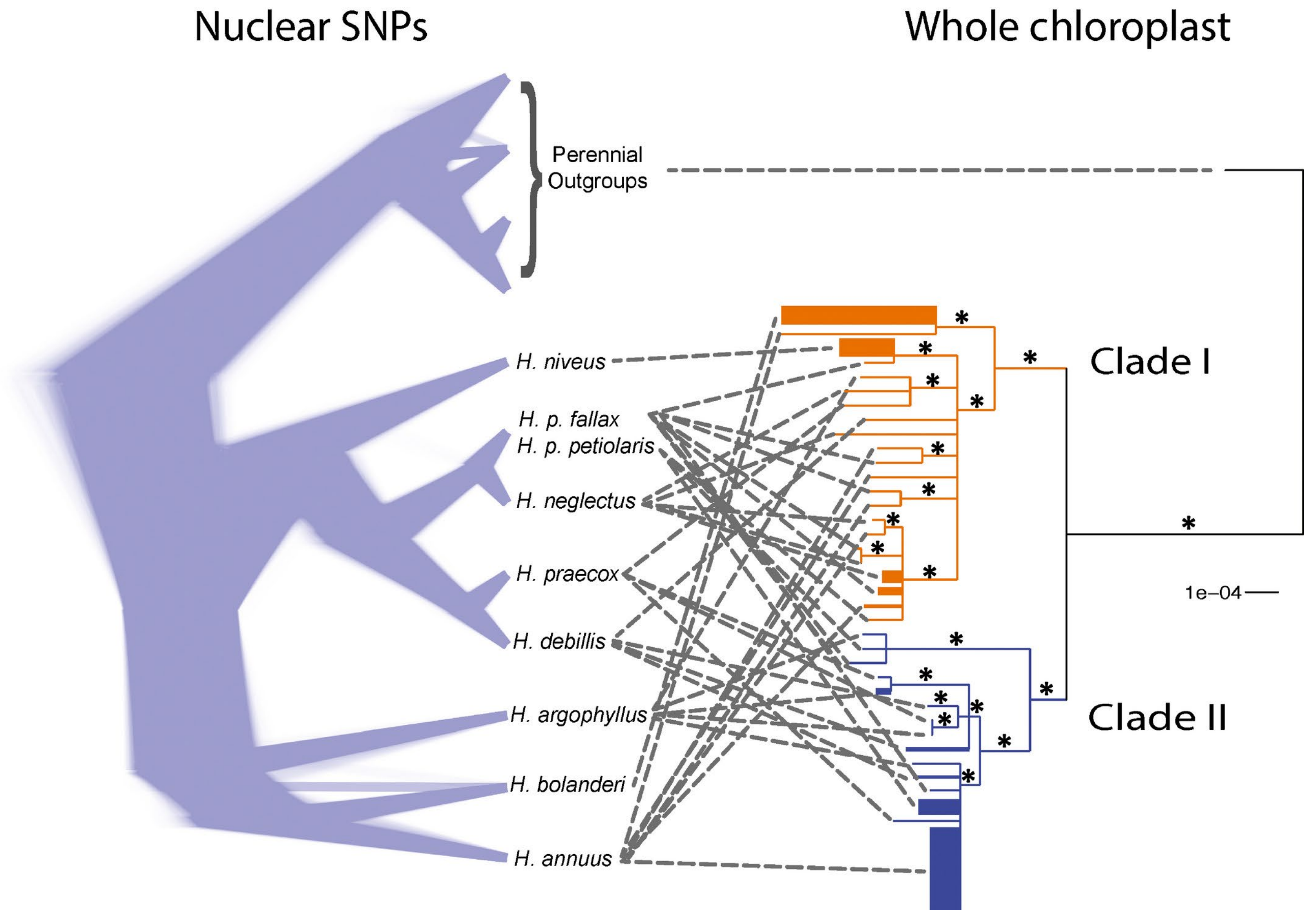


Incomplete lineage sorting



Introgression

Discordance between nuclear and chloroplast phylogenies in sunflower due to introgression



Genetic distance between chloroplasts in different species was smaller that expected under incomplete lineage sorting

New Phytologist, Volume: 221, Issue: 1, Pages: 515-526, First published: 23 August 2018, DOI: (10.1111/nph.15386)

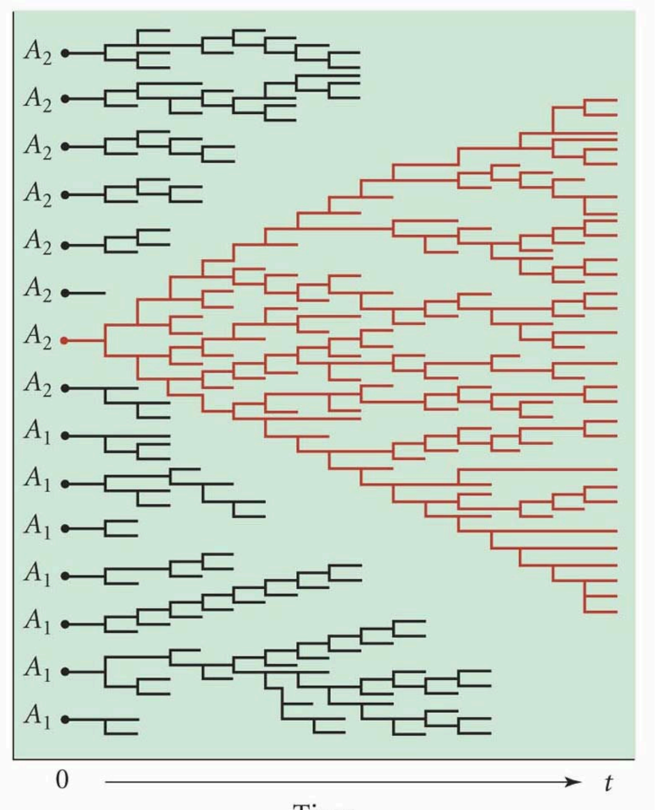
New Approach: Coalescent theory

- 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

All gene copies in a population ultimately are descended from a single ancestral copy

(Or, all gene copies **coalesce** back to a single common ancestor)

Futuyma *Evolution* 2009, fig. 10.1



Coalescent theory

We go backwards in time from the present

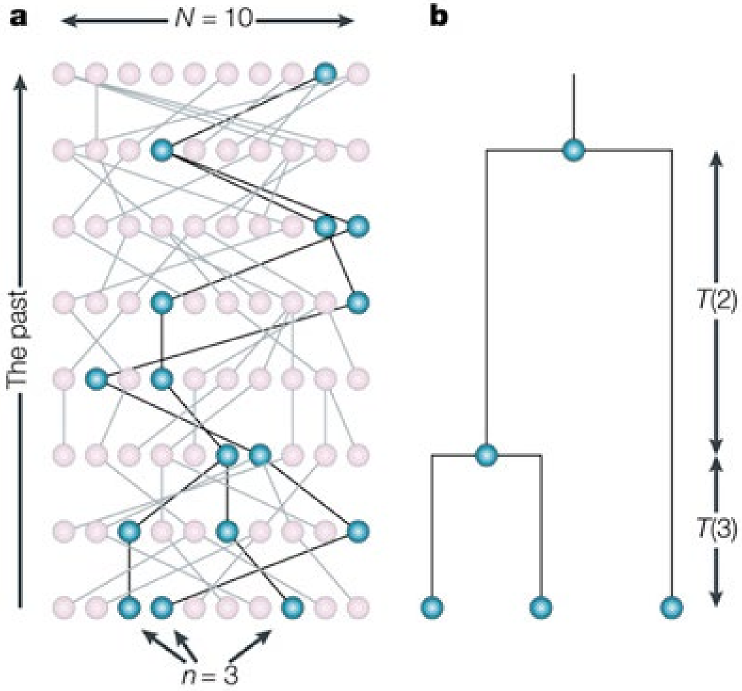
Lineages can randomly “pick” their parents as we go back in time (no selection)

When two lineages pick the same parent they coalesce

All lineages will eventually coalesce to MRCA (most recent common ancestor of 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

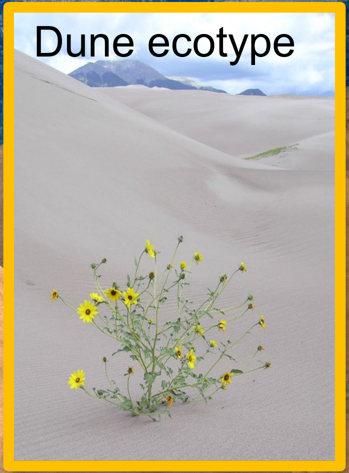
-explicitly considers genealogical processes

Phylogeography: coalescent theory

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

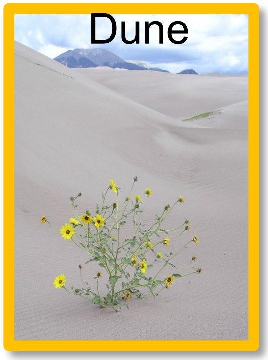
Coalescent theory - example

Sunflower ecotypes: *H. petiolaris*

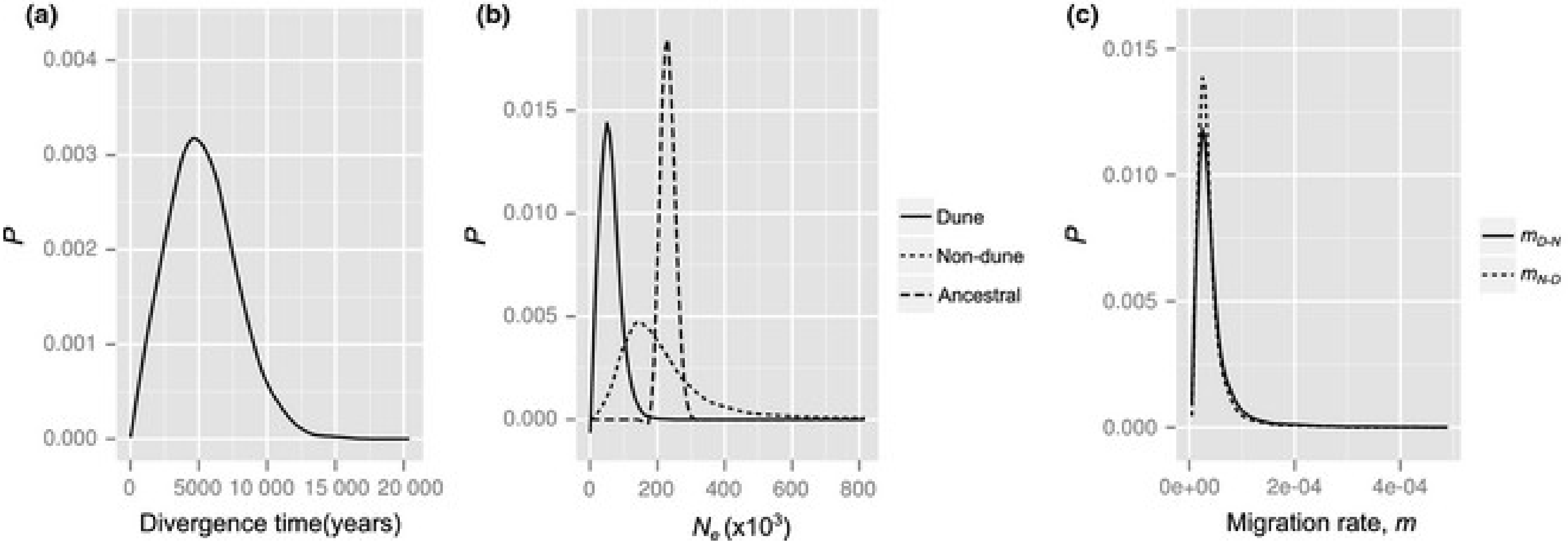


Photos: NPS/Patrick Myers, Kate Ostevik, J.D. Herndon

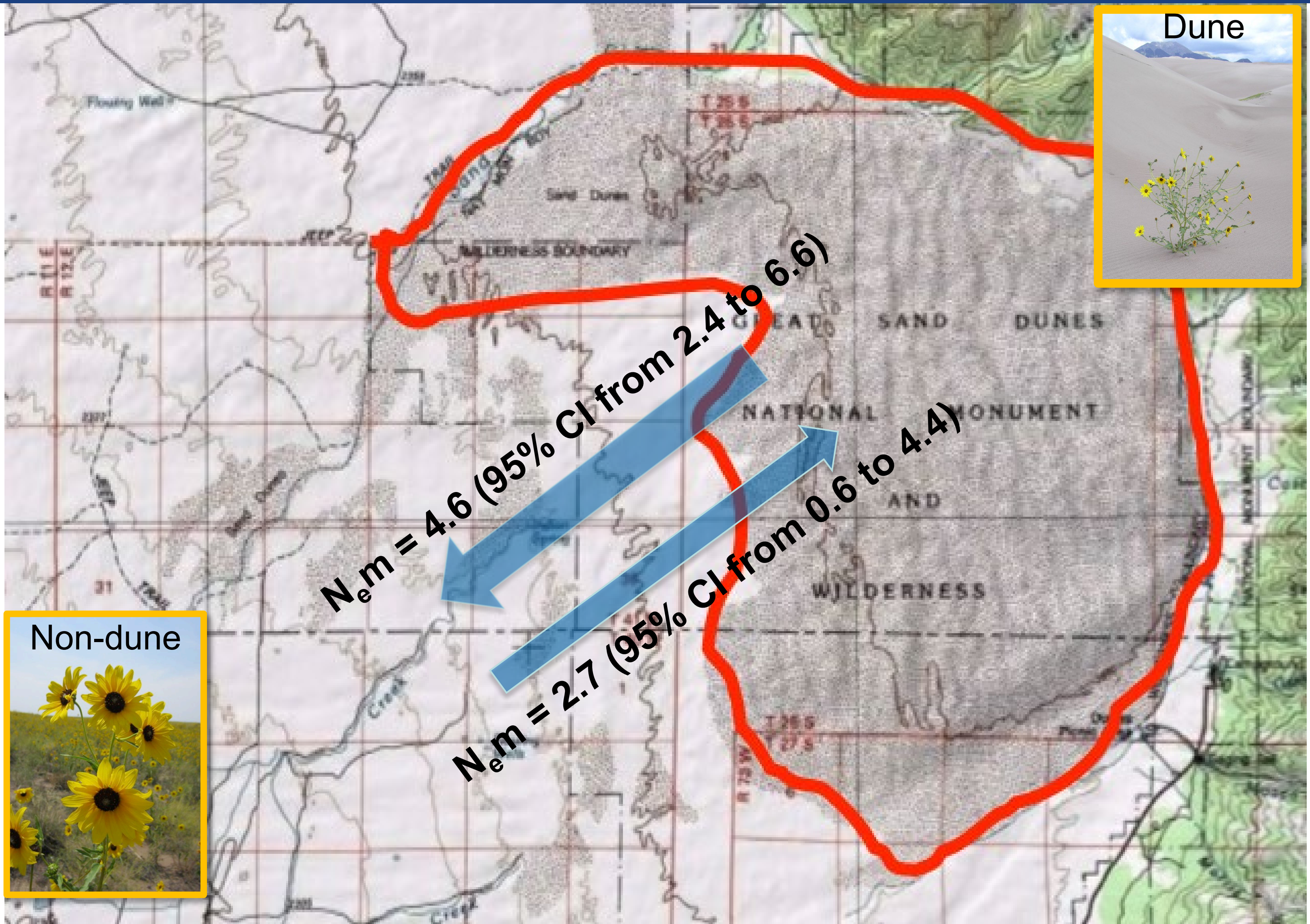
Coalescent theory – Isolation with migration model



Andrew et al. (2013, *Molecular Ecology*)



Coalescent theory – Isolation with migration model



$N_e m = 4.6$ (95% CI from 2.4 to 6.6)

$N_e m = 2.7$ (95% CI from 0.6 to 4.4)

Gene flow moderate and asymmetric

Phylogeography: summary

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

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 relative importance of incomplete lineage sorting versus introgression in explaining gene tree heterogeneity?