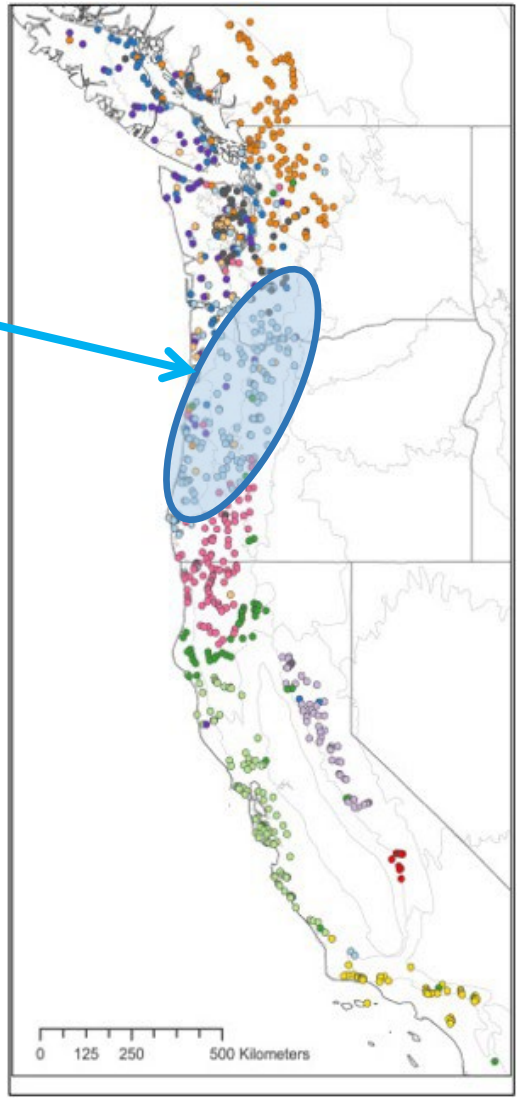
Genetic population assignment:
potential for seed provenance ID?

Hayley Tumas (she/her)
Sally Aitken's Lab | University of British Columbia
on the traditional & unceded territory of the xʷməθkʷəy̓əm (Musqueam)
Challenges to our Future Tree Seed Supply
June 23, 2022

What is population assignment?

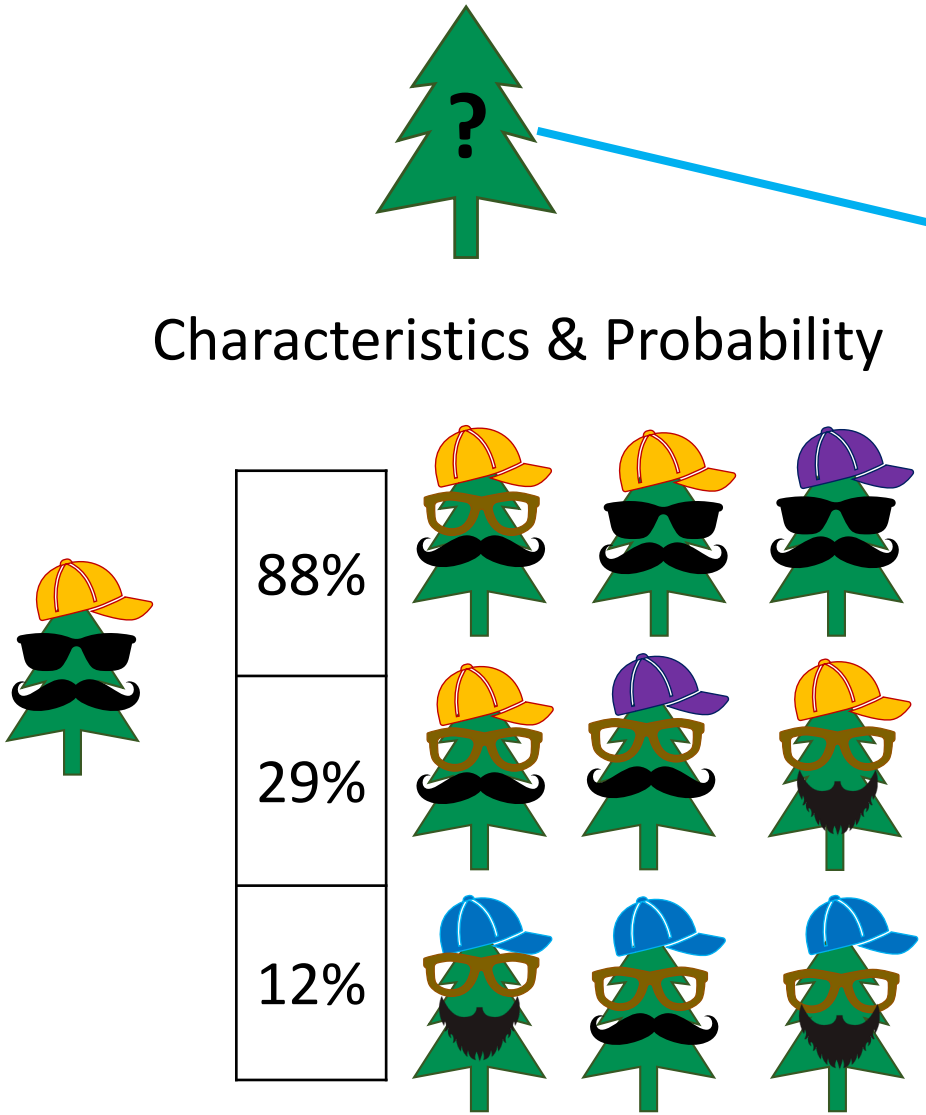


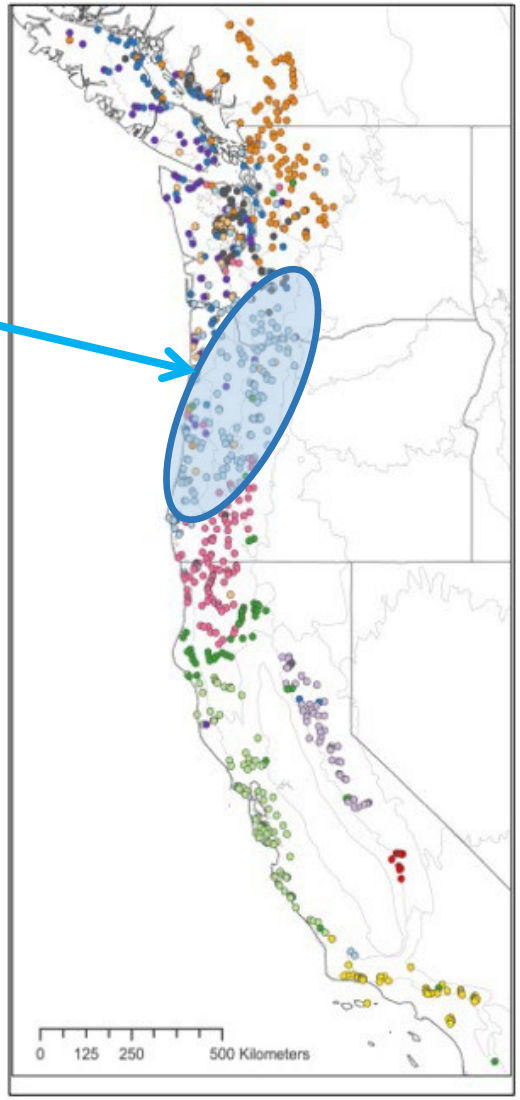
Provenance identification*

*if populations sufficiently differentiated at finer geographic scale

Cronn et al. 2021

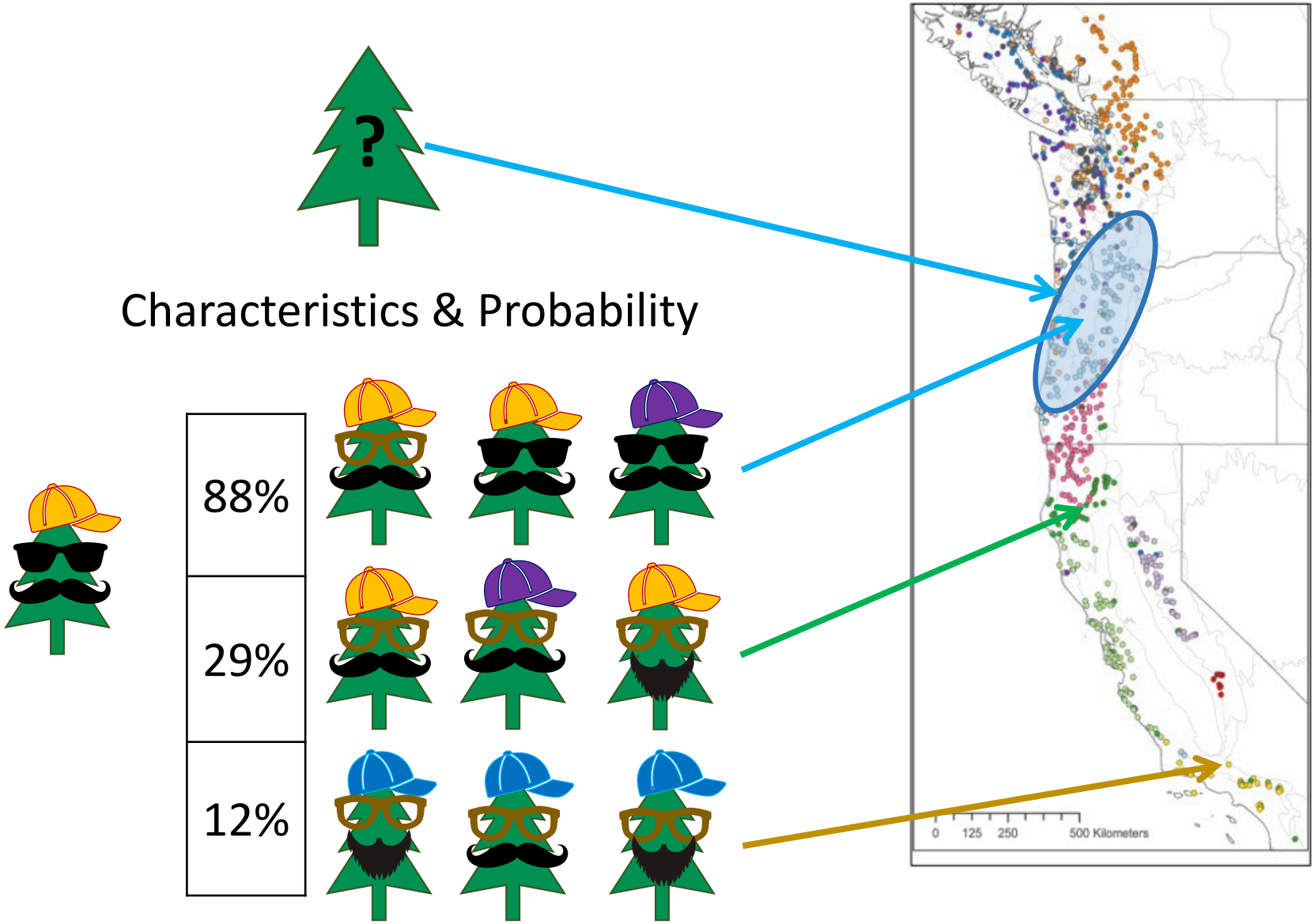
What is population assignment?





Cronn et al. 2021

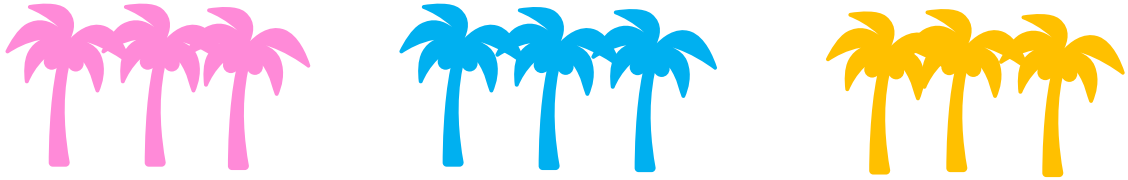
What is population assignment?



Cronn et al. 2021

Why population assignment?

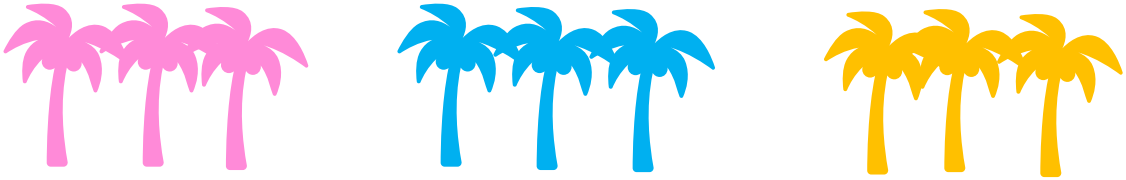
Illegal harvest



Goncalves Nazareno & Sedrez dos Reis 2013

Why population assignment?

Illegal harvest

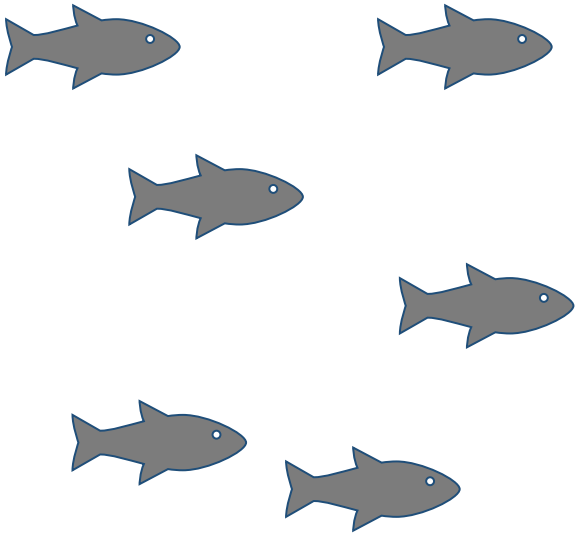


Goncalves Nazareno & Sedrez dos Reis 2013

Why population assignment?

Illegal harvest

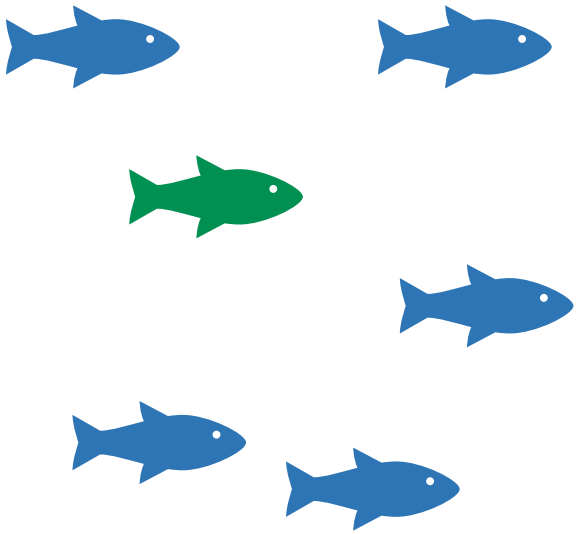
Conservation & Management



Why population assignment?

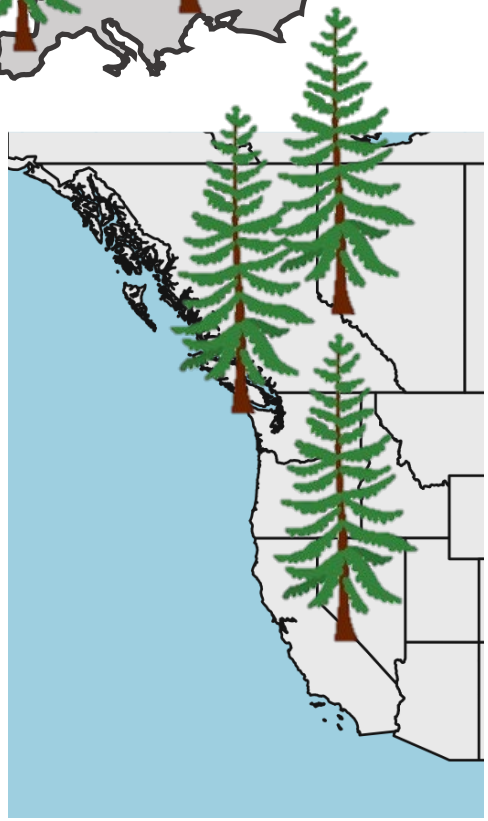
Illegal harvest

Conservation & Management



Why population assignment?

Illegal harvest



Conservation & Management

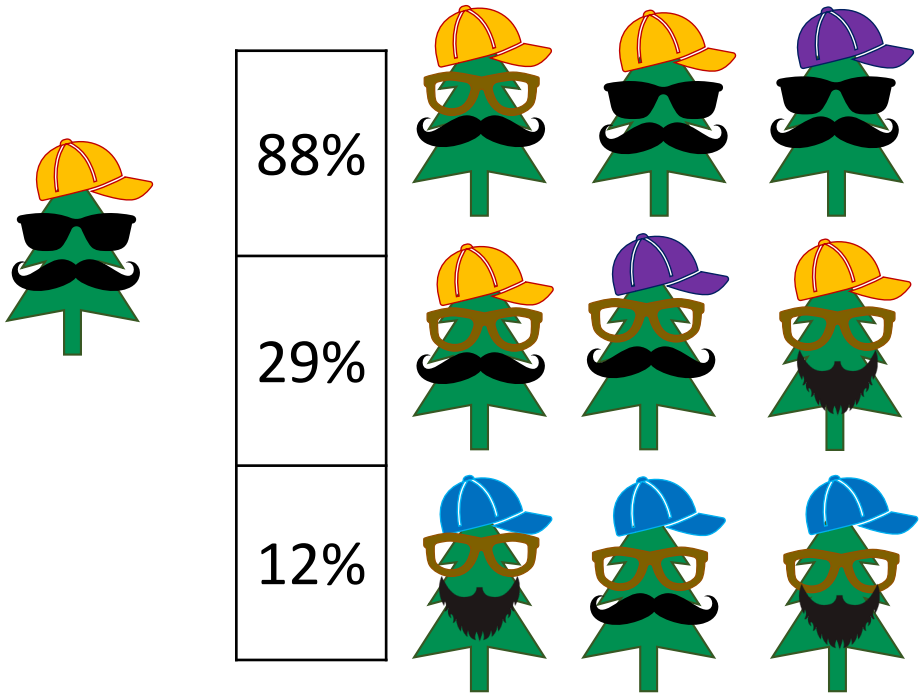
Seed provenance ID

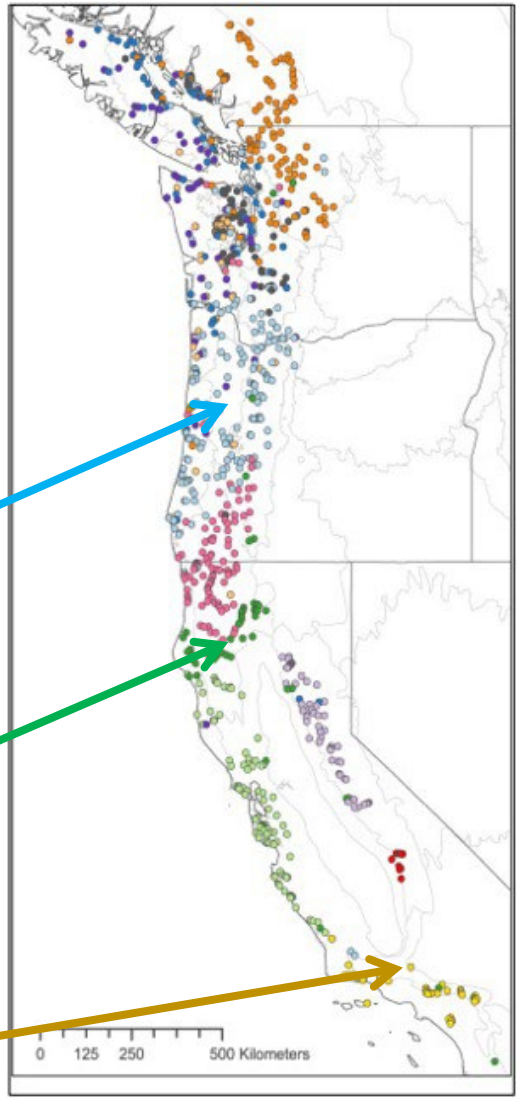
Hintsteiner et al. 2018

How to do population assignment?

What do we need?

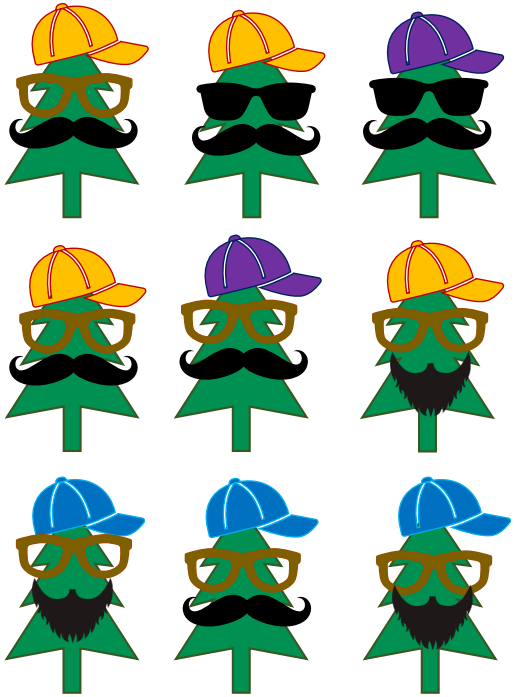
1. Spatially referenced genotypes of samples from all possible populations
2. Population structure





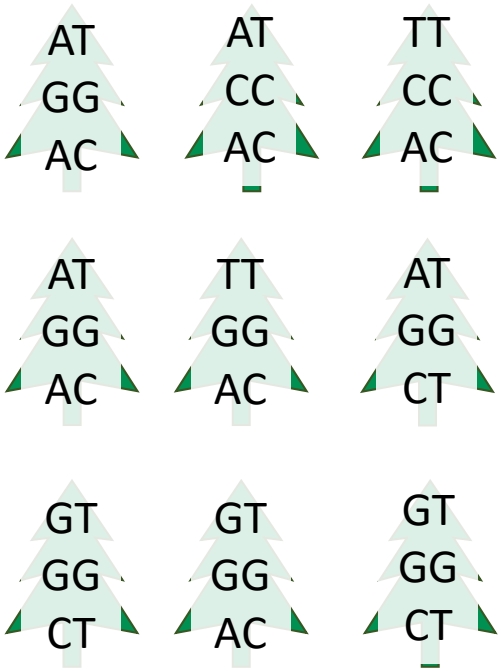
Cronn et al. 2021

Step 1: Develop genetic markers



Step 1: Develop genetic markers

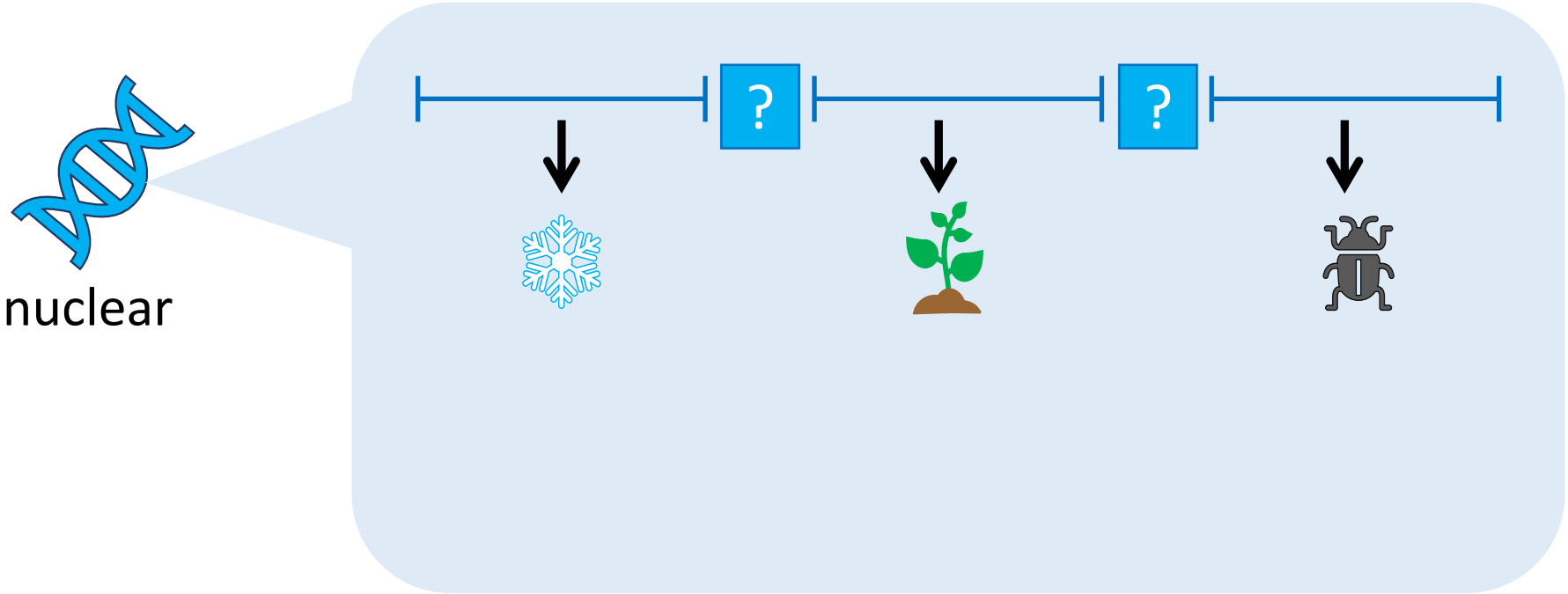
Use genetic markers

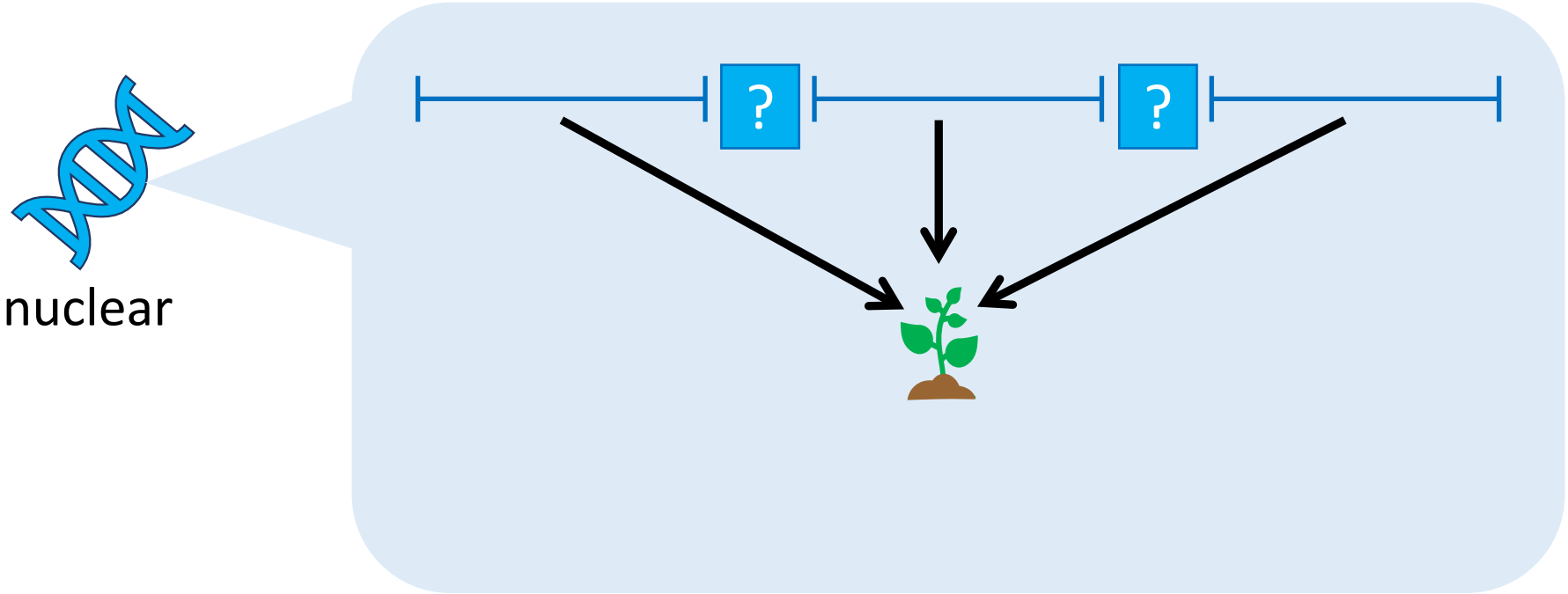


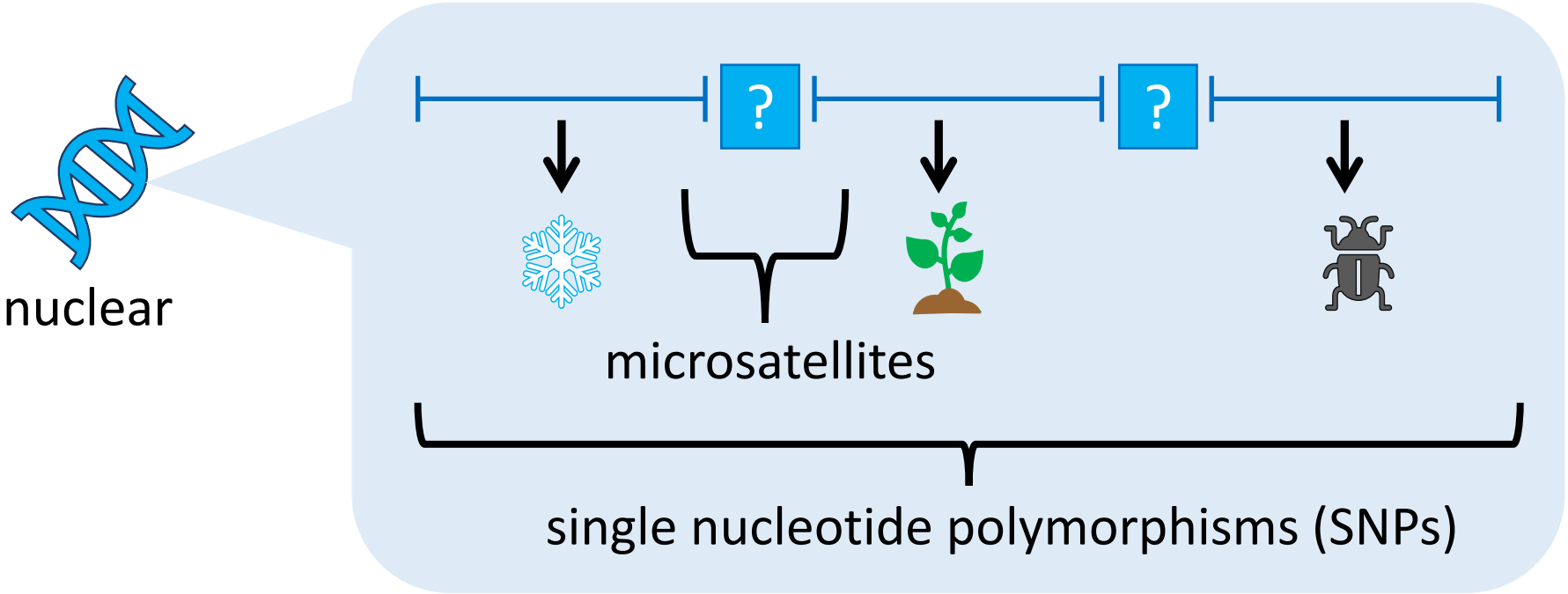
Highly variable to pick up differences

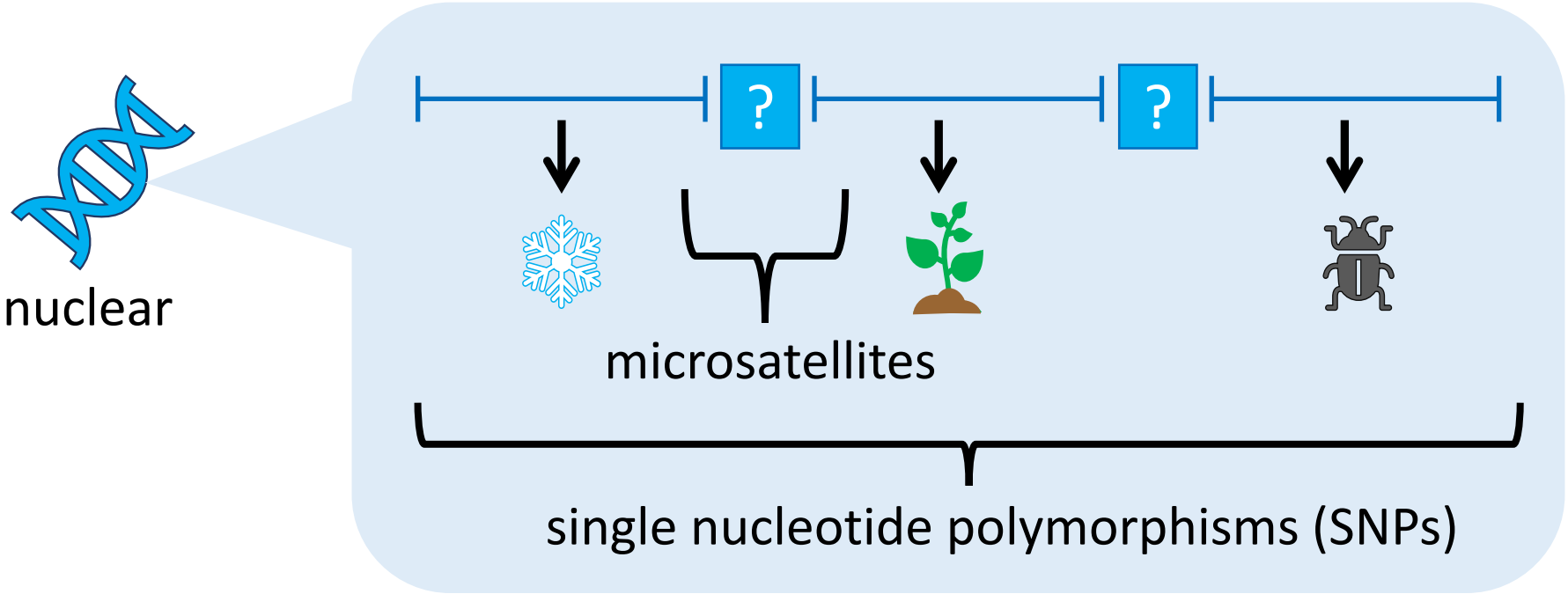
OR

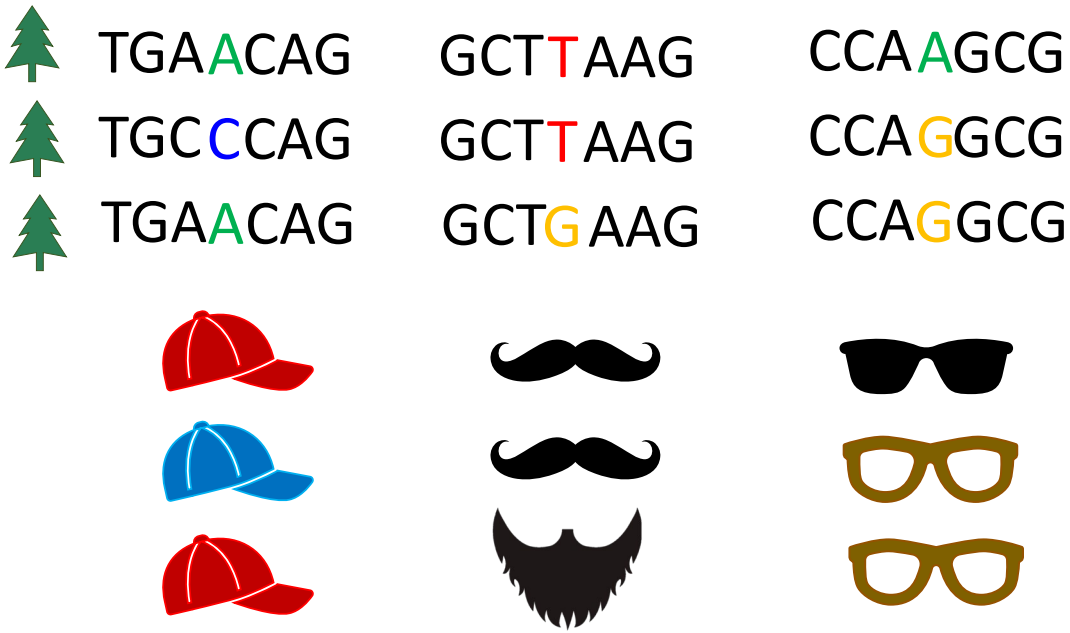
Low variability, but geographically restricted

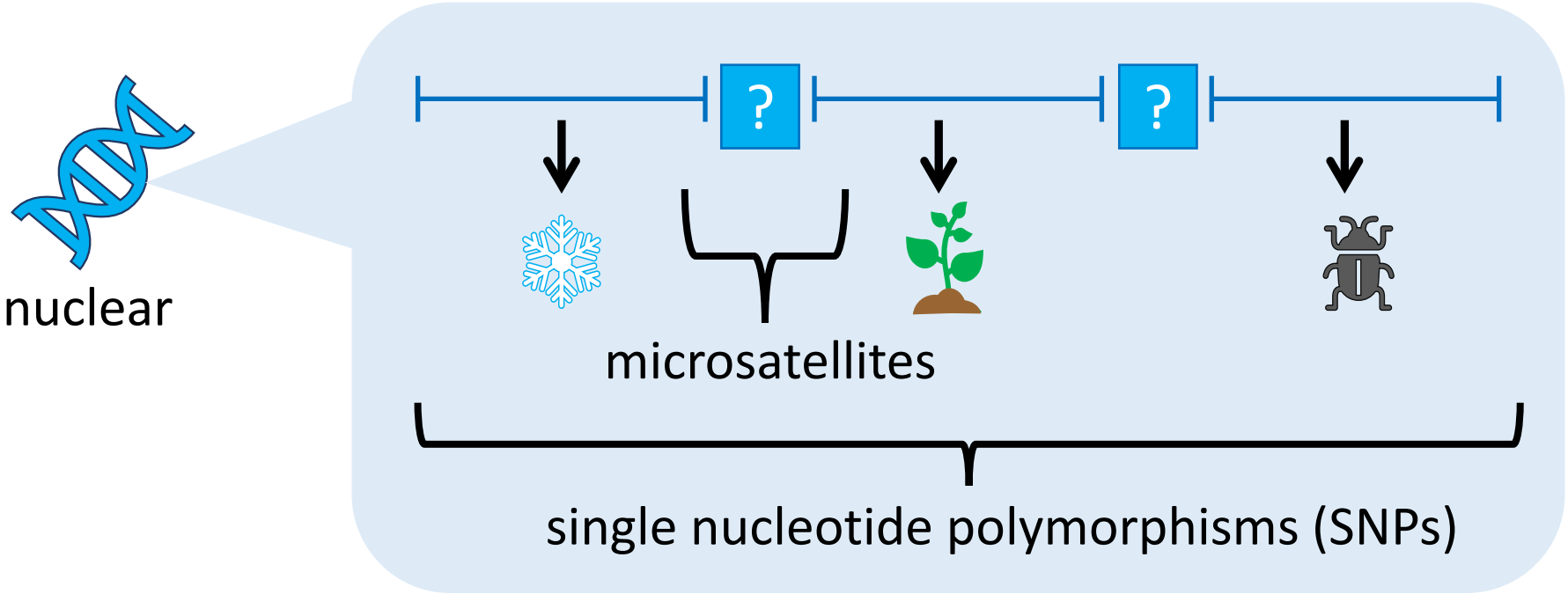

nuclear
?
?


nuclear
?
?


nuclear
?
?
microsatellites
single nucleotide polymorphisms (SNPs)

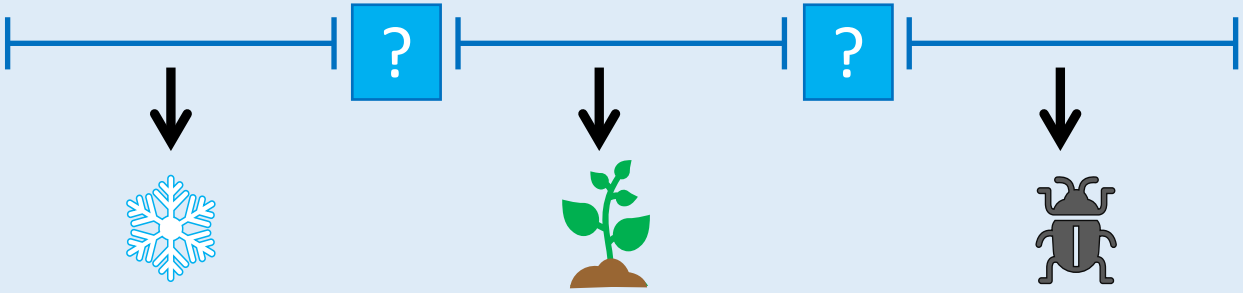

nuclear
?
?
microsatellites
single nucleotide polymorphisms (SNPs)


TGAACAG
GCTTAAG
CCAAGCG
TGCCCAG
GCTTAAG
CCAGGCG
TGAACAG
GCTGAAG
CCAGGCG



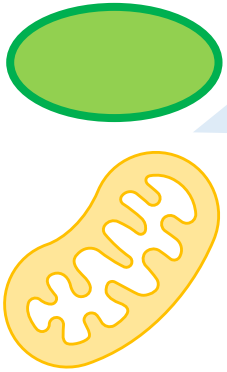
- Inherited from both parents
  - 2 alleles (Mom + Dad) ex. Aa or CC
- Species-specific
- Look at panels of 10s (msats) or 100s-1000s (SNPs)
- Highly variable (masts) or spread throughout genome (SNPs)

nuclear

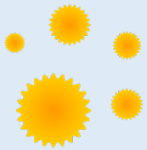


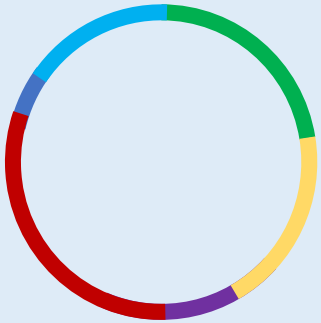
- Target 10s– 1000s
- 2 alleles
- Higher variability
- Low geographic restriction

organellular



or





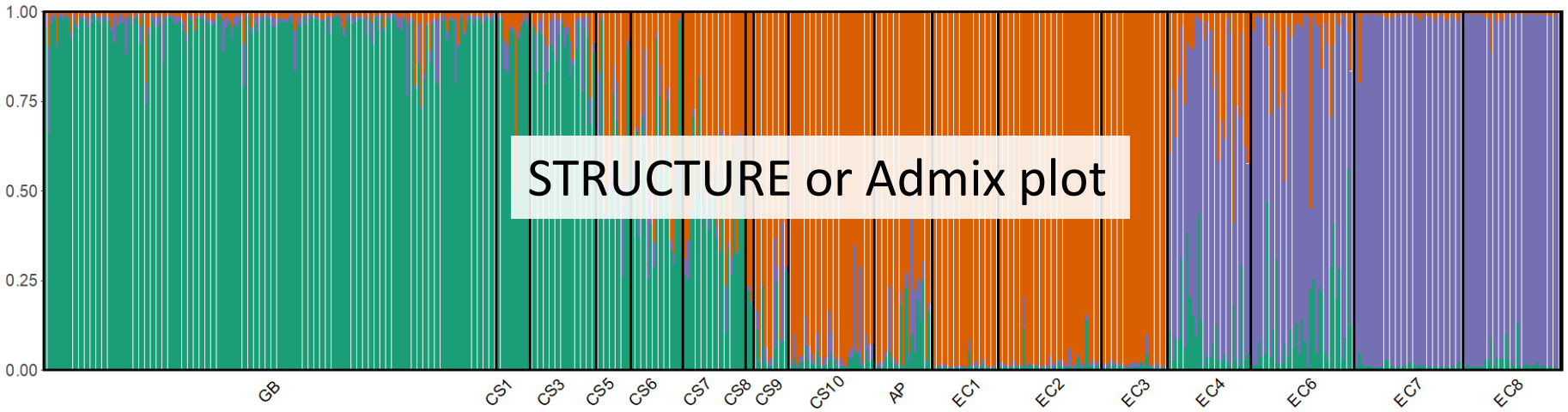
- Target & sequence one region
- One allele
- Lower variability
- Greater geographic restriction

Step 2: Assess population structure

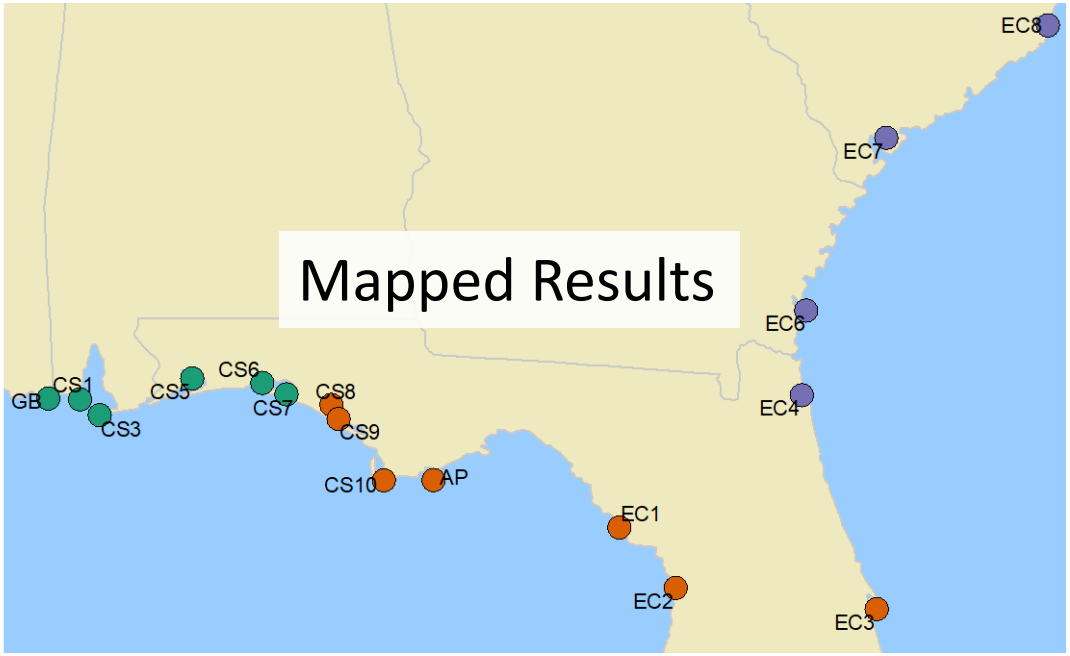
Use genet

Need to genotype all possible populations

Step 2: Assess population structure

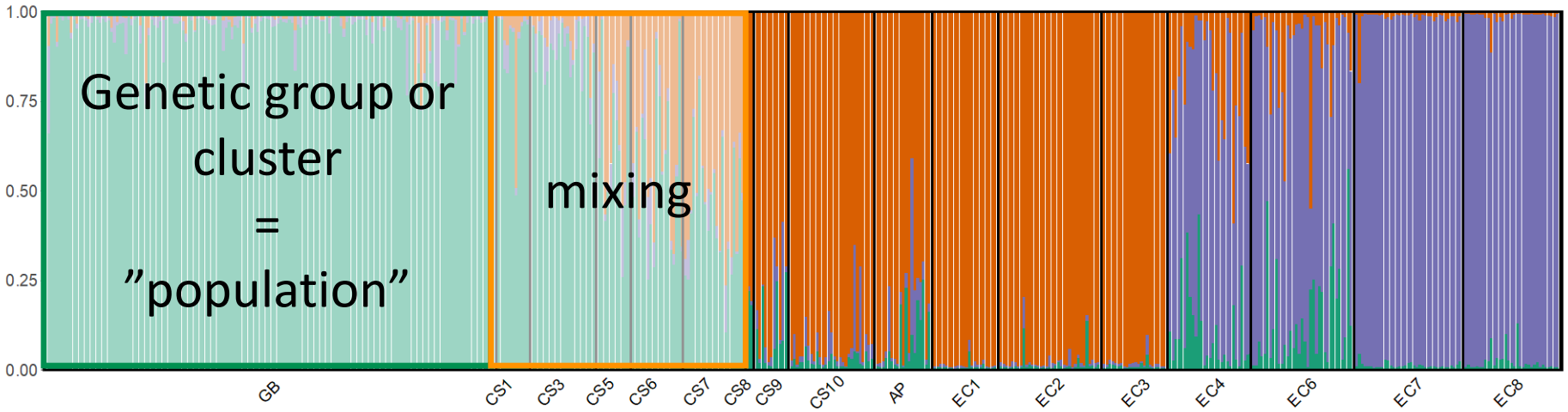


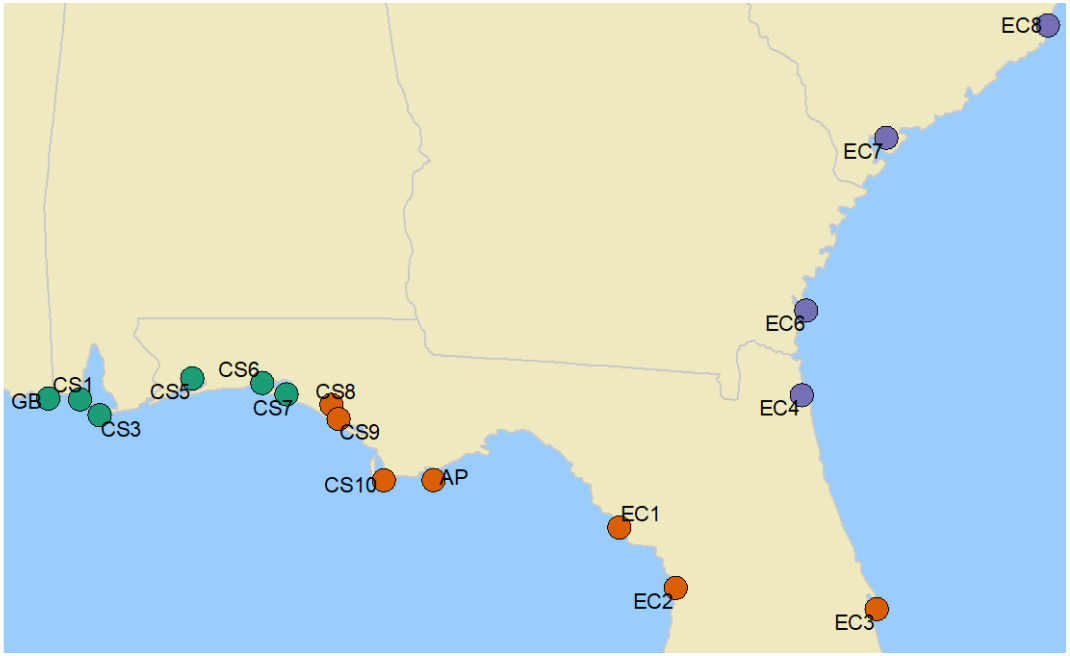
STRUCTURE or Admix plot



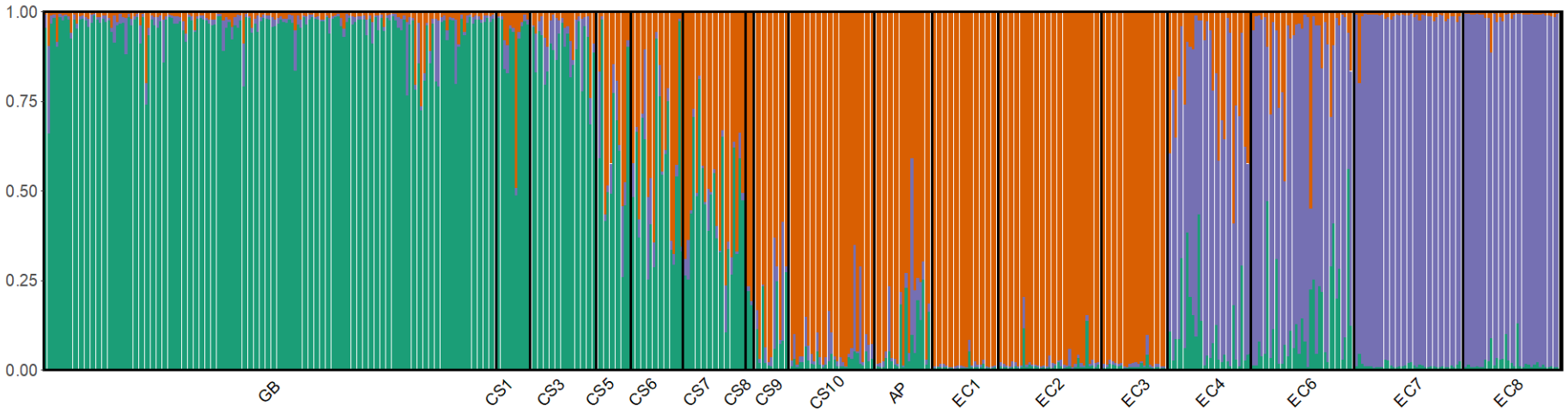
Mapped Results

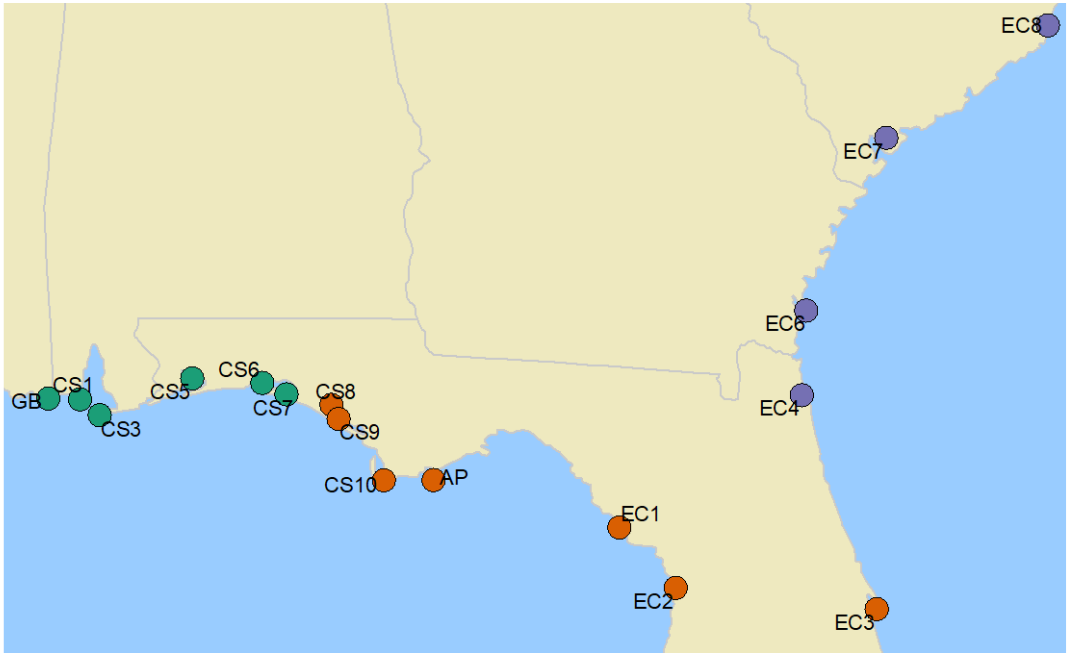
Step 2: Assess population structure



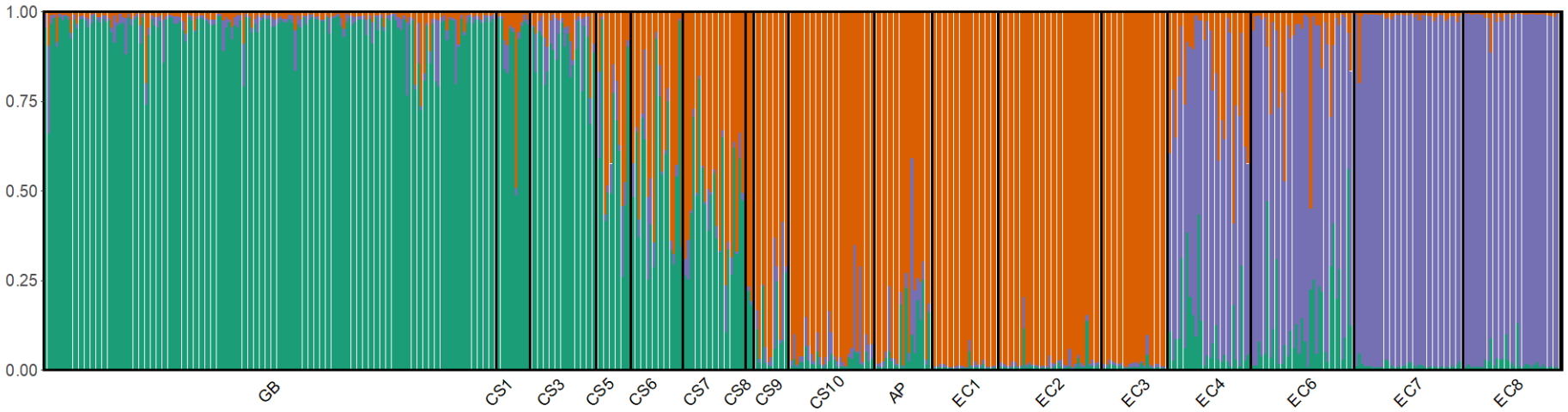


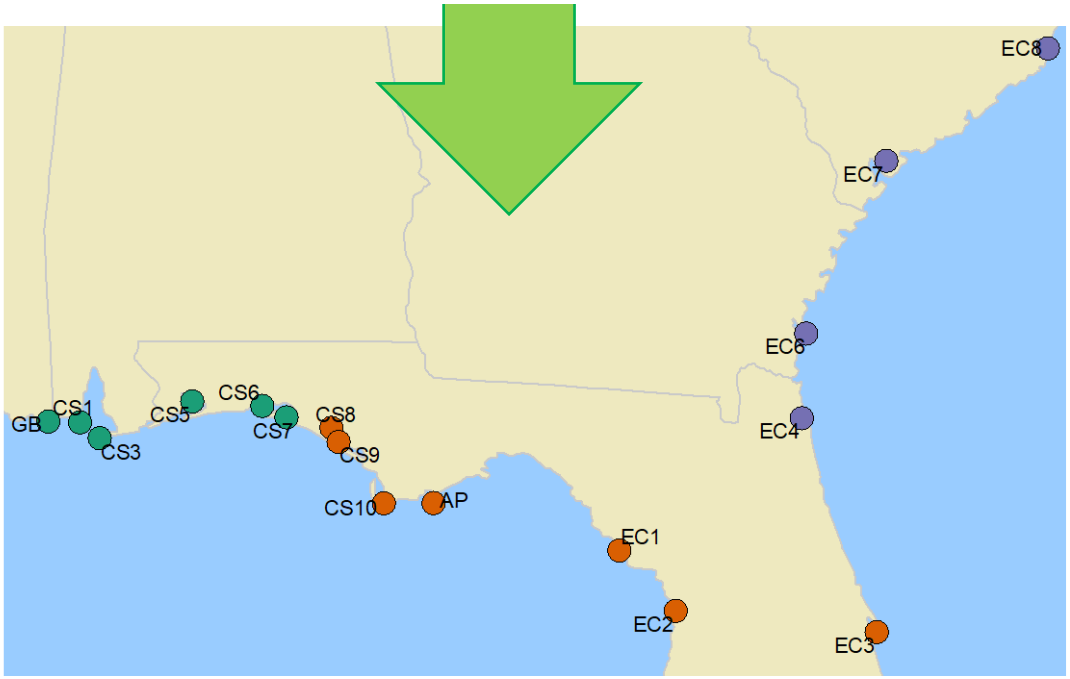
Step 2: Assess population structure





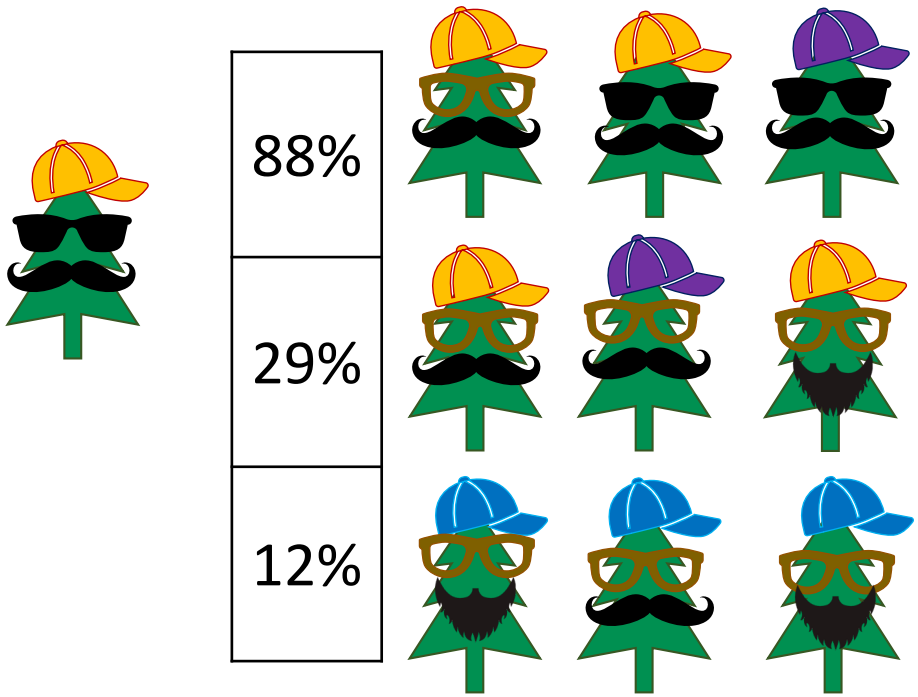
Step 2: Assess population structure



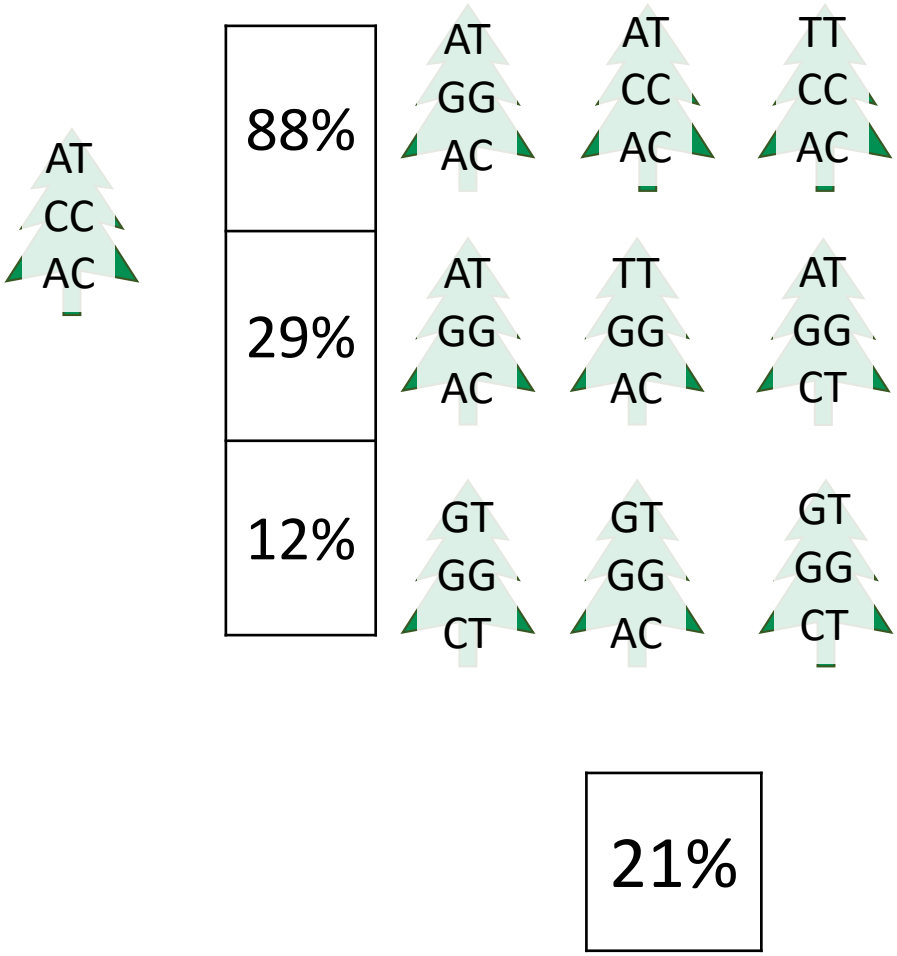


Step 2: Assess population structure

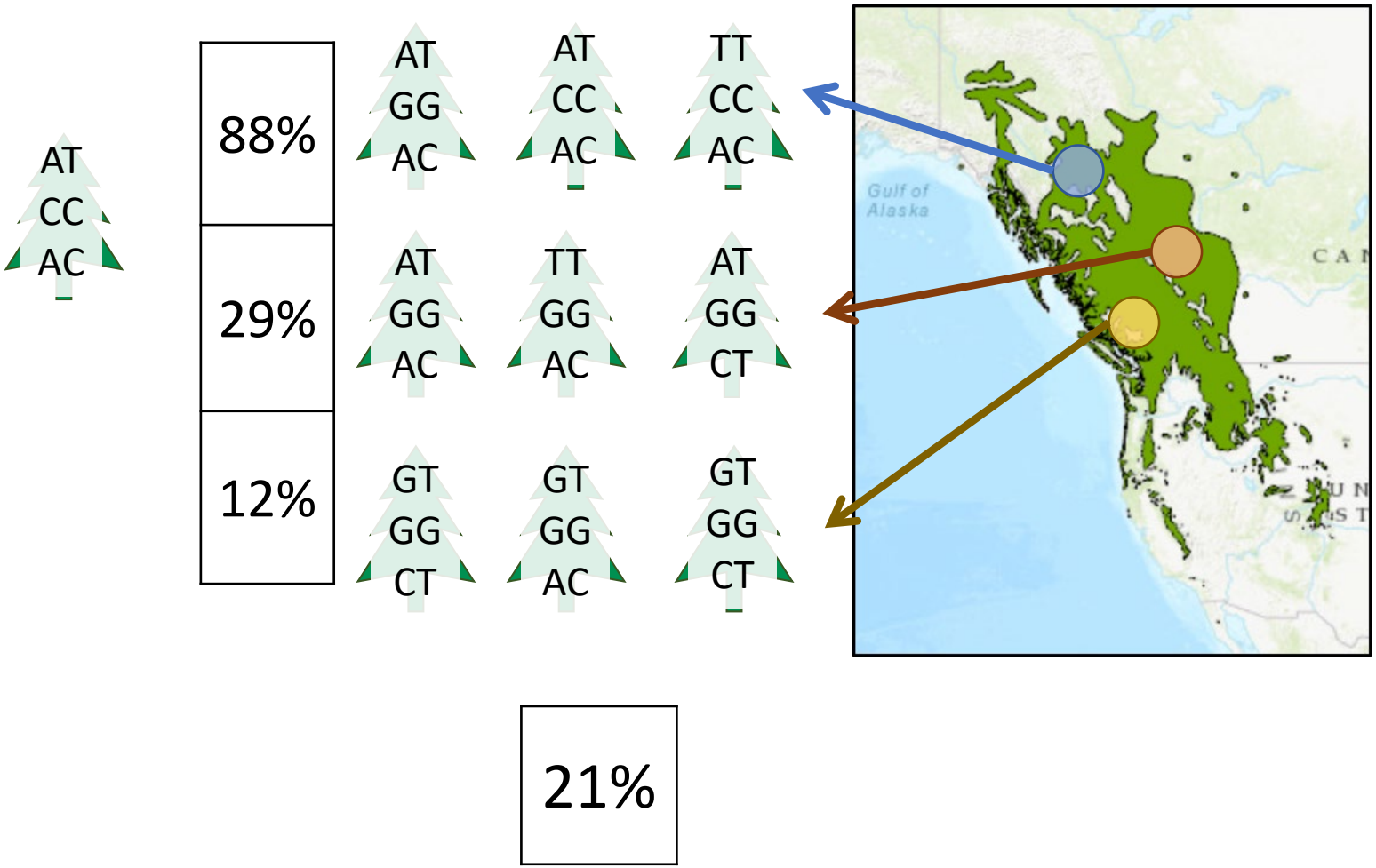
Step 3: Assign your sample



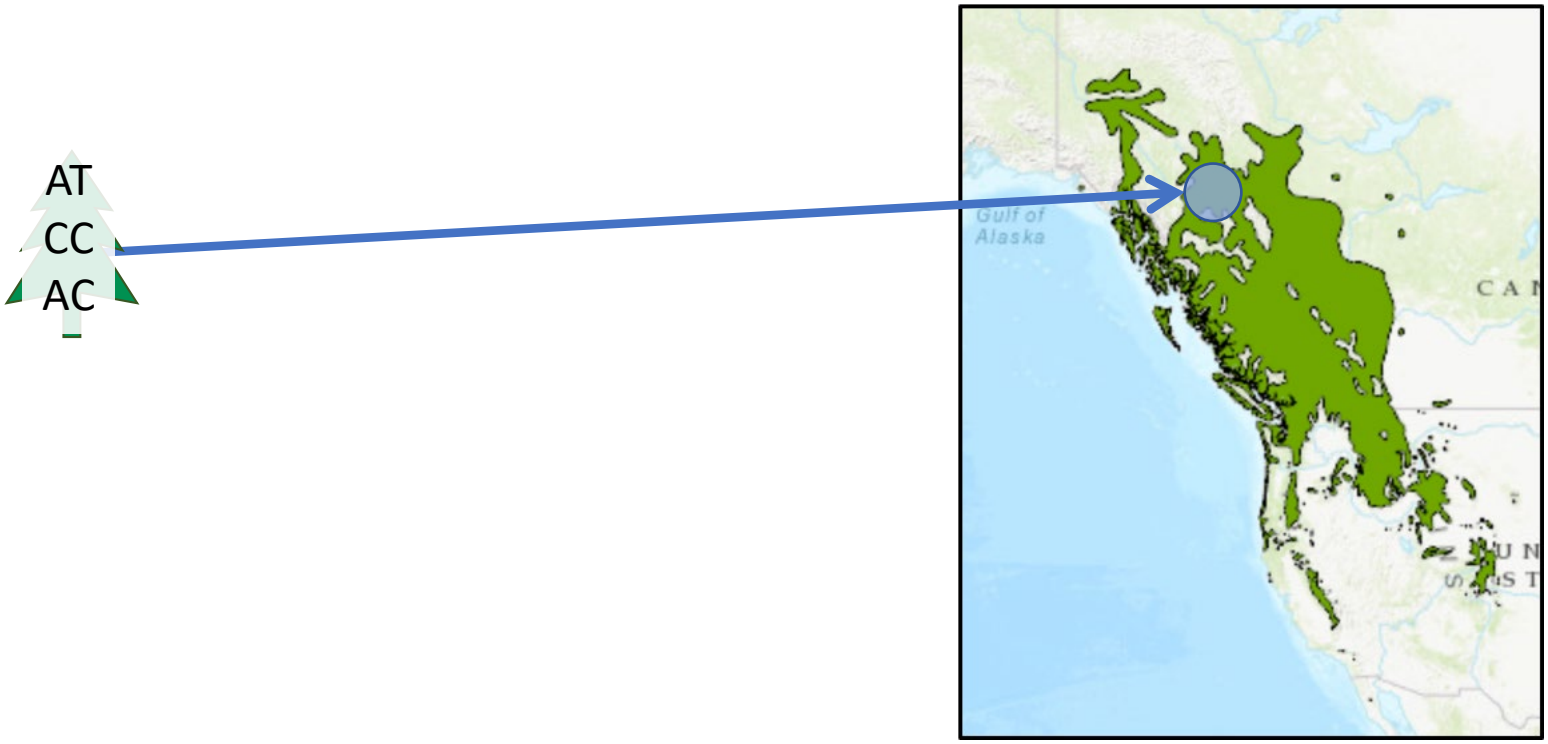
Step 3: Assign your sample



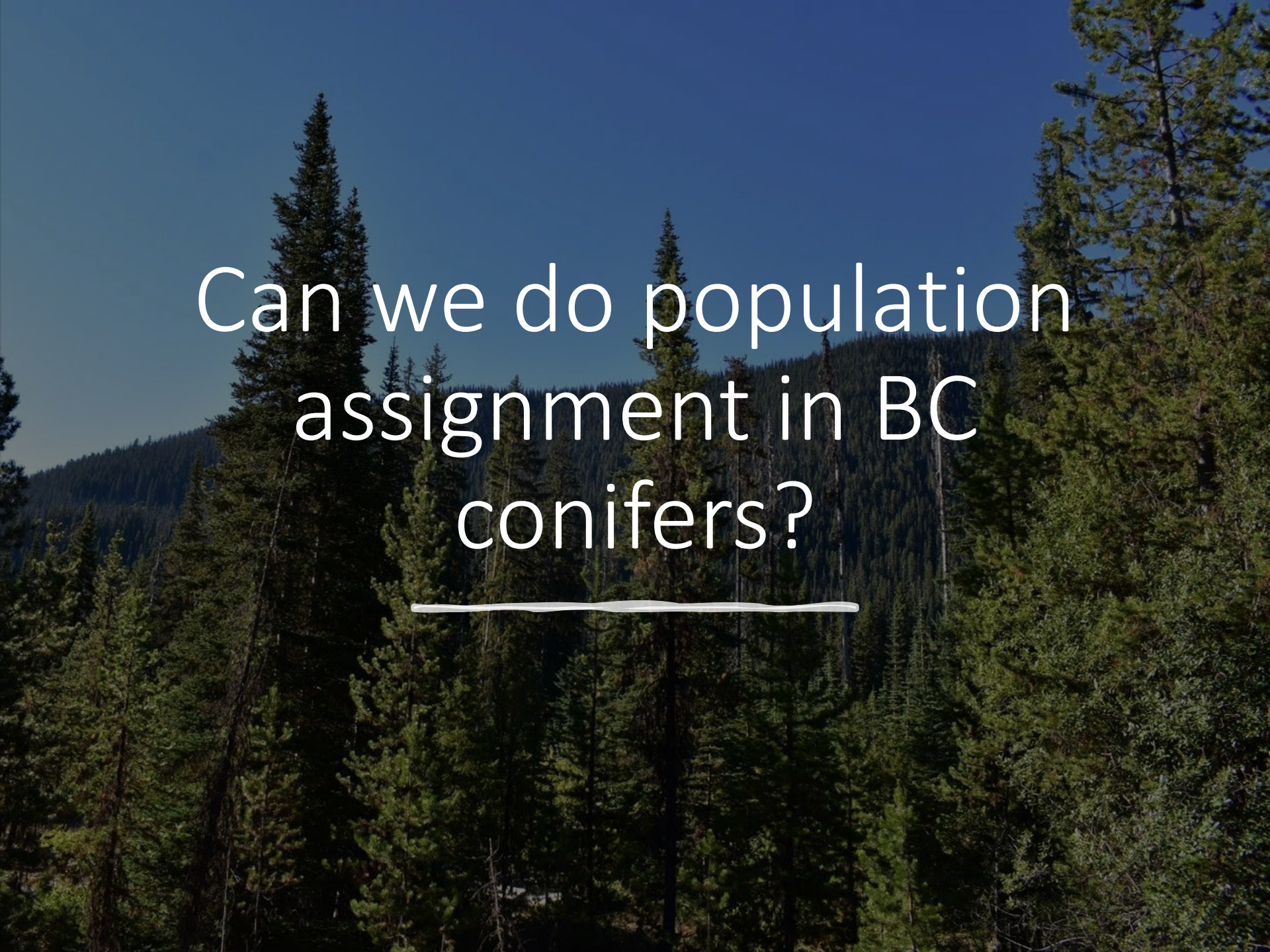
Step 3: Assign your sample



Step 3: Assign your sample



21%



Can we do population assignment in BC conifers?

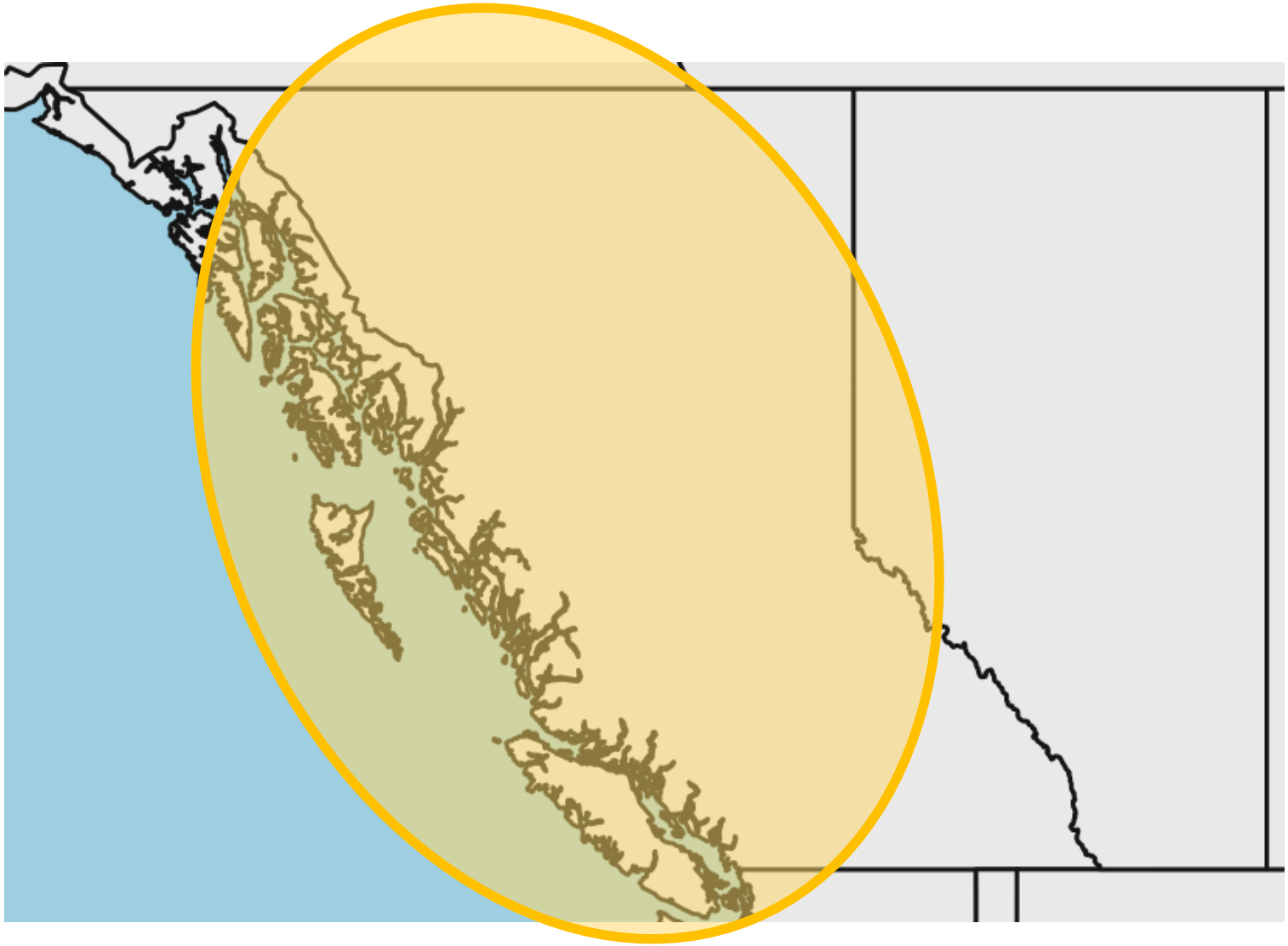
It depends:

- Scale of interest
- Life history

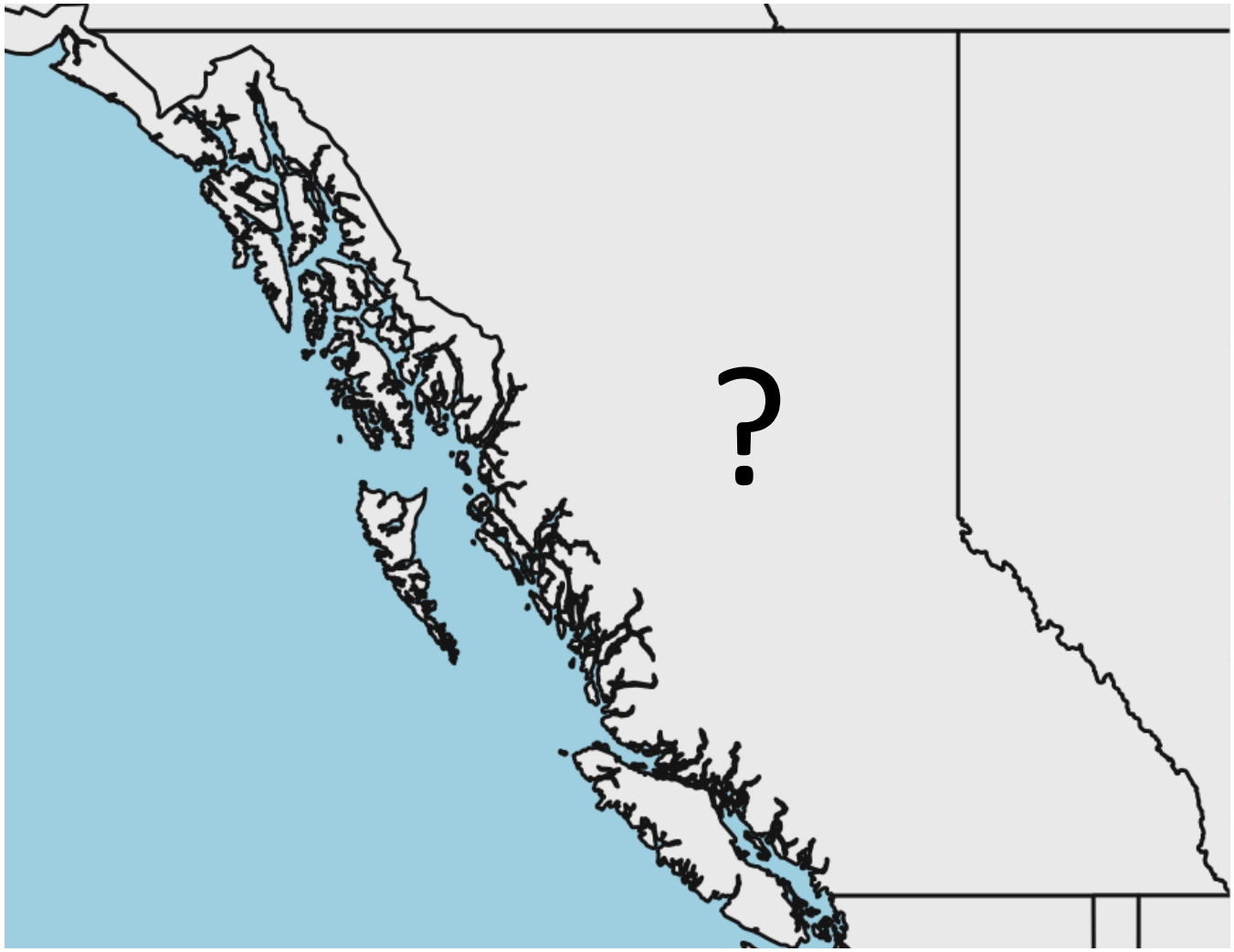
How close can you get?

How close can you get?

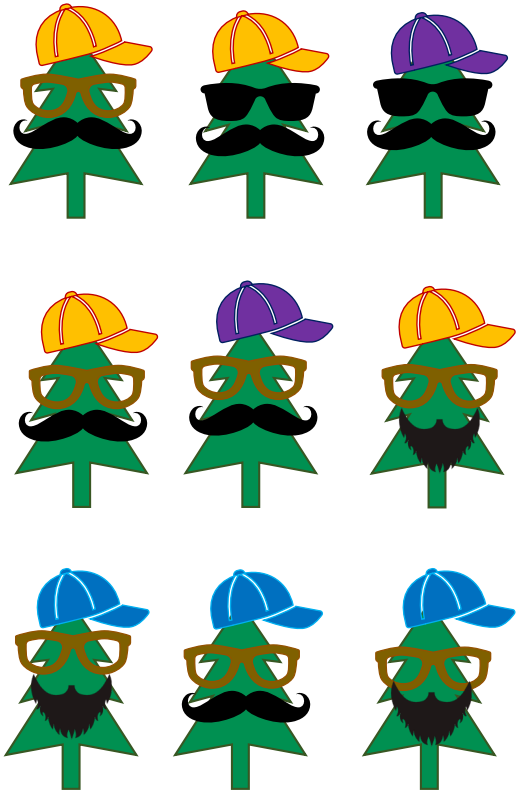
How close can you get?

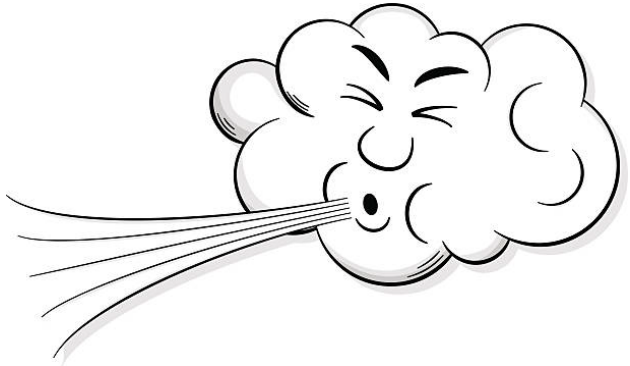


How close can you get?

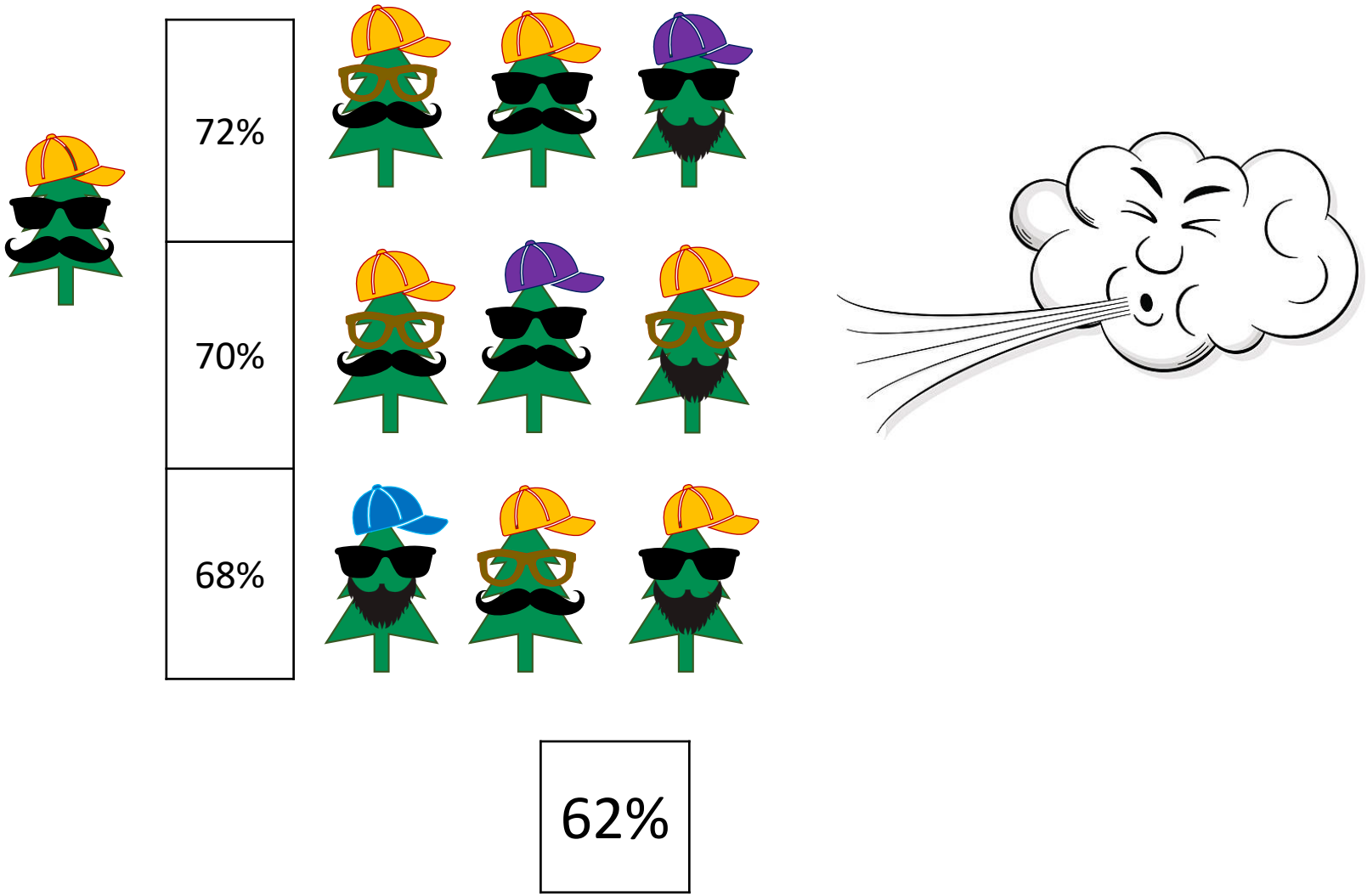


Conifer Life History

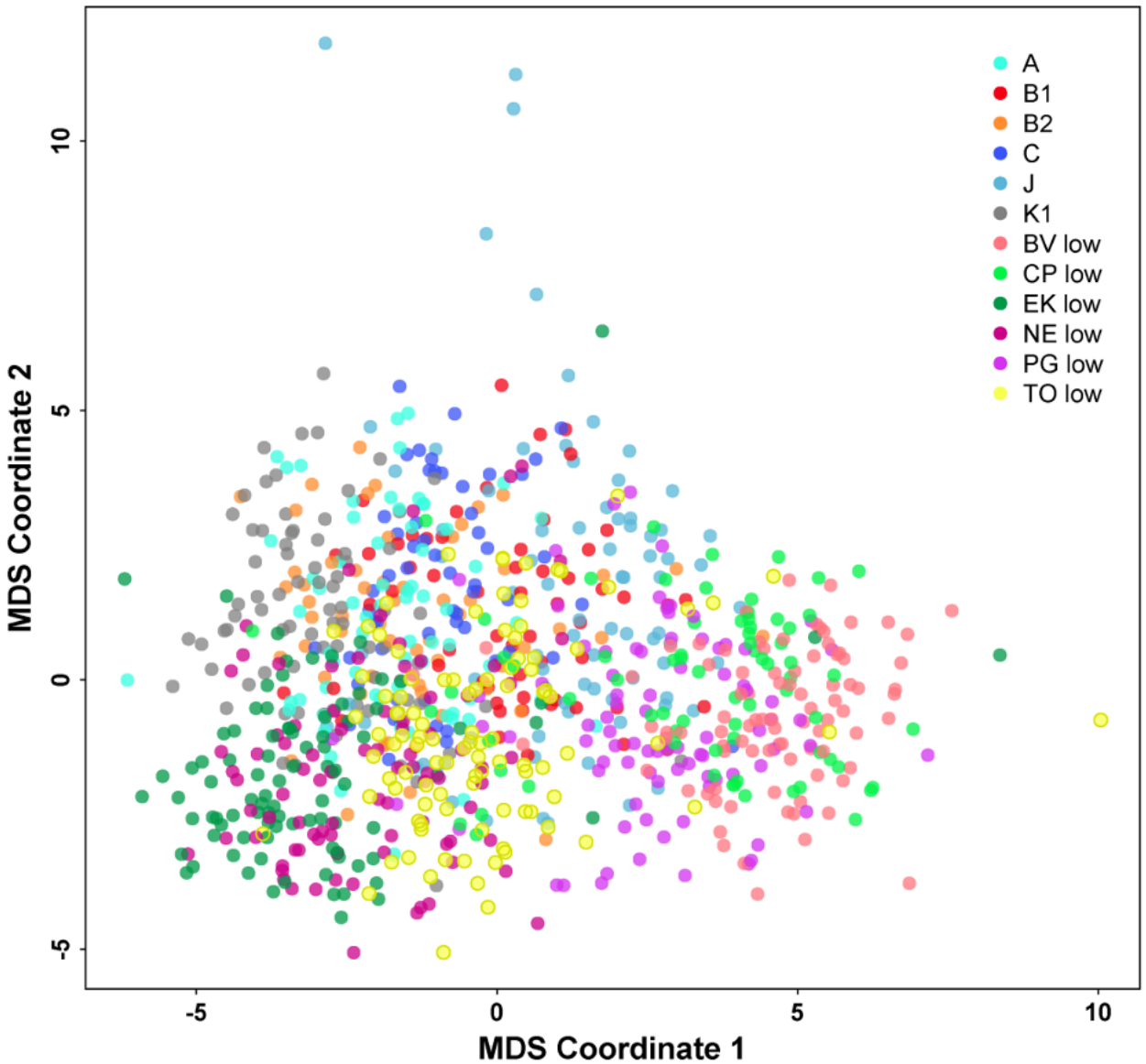




Conifer Life History

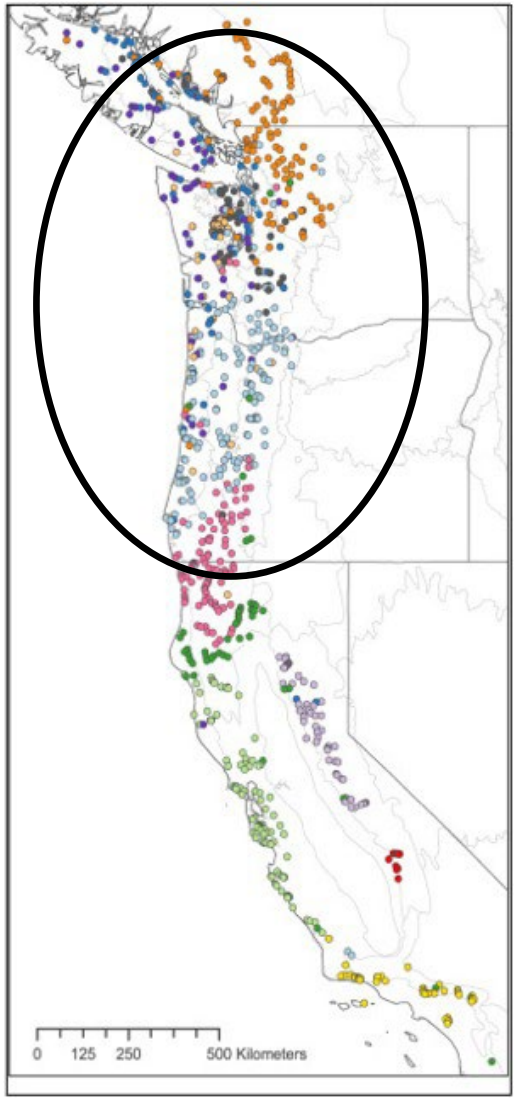


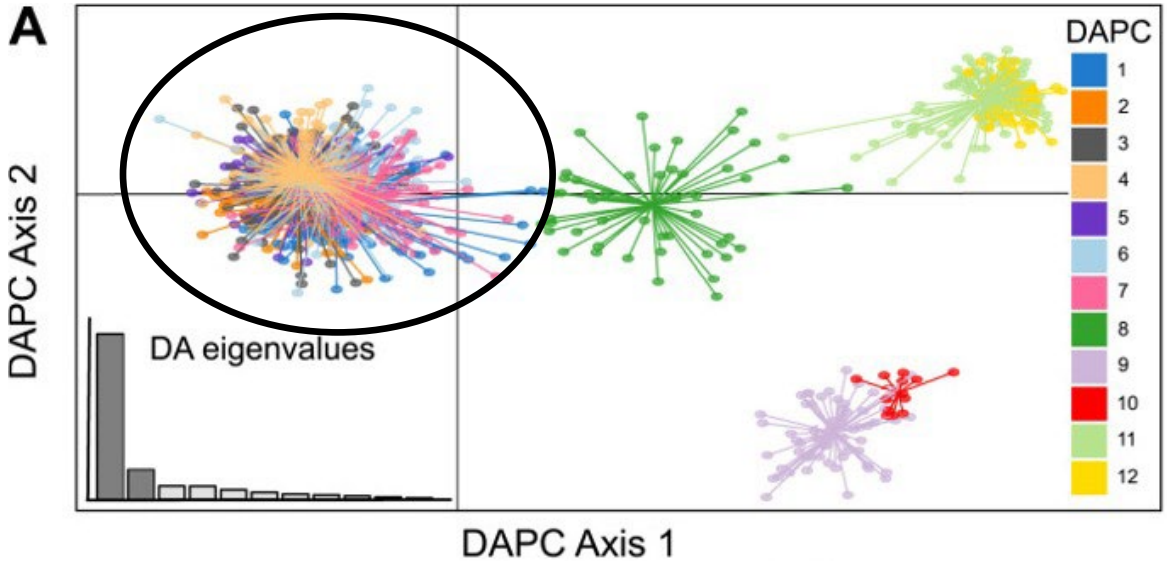
Conifer Life History



MacLachlan et al. 2021

Tree Life History: Big Leaf Maple





50.3% classification error

=

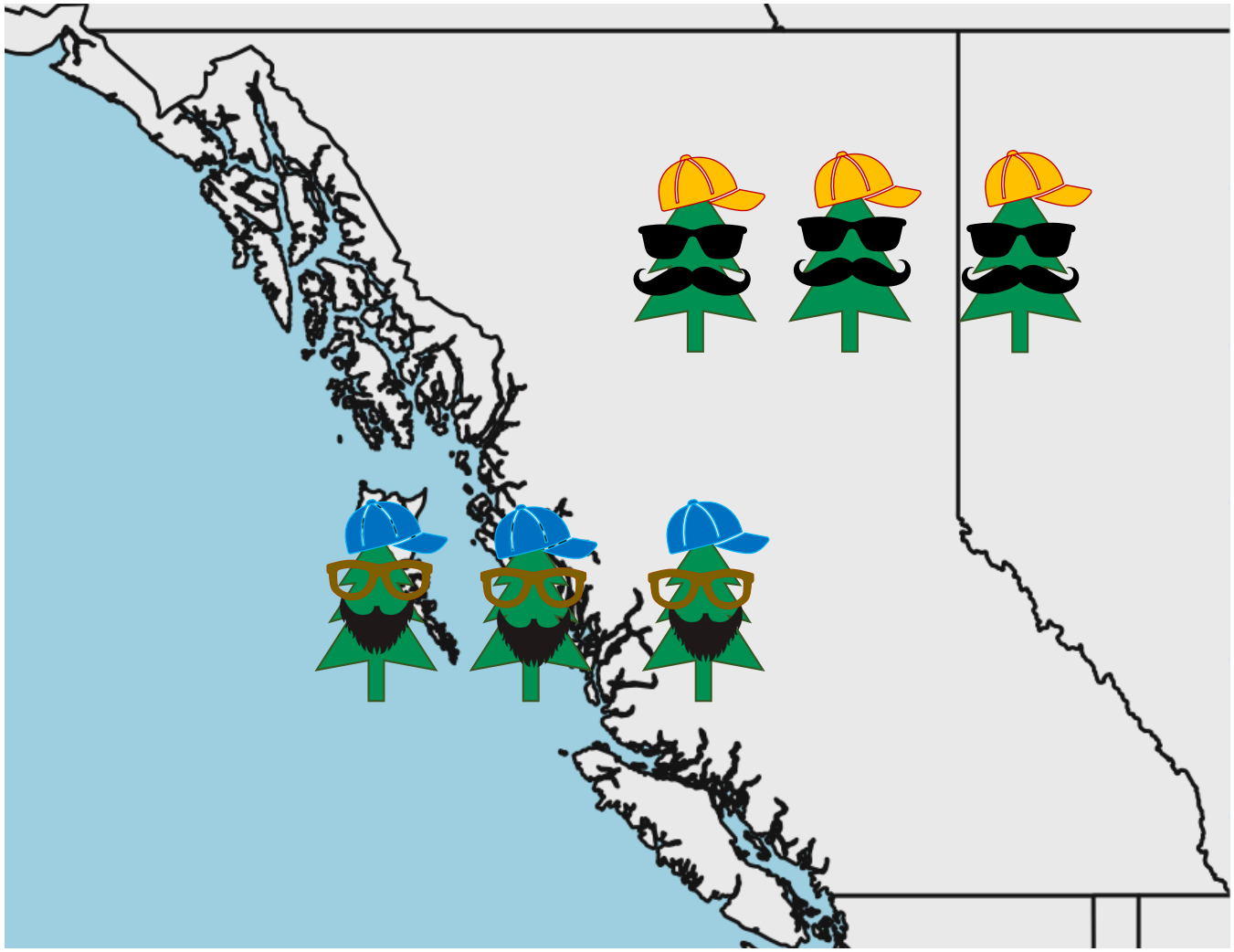
Incorrect geographic assignment 50% of the time

Cronn et al. 2021

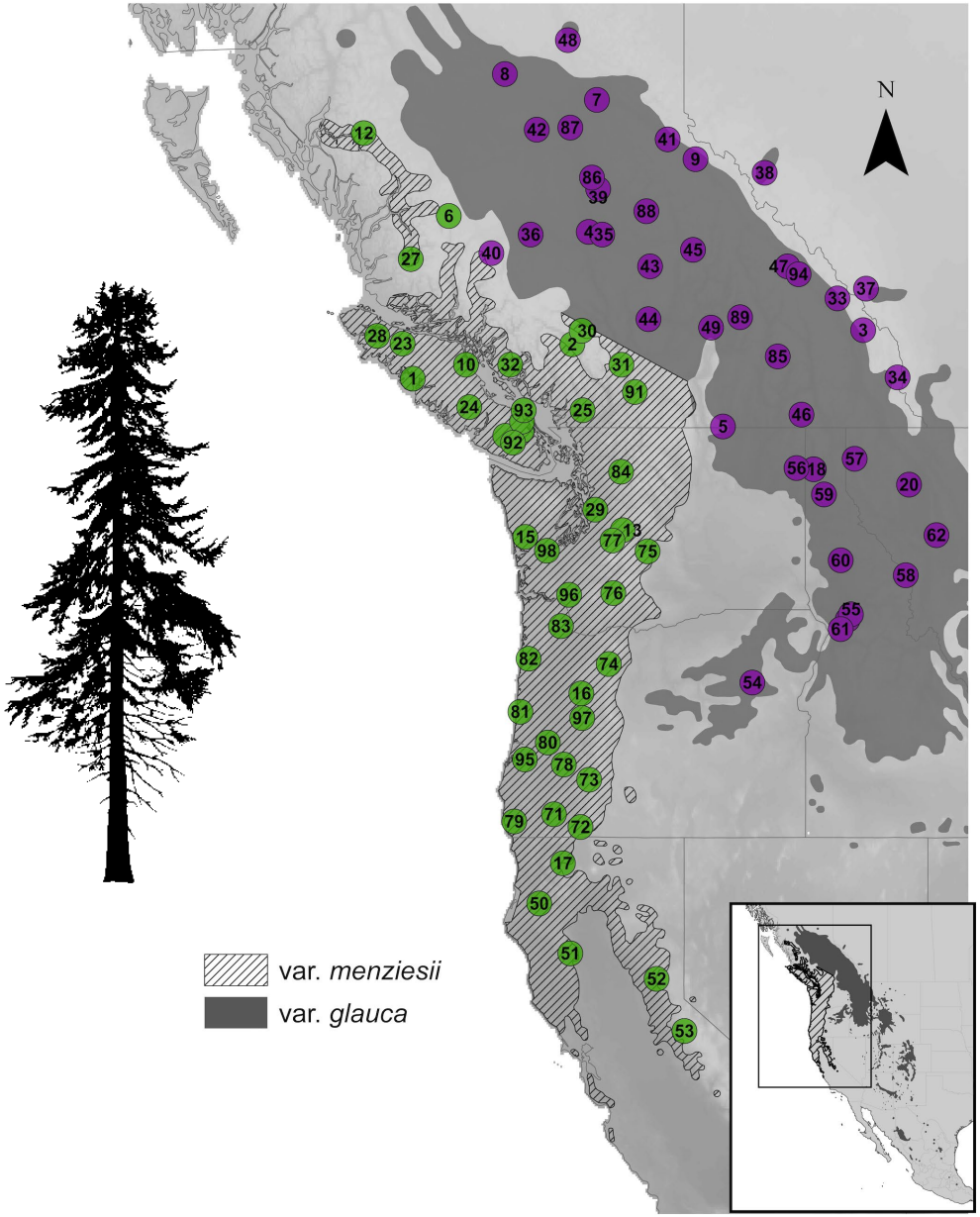
So, can we do population assignment in BC conifers?

it depends…

Different Varieties



Different Varieties: Douglas-fir



CoAdapTree

Healthy Trees for Future Climates

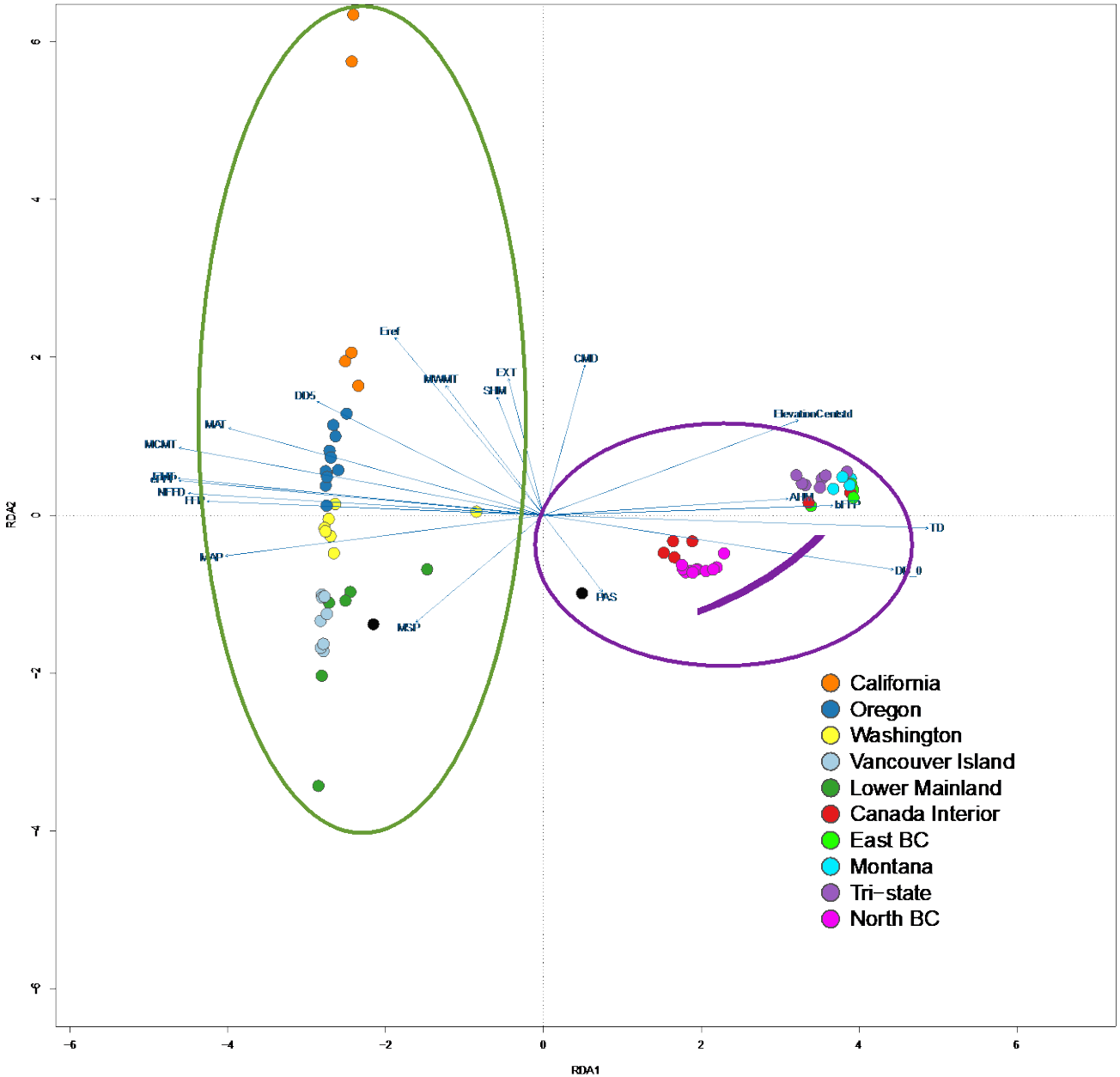
Rafael Candido-Ribeiro
UBC

Brandon Lind
UBC

Sally Aitken
UBC

Sam Yeaman
UCalgary

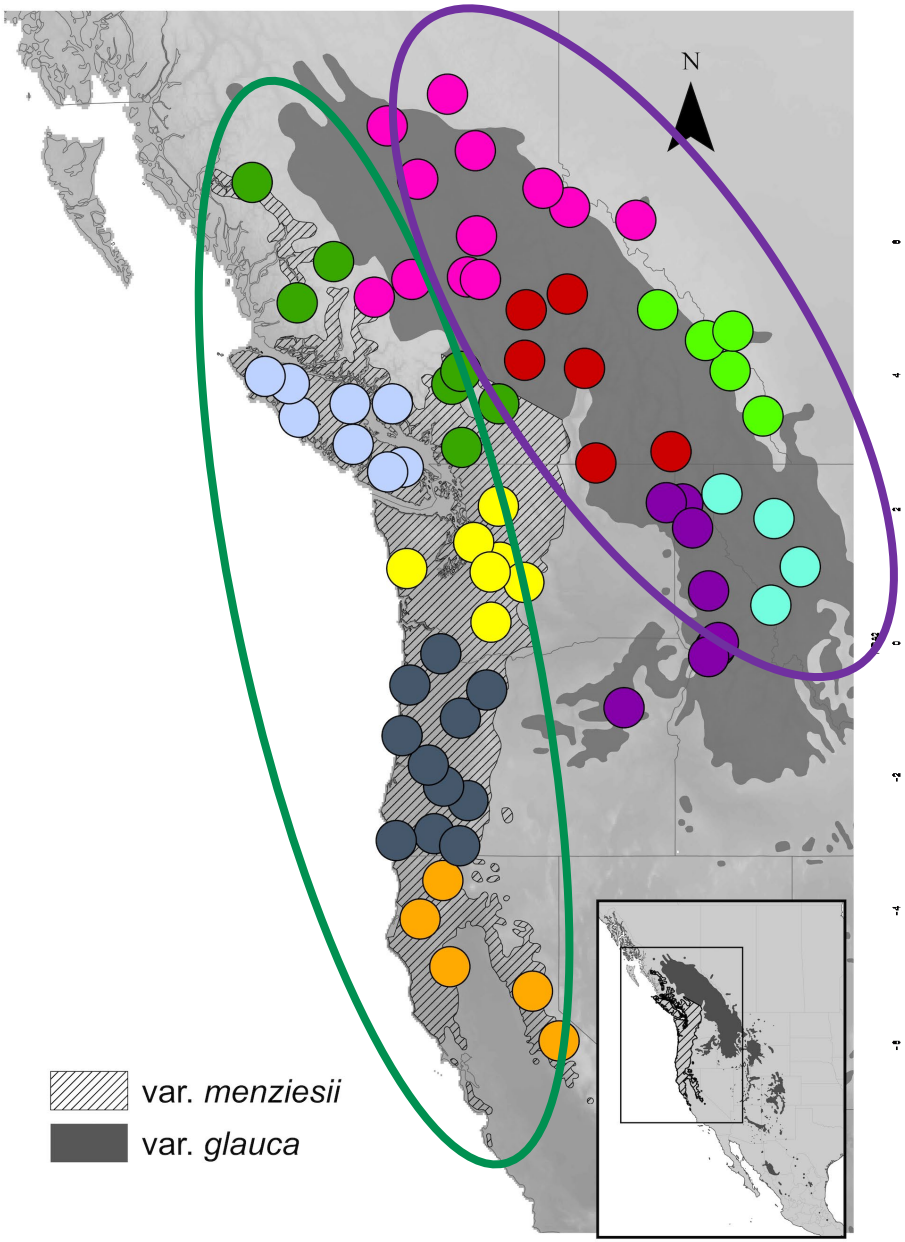
Different Varieties: Douglas-fir

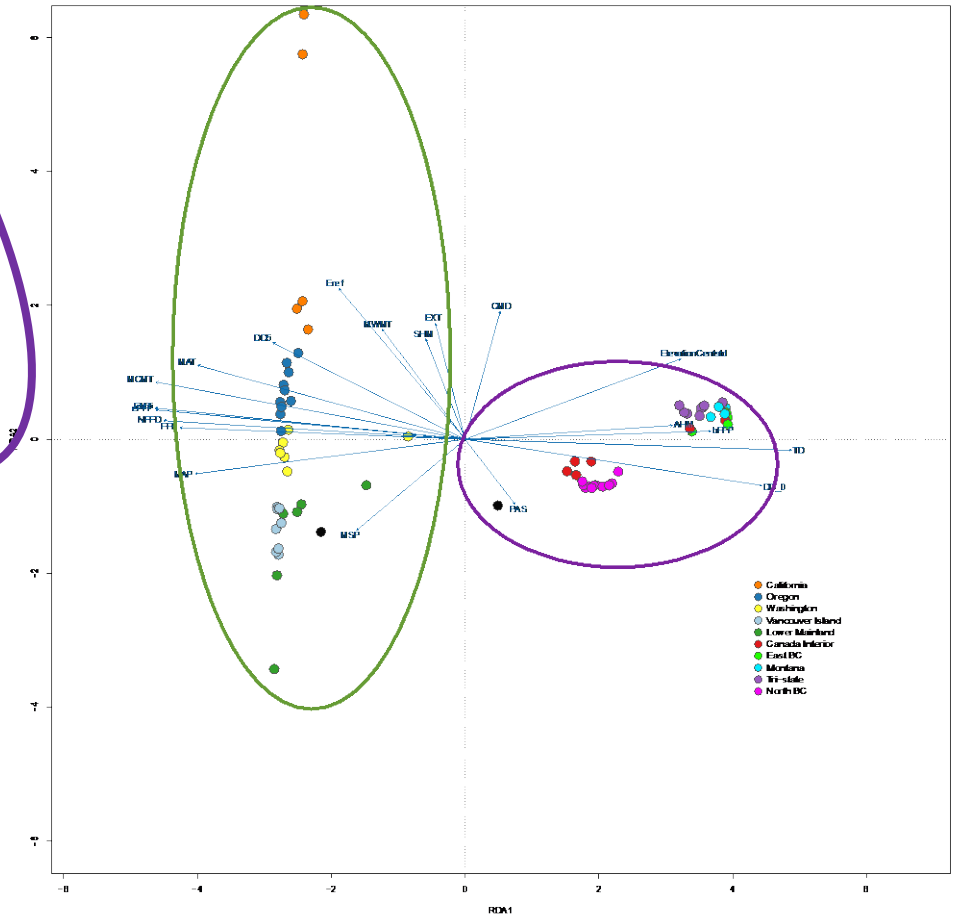


coastal

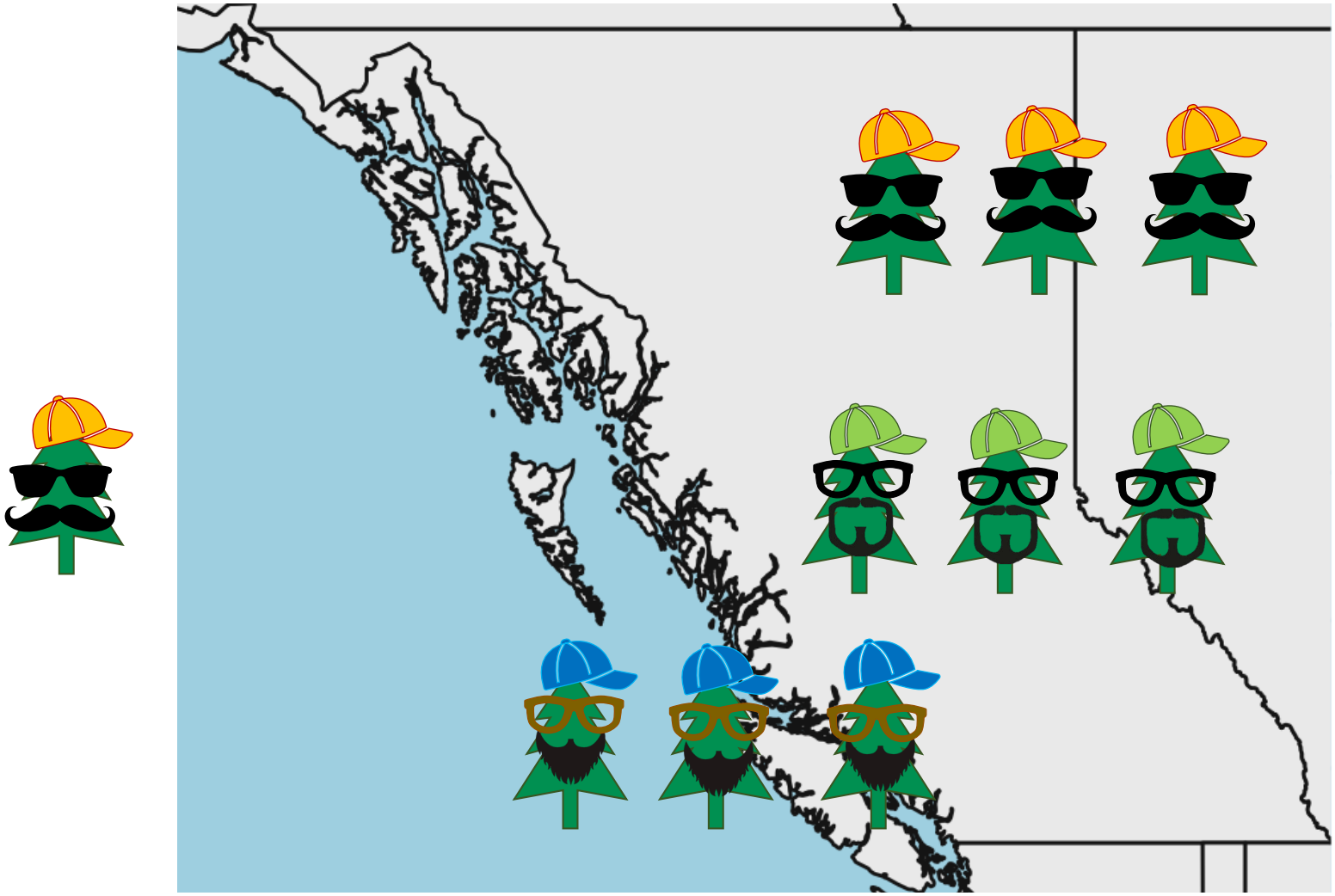
interior

Different Varieties: Douglas-fir

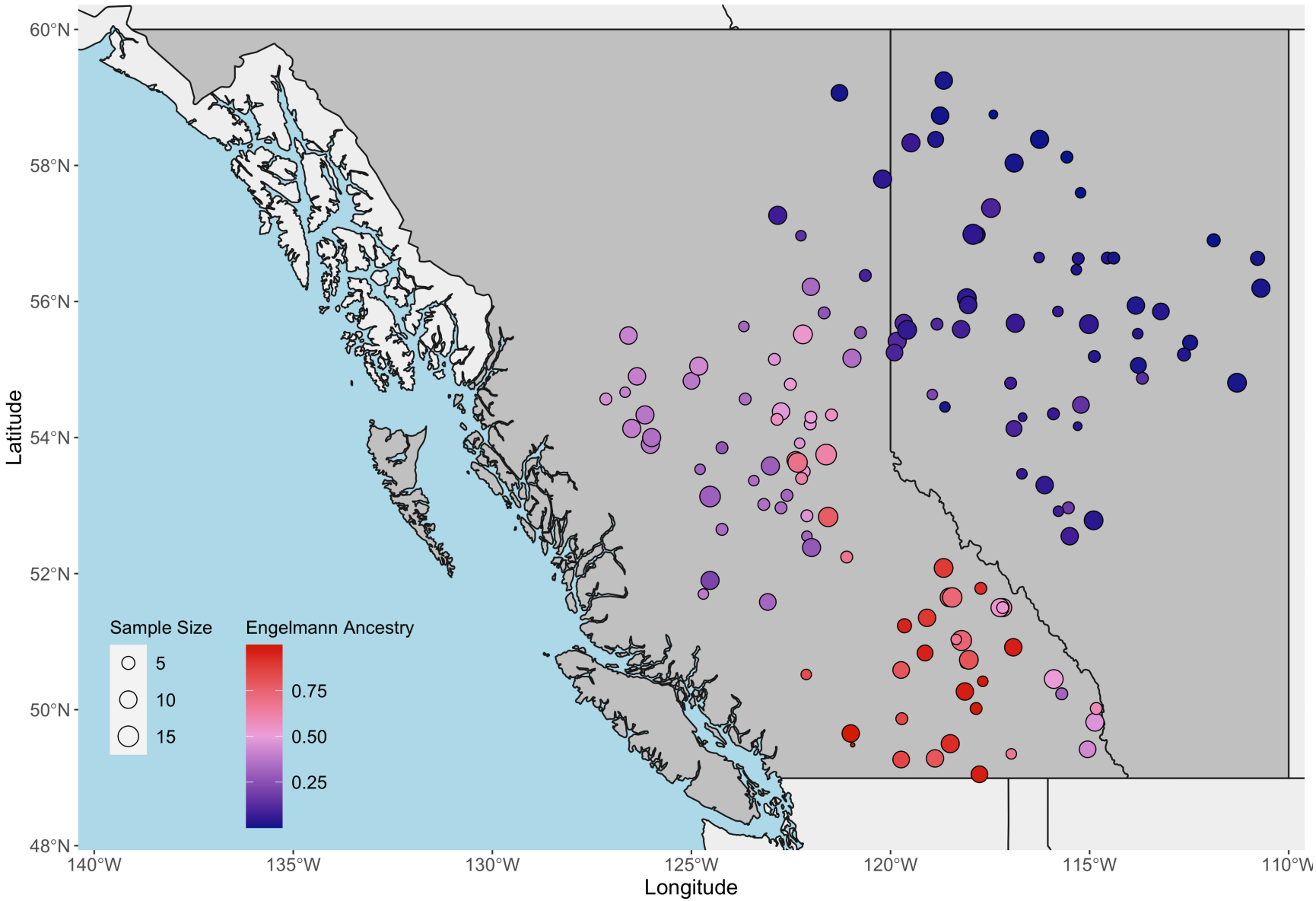




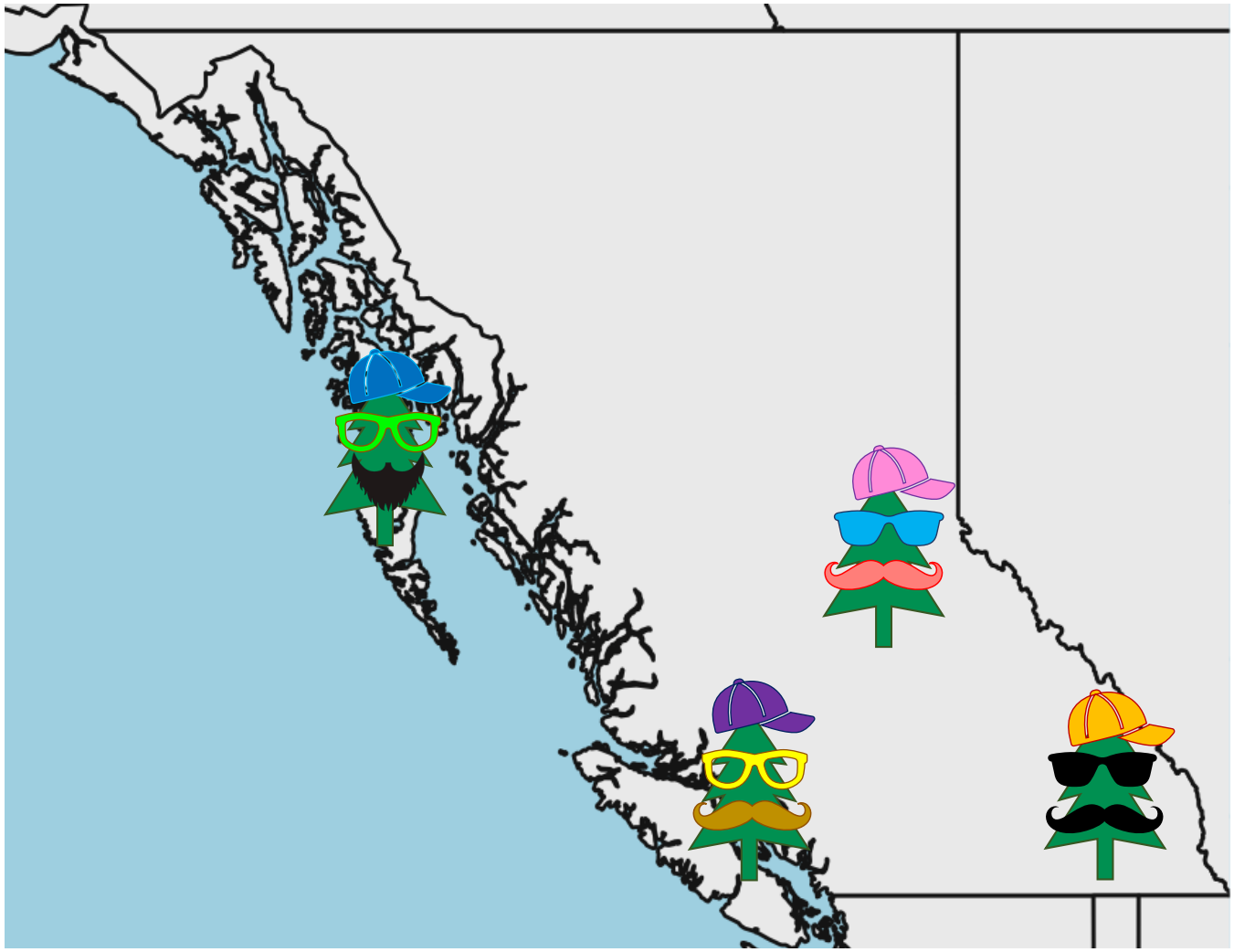
Hybridization



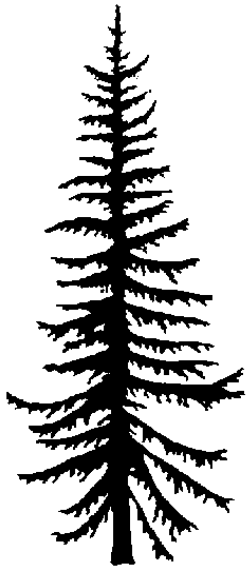
Hybridization: Interior Spruce

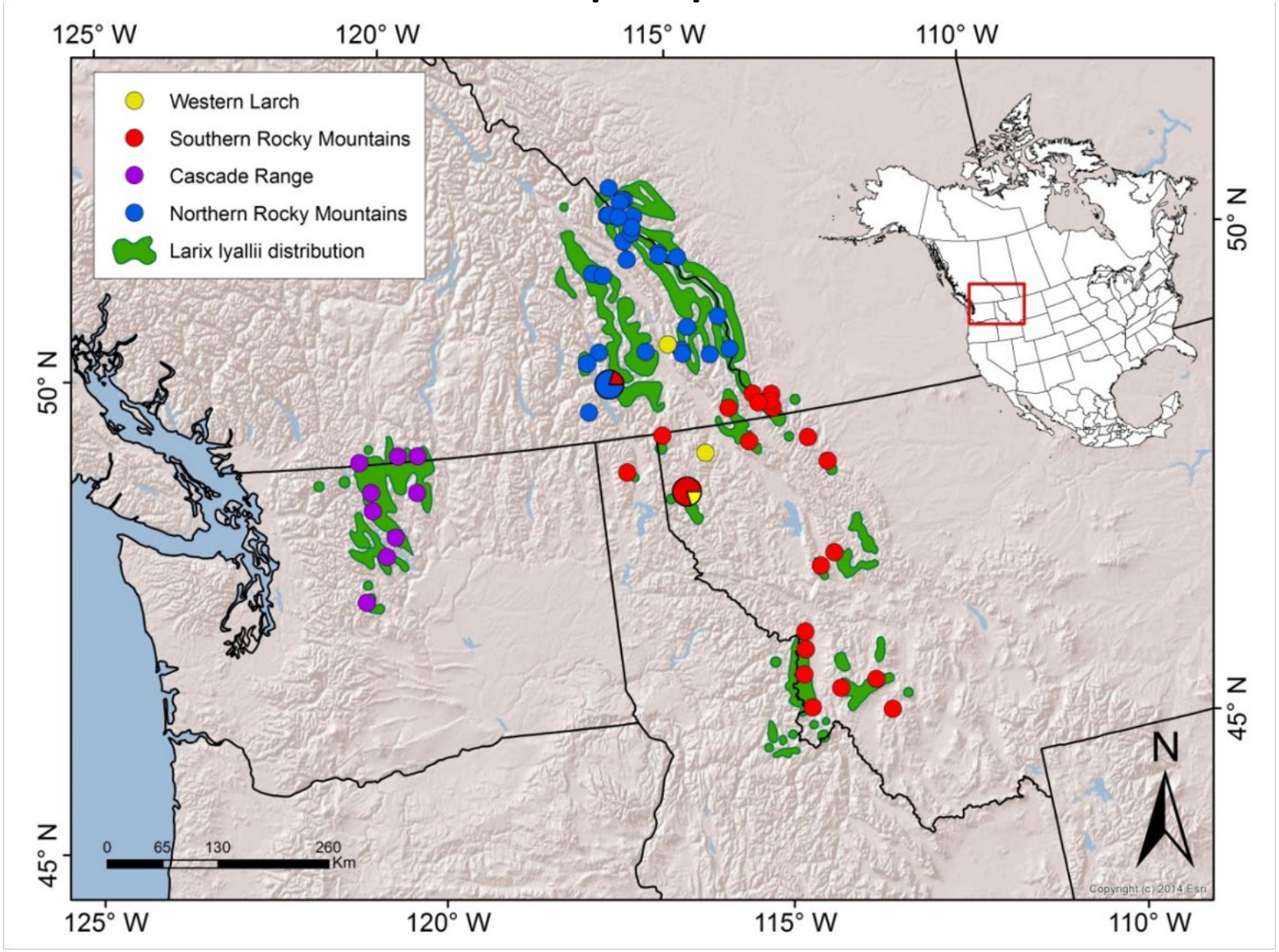


Small remnant populations



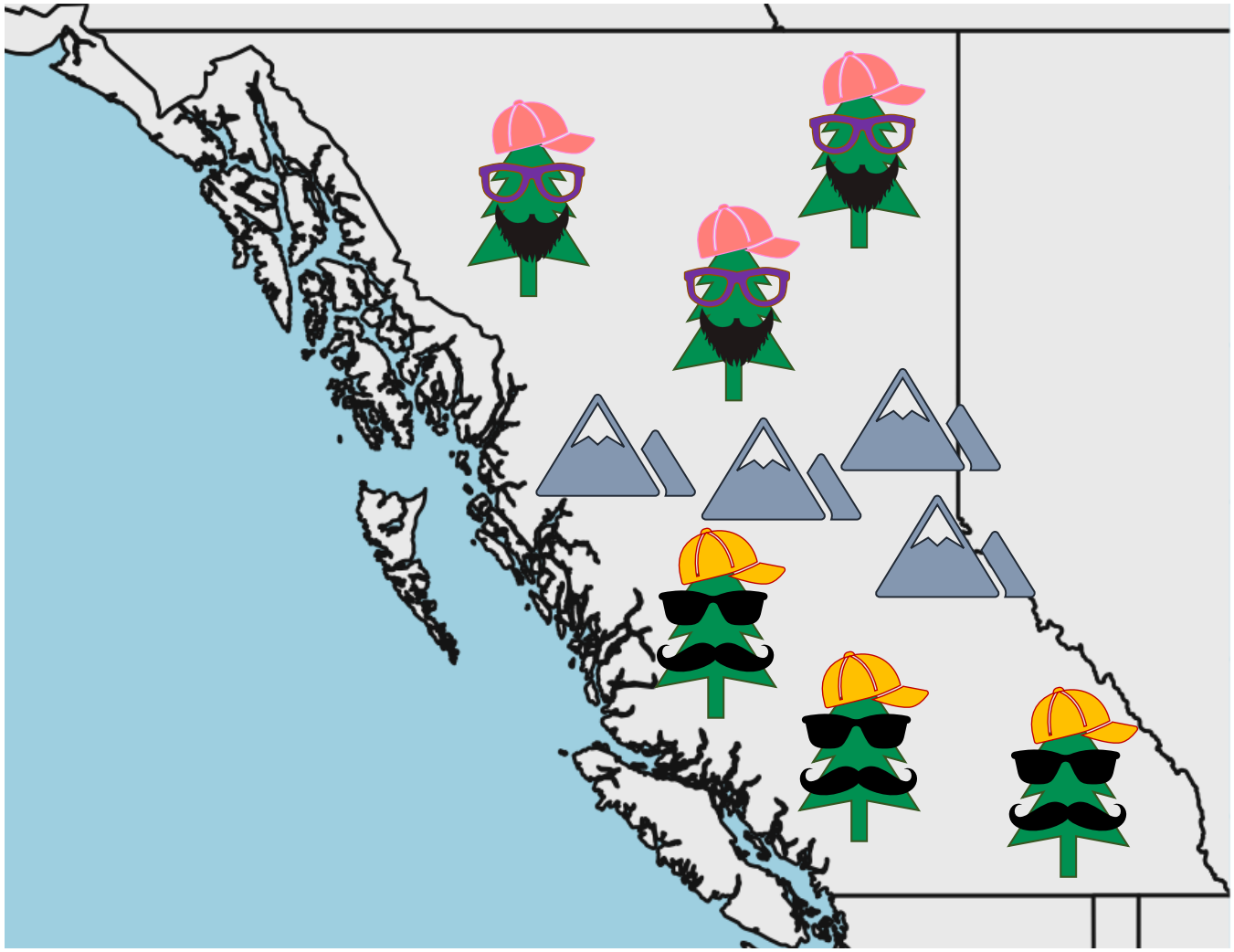
Small remnant populations



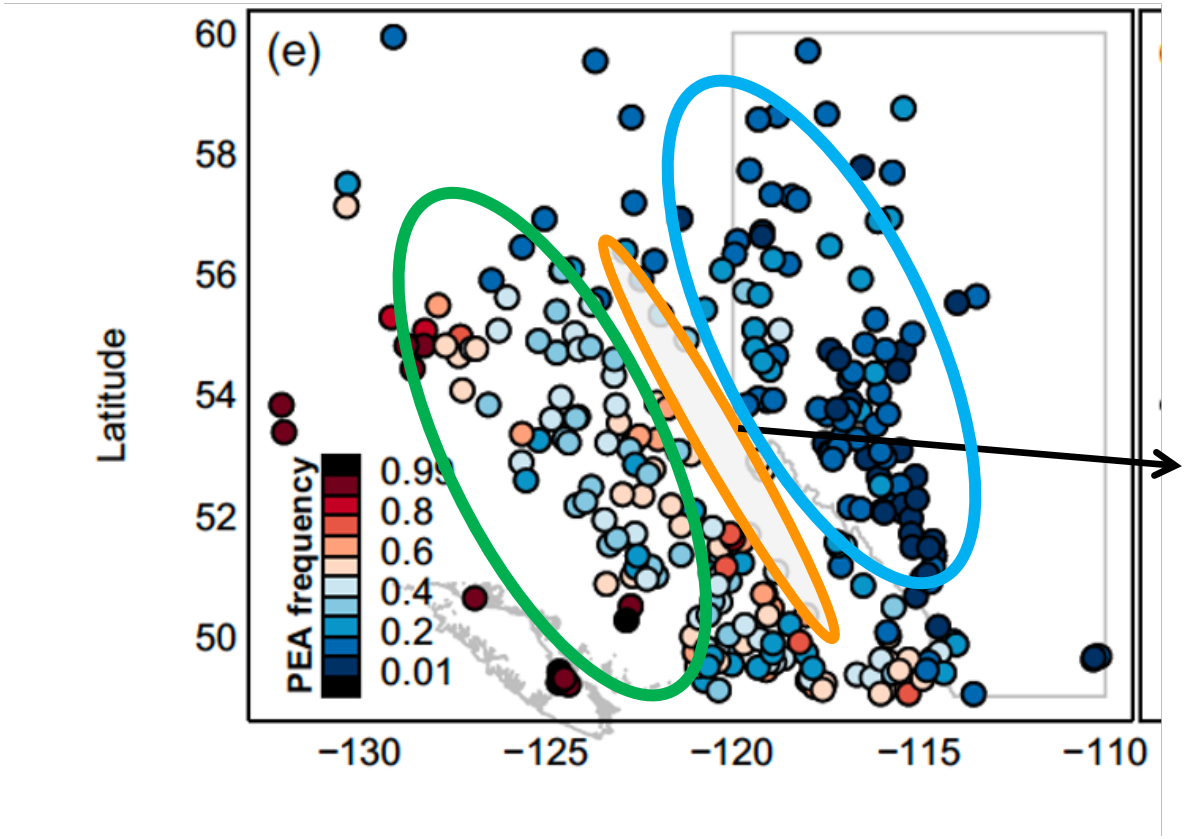


Vance 2019

Geographic Barriers

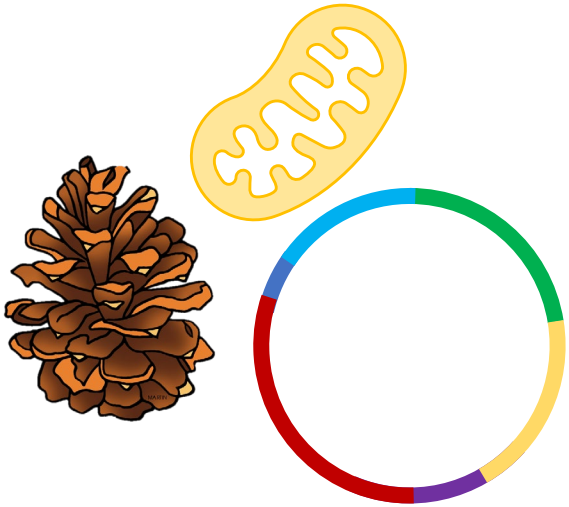


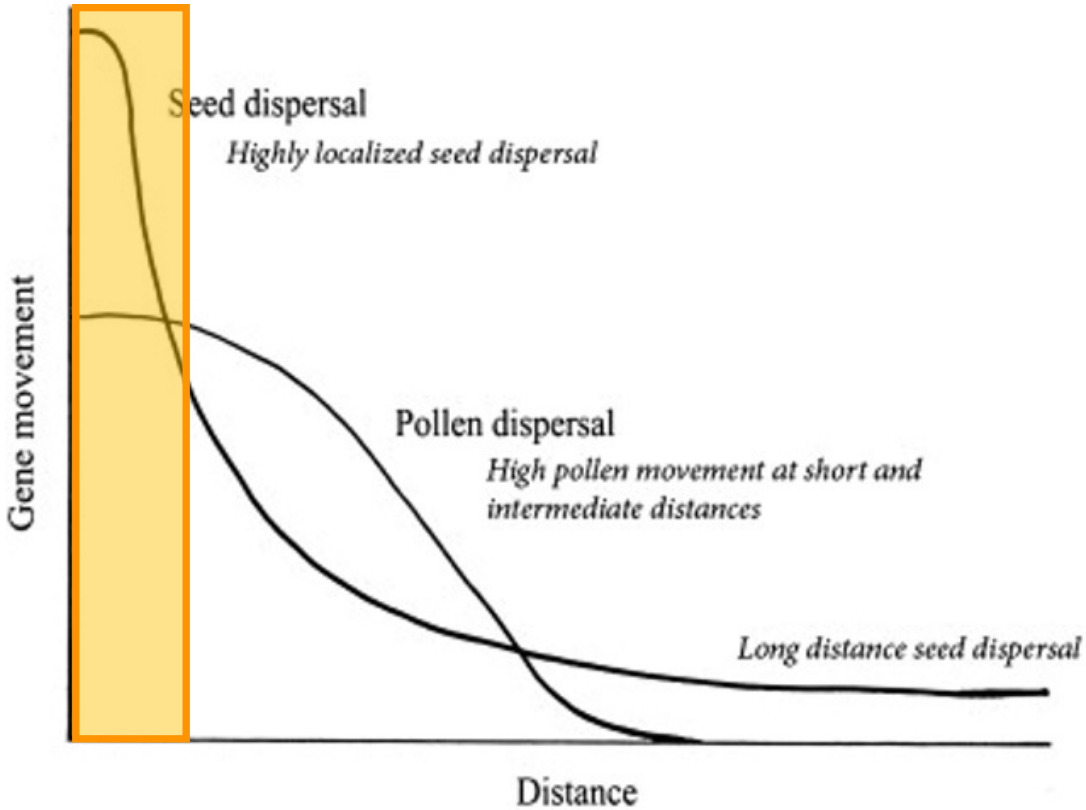
Geographic Barriers: Lodgepole Pine



Mahony et al 2019

Mitochondrial Markers

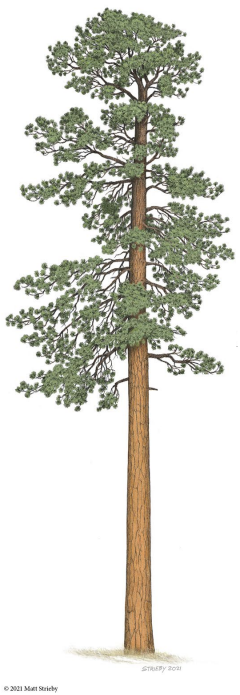


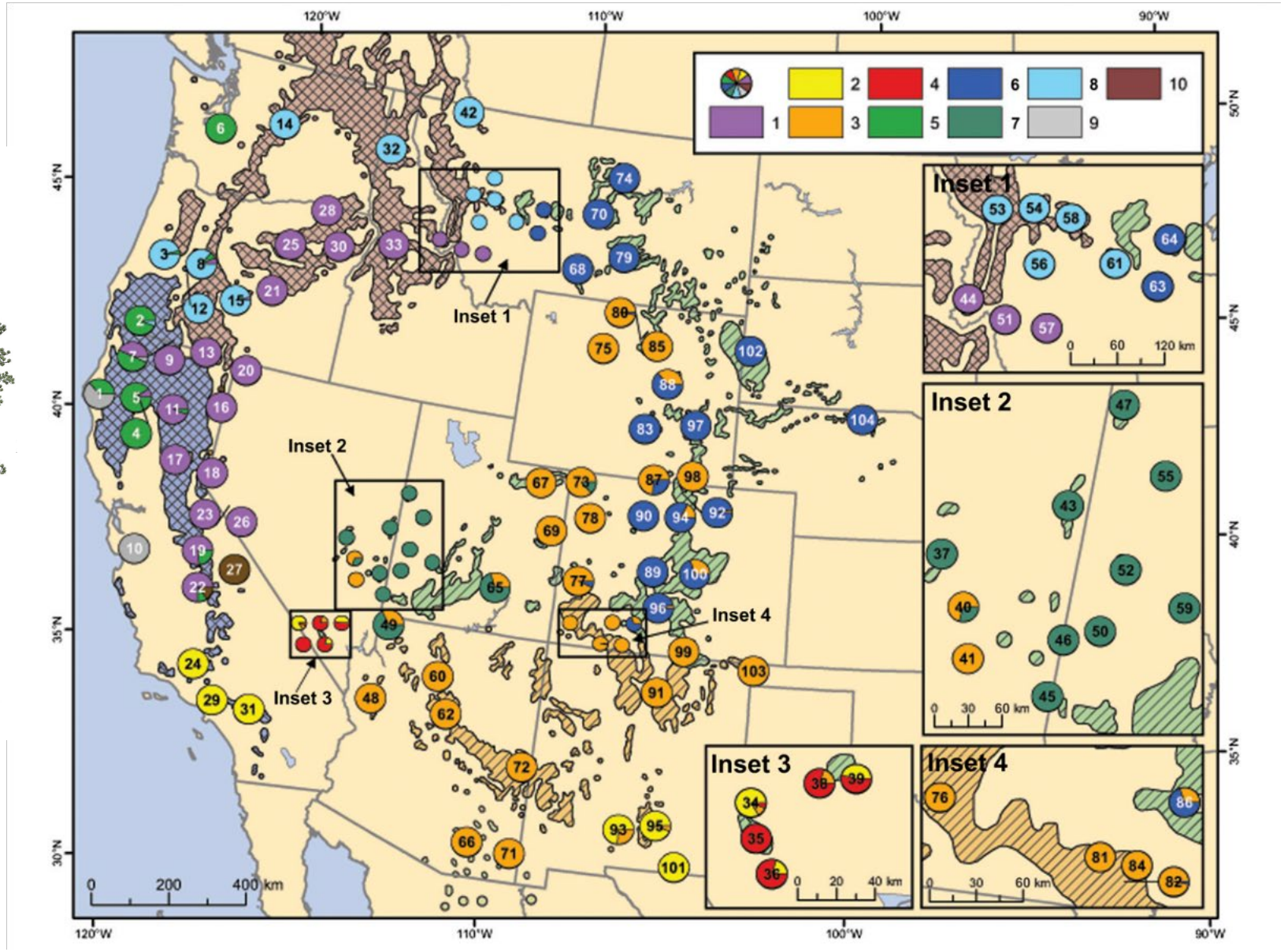


Seeds tend to stay closer to mom
*still depends on variation across moms

Hamrick & Trapnell 2011

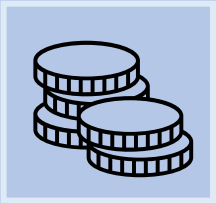
Mitochondrial Markers





Potter et al. 2013

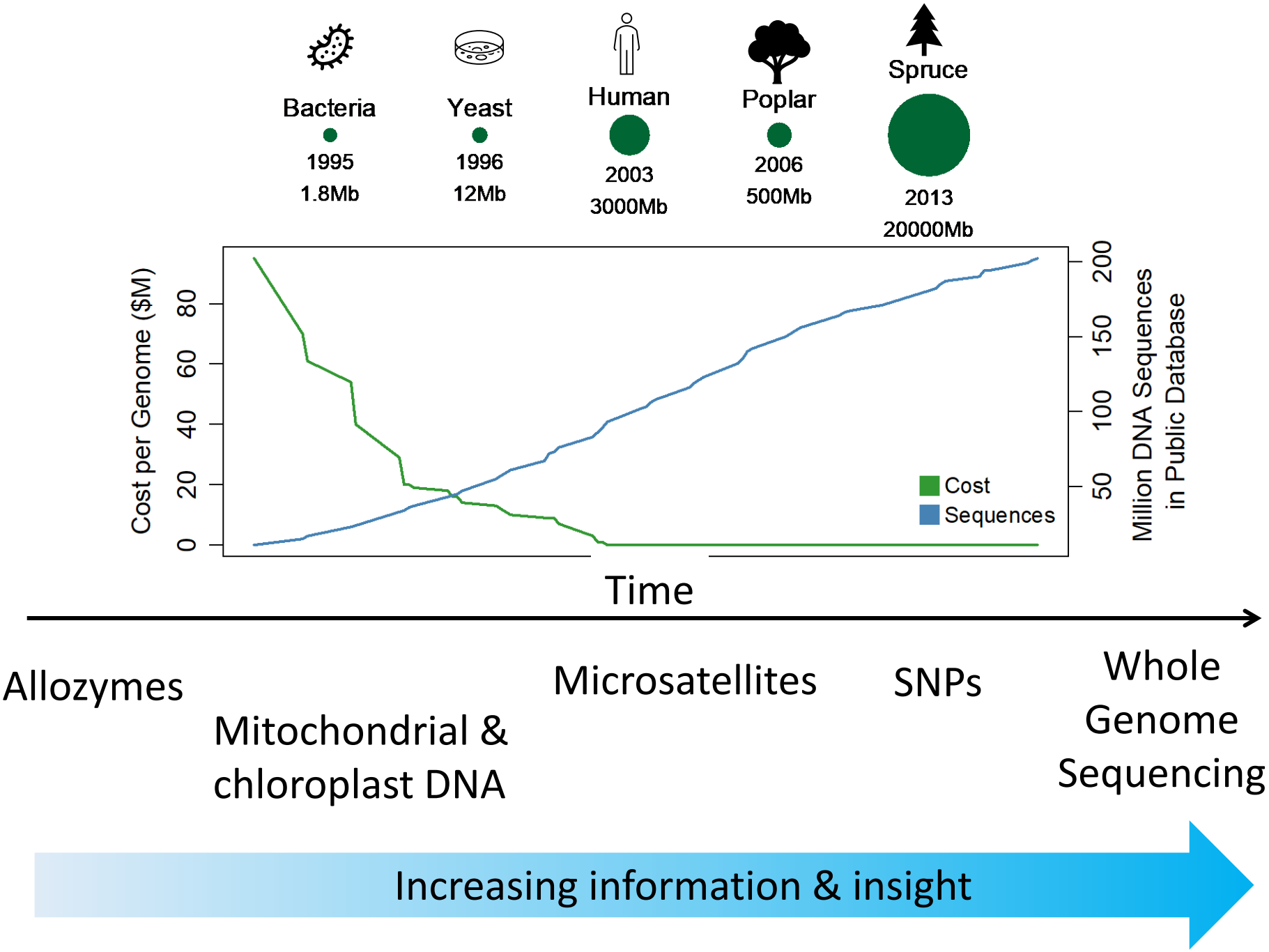
Conclusions



- Any provenance identification would take investment

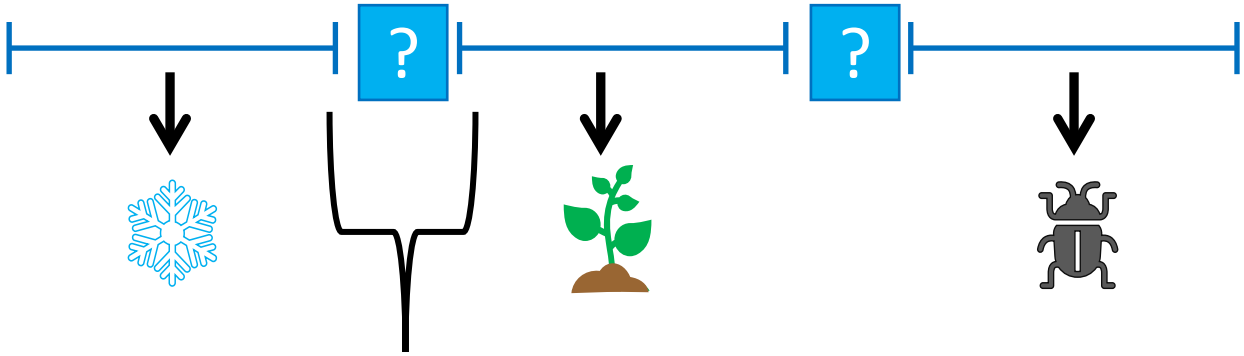
- Accuracy of ID highly dependent on scale and life history

BUT…



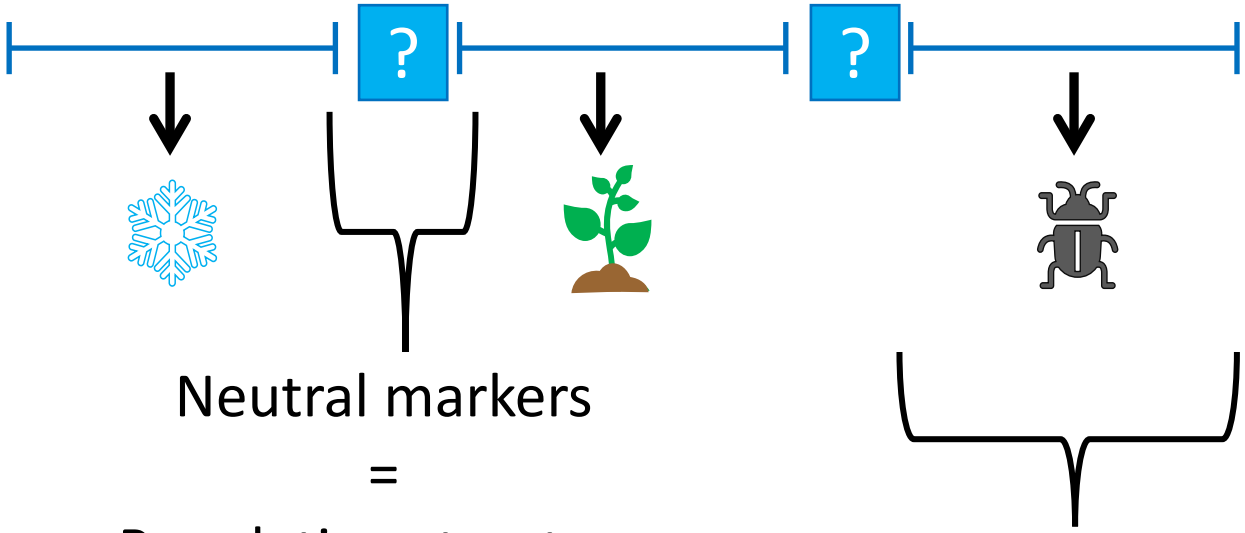
[1] National Human Genome Research Institute [2] National Center for Biotechnology Information

Using adaptive markers



Use neutral markers to examine demographic processes like migration, population size, and population structure

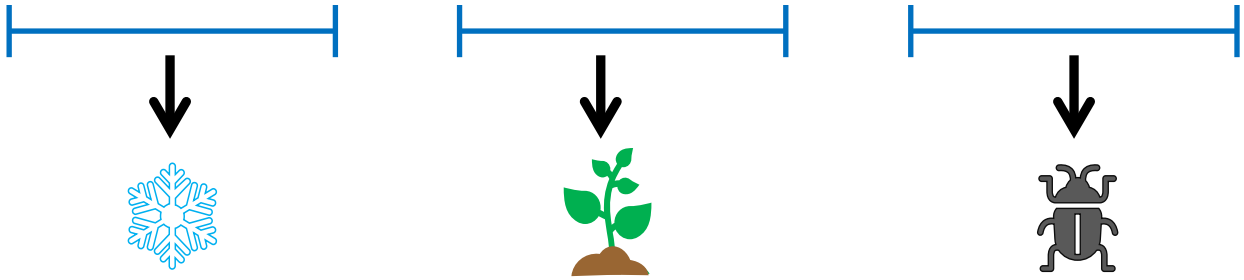
Using adaptive markers

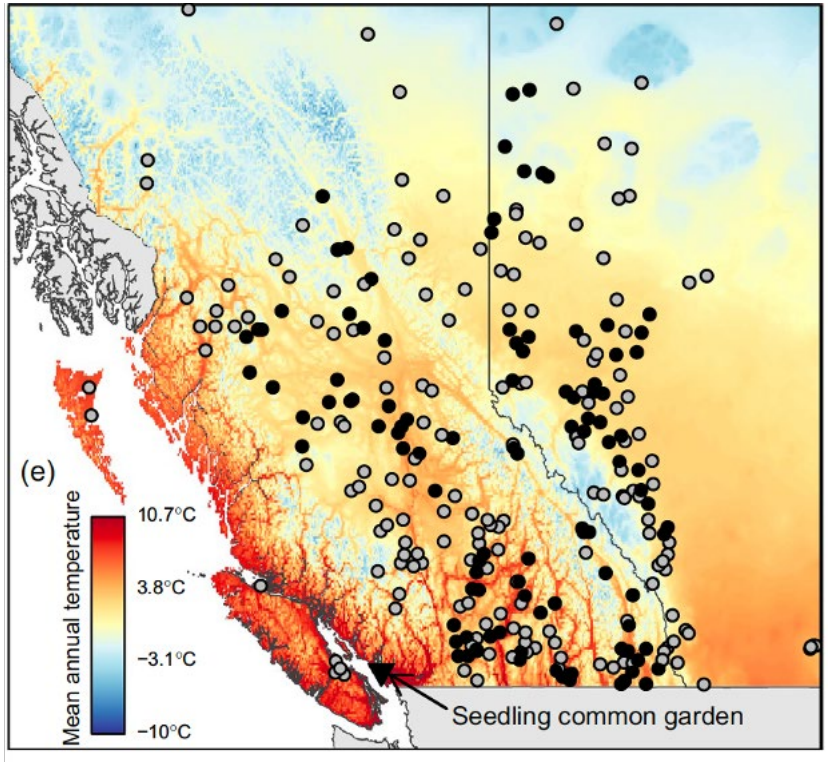


Neutral markers
=
Population structure

Adaptive markers could reveal finer scale geographic structure

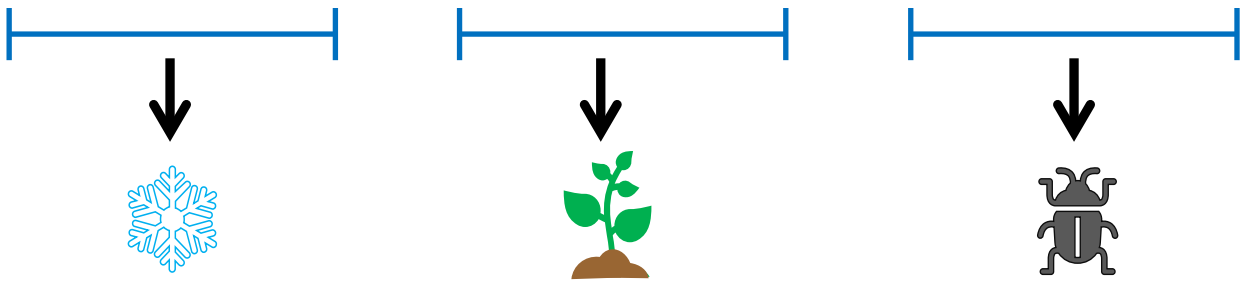
Using adaptive markers

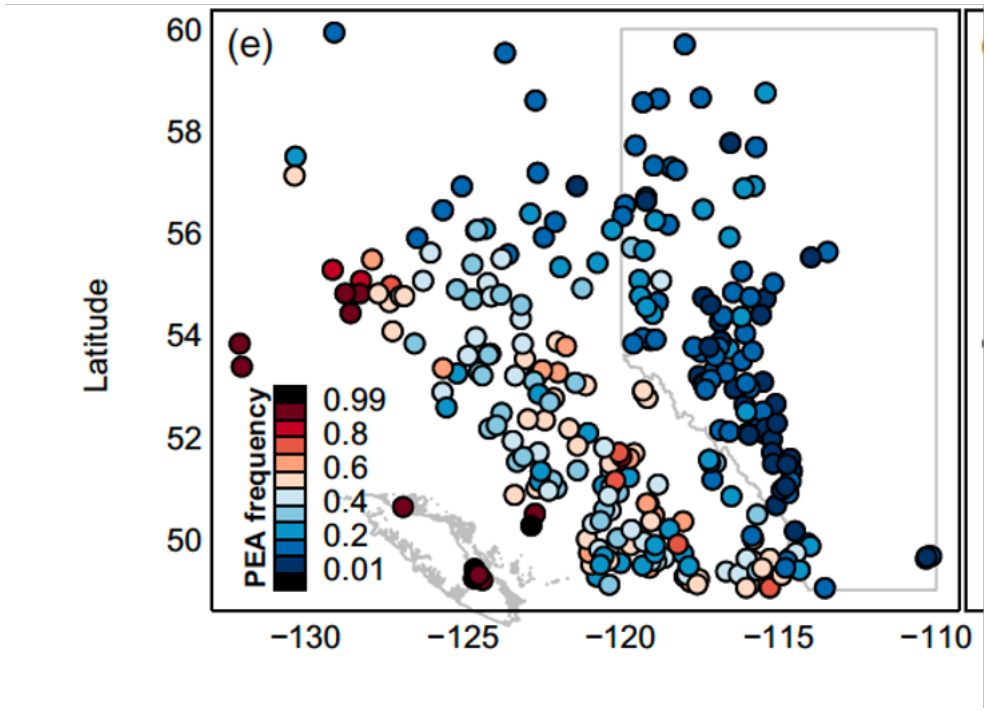




Mahony et al 2019

Using adaptive markers



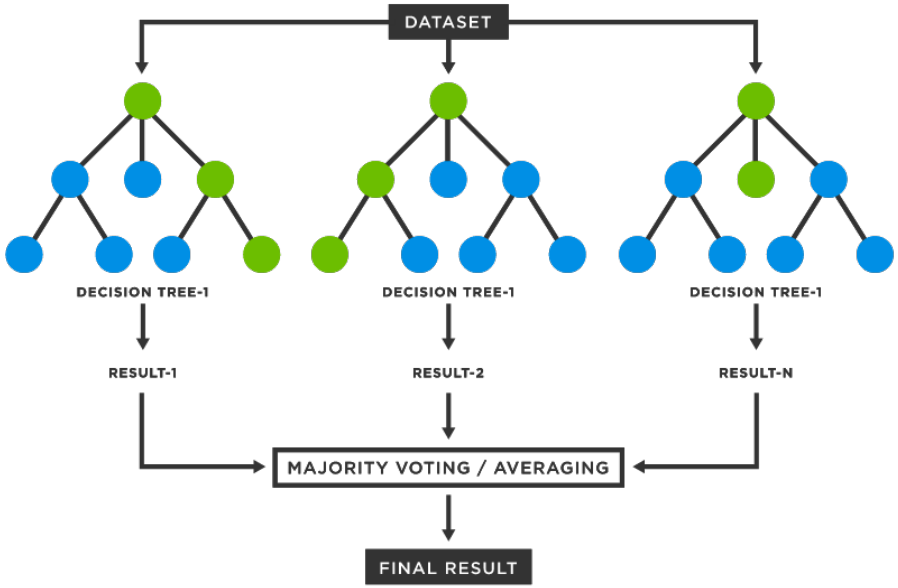


SNPs
are your friend

Mahony et al 2019

Advances in assignment methods

Random forest



Machine Learning

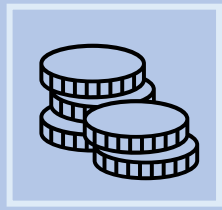
performed up to 7.8 and 11.2 percentage points better than F_{ST}-selected panels of similar size

- Sylvester et al. 2017

this approach yields median test errors of 16.9km, 5.7km, and 85km

- Battery et al. 2020

Conclusions Amended*



- Any provenance identification would take investment

- Accuracy of ID highly dependent on scale and life history

*** Investment will decrease and accuracy will increase ***

Questions?

or feel free to ask me later:

hayley.tumas@ubc.ca
@HayleyTumas