USDA Forest Service
Pacific Northwest Research Station

The ABCs of DNA ~~GMOs~~:
Demystifying Genomics Jargon

Rich Cronn

Research Geneticist

PNW Research Station Corvallis, OR

richard.cronn@usda.gov

Tree Improvement 101 Webinar Series

https://forestrywebinars.net/sponsor-pages/trees-101•
October 25, 2024

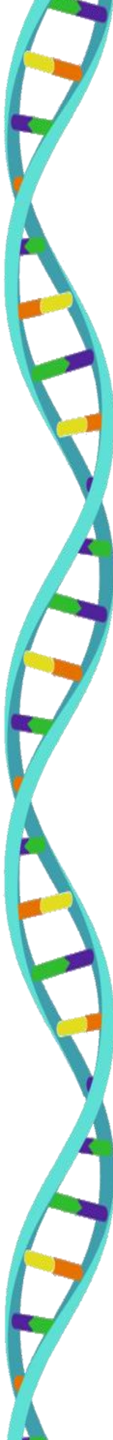
ABC’s of DNA

Outline

1. Genetics basics of DNA and genetic markers; examples showing how data are produced.
2. Genomics basics and genome data; examples showing how data are produced.
3. Examples how genetic and genomic approaches are used in orchard and forest management.

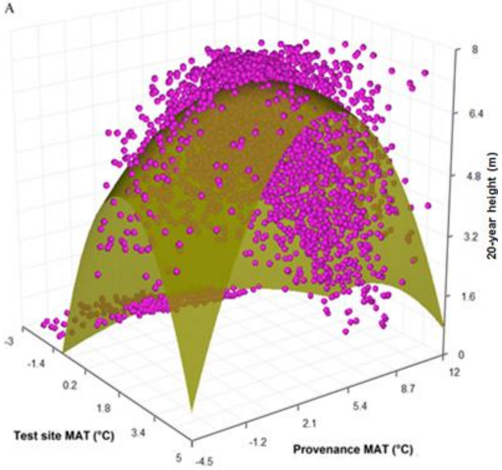
Goals

1. Provide an overview of jargon used in genetics.
2. Show relevant examples of modern applications.



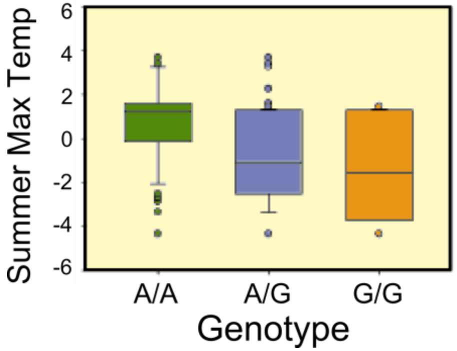
Two views of genetics

Quantitative



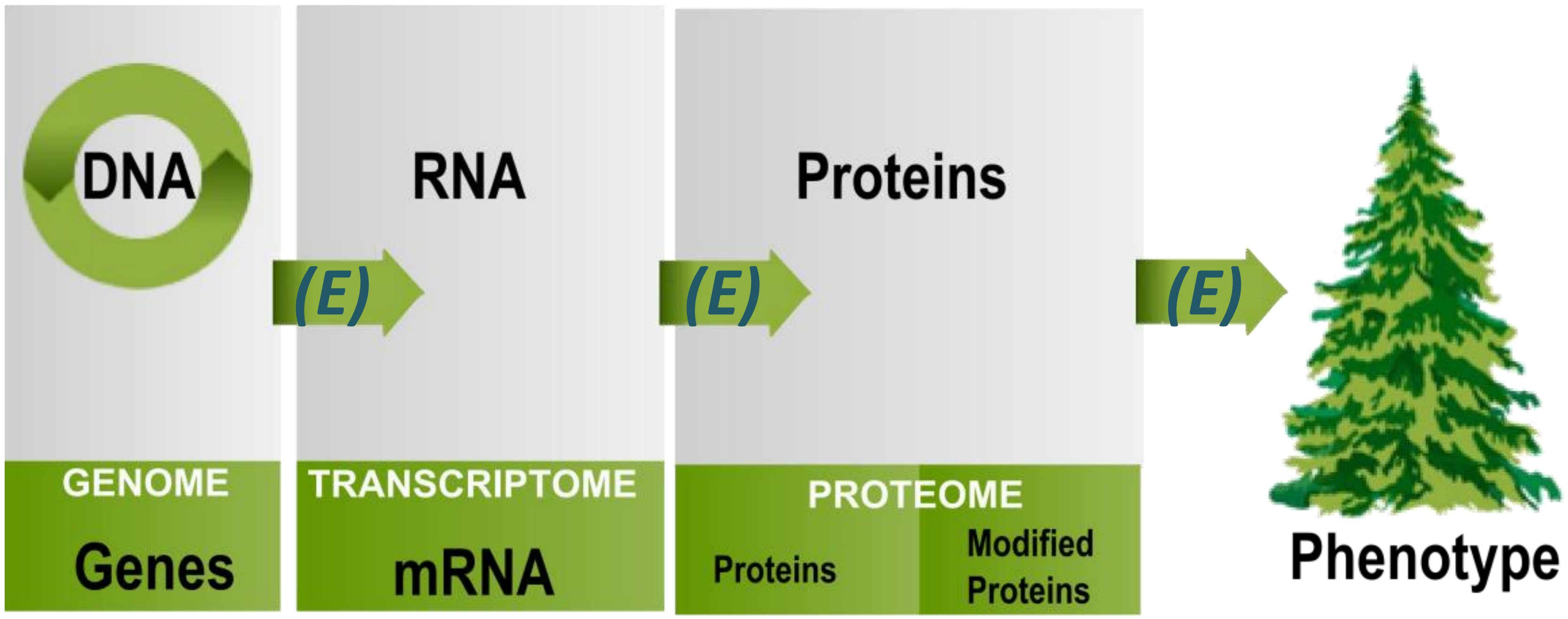
Traits controlled by large numbers of genes, small effects

Molecular



Traits controlled by small numbers of genes, large effects

Genetics Basics – G, E, P



1. "Central Dogma":
DNA encodes RNA »
RNA encodes proteins »
Proteins encode
biosynthesis, metabolism

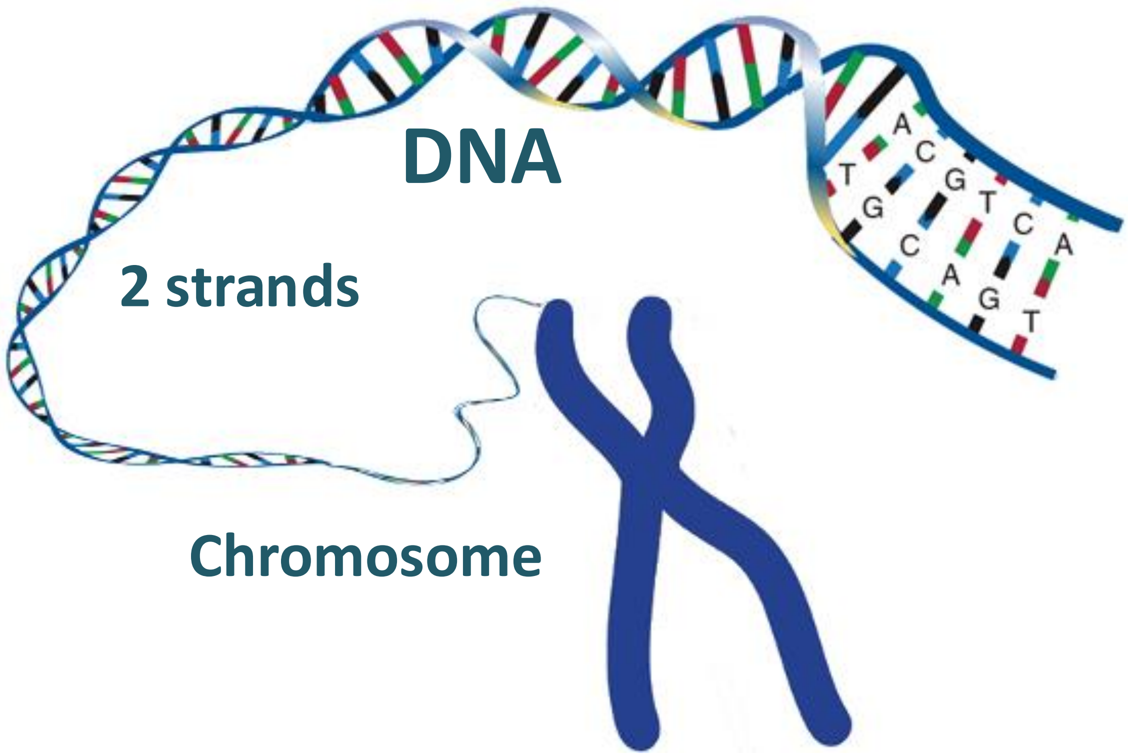
2. Molecules are inherited, but the phenotype is influenced by genetics *and* environment.

P = $f(G,E)$

Genetics Basics - DNA

1. DNA: double-stranded molecule made of 4 bases, organized into chromosomes, basis of inheritance

2. DNA is present in all cells and tissues *(leaves, bark, roots, seeds, wood)*



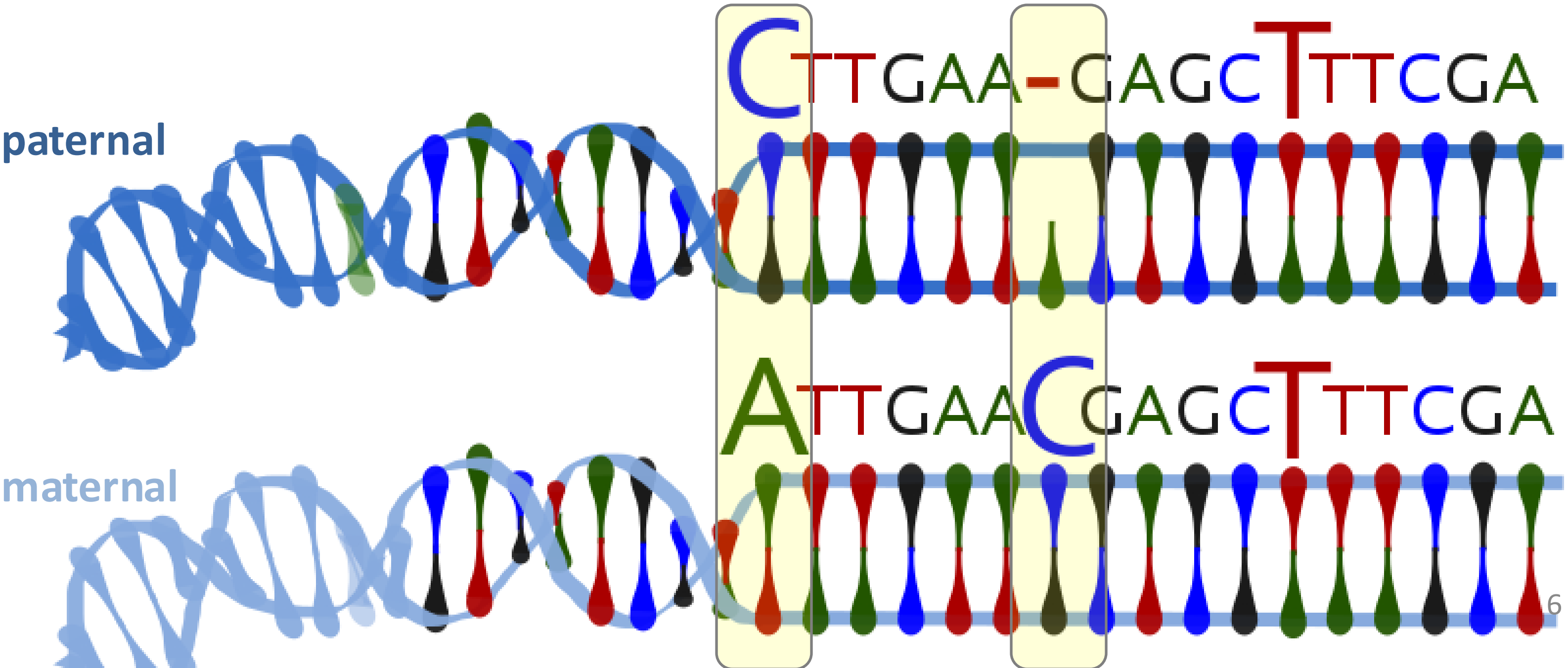
DNA

2 strands

Chromosome

Genetics Terms – DNA

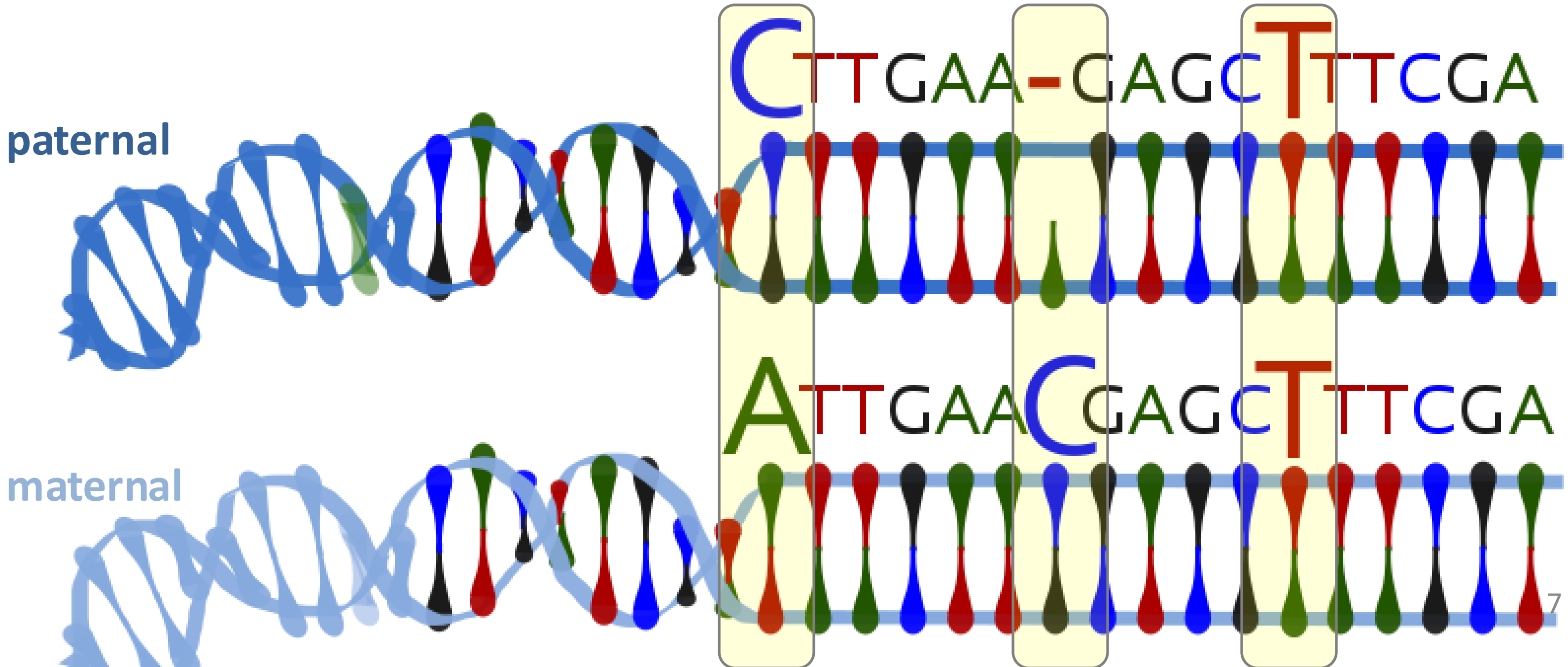
- ***Single nucleotide polymorphisms* or *SNPs*: variation at one ,** **base position** (e.g., C/A)
- ***Insertion/deletion polymorphism*: variation where one or more bases are added/missing** (e.g., -/C)



Genetics Terms - DNA

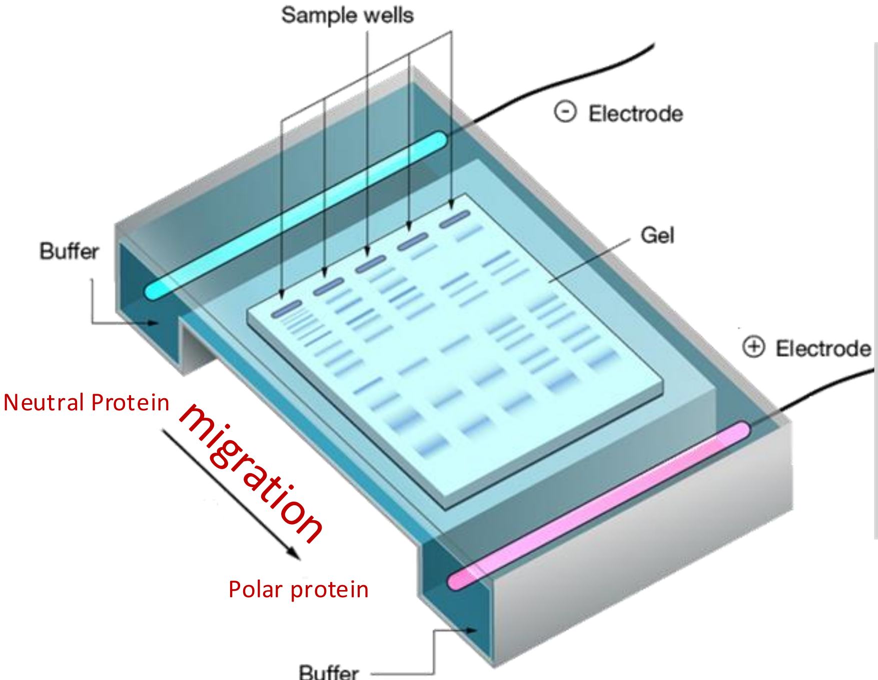
Variation in an <u>individual</u> can be defined two ways:

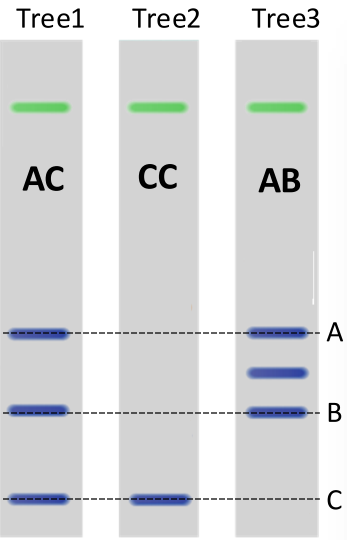
- ***Genotype:* Pairs of bases at defined marker location**
  (e.g., C/A, -/C, or T/T)
- ***DNA Profile*: Combined genotypes across <u>all</u> marker locations**
  (e.g., C/A·-/C·T/T)

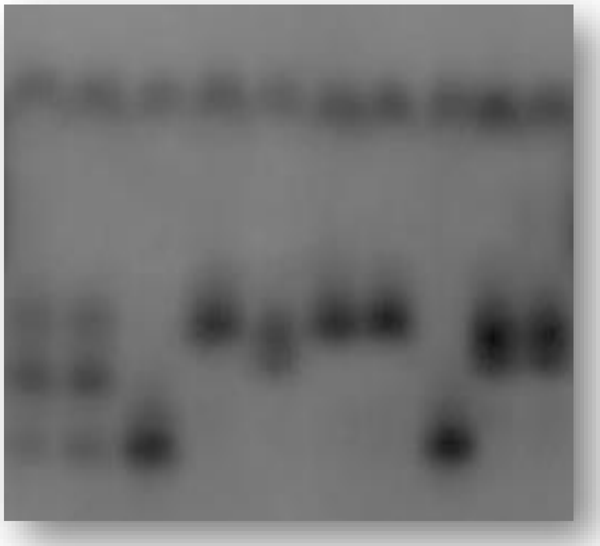


Revealing genetic variation: Proteins

- ○ ***Allozymes*** **(*isozymes*), 1957 to 2000(+/-).**
- ○ **Variation in amino acid sequences (*alleles*); variants migrate differently under a charged field.**
- ○ **Visualized by *gel electrophoresis and* enzymatic reactions.**

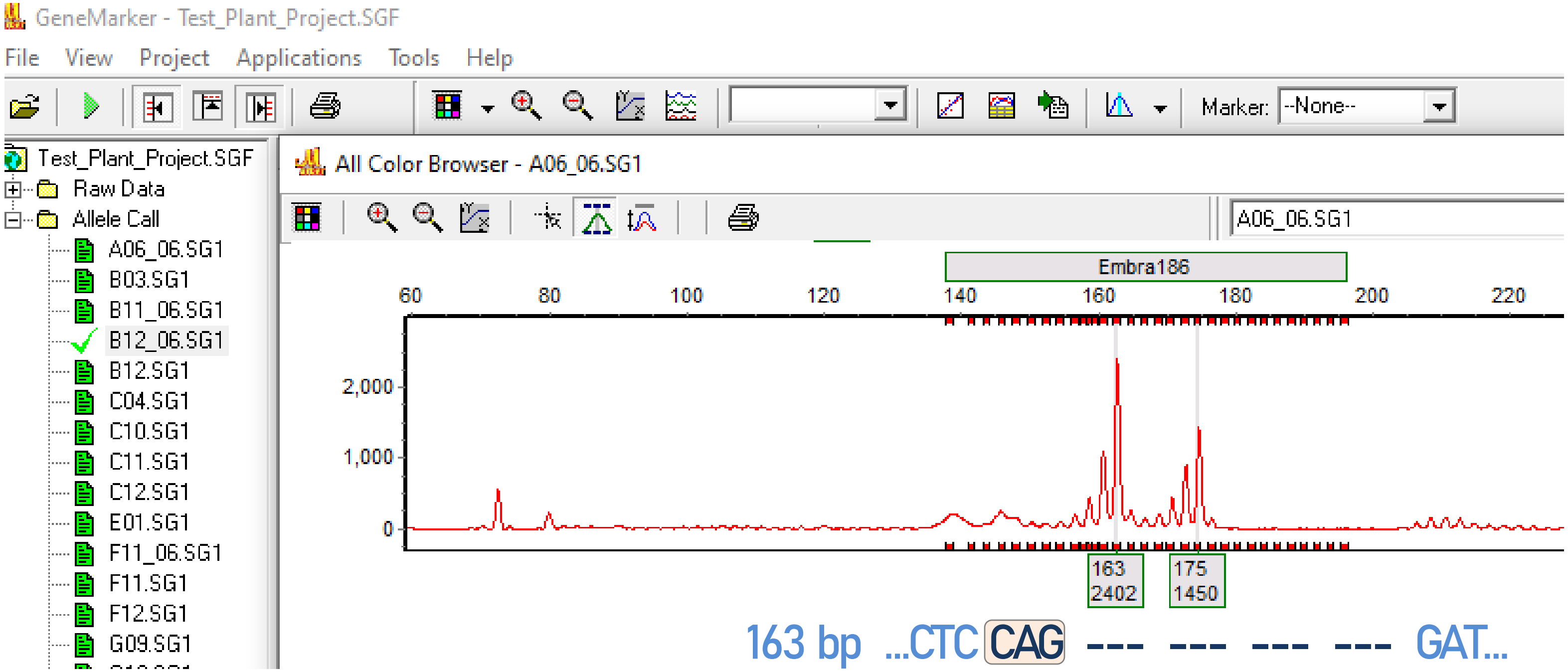






Revealing genetic variation: Microsatellites

- **DNA fragment methods – *microsatellites*, 1989 to 2024(+/-).**
- **Short, tandemly-repeated DNA sequences; highly-variable.**
- **Examples: AC, GT (2bp), GAT, CAG (3 bp), GGAT (4 bp), etc.**

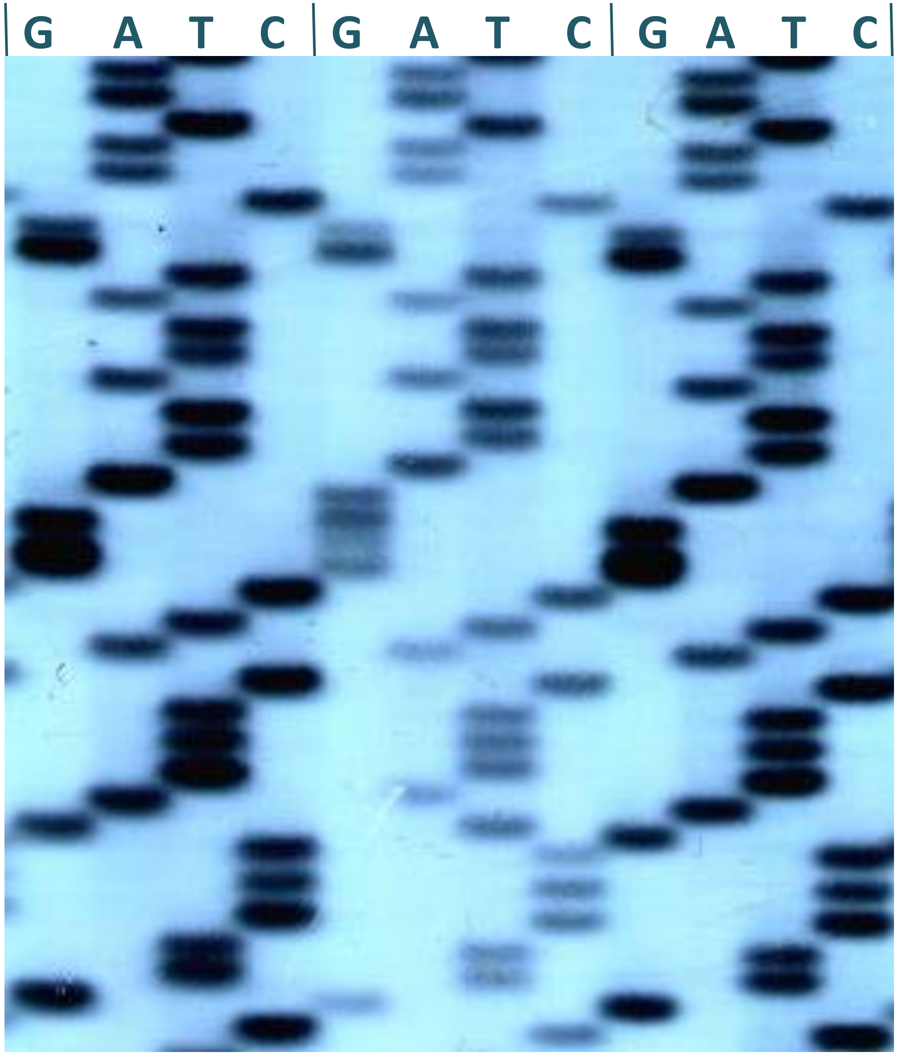


163 bp ...CTC CAG --- --- --- --- GAT...

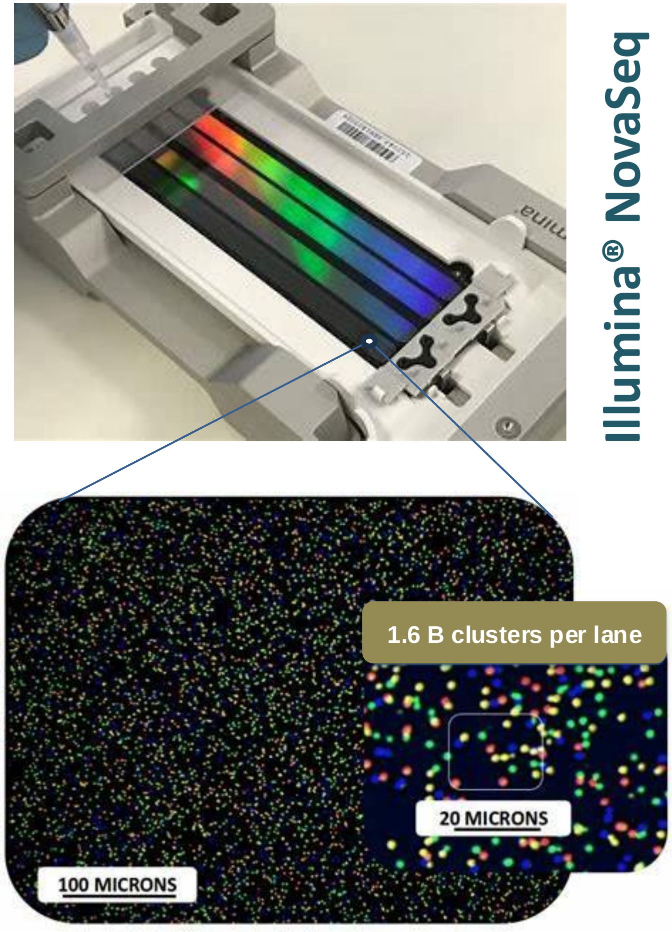
175 bp ...CTC CAG CAG CAG CAG CAG GAT...

Revealing genetic variation: Sequences

DNA sequencing, Sanger (1977^{+})

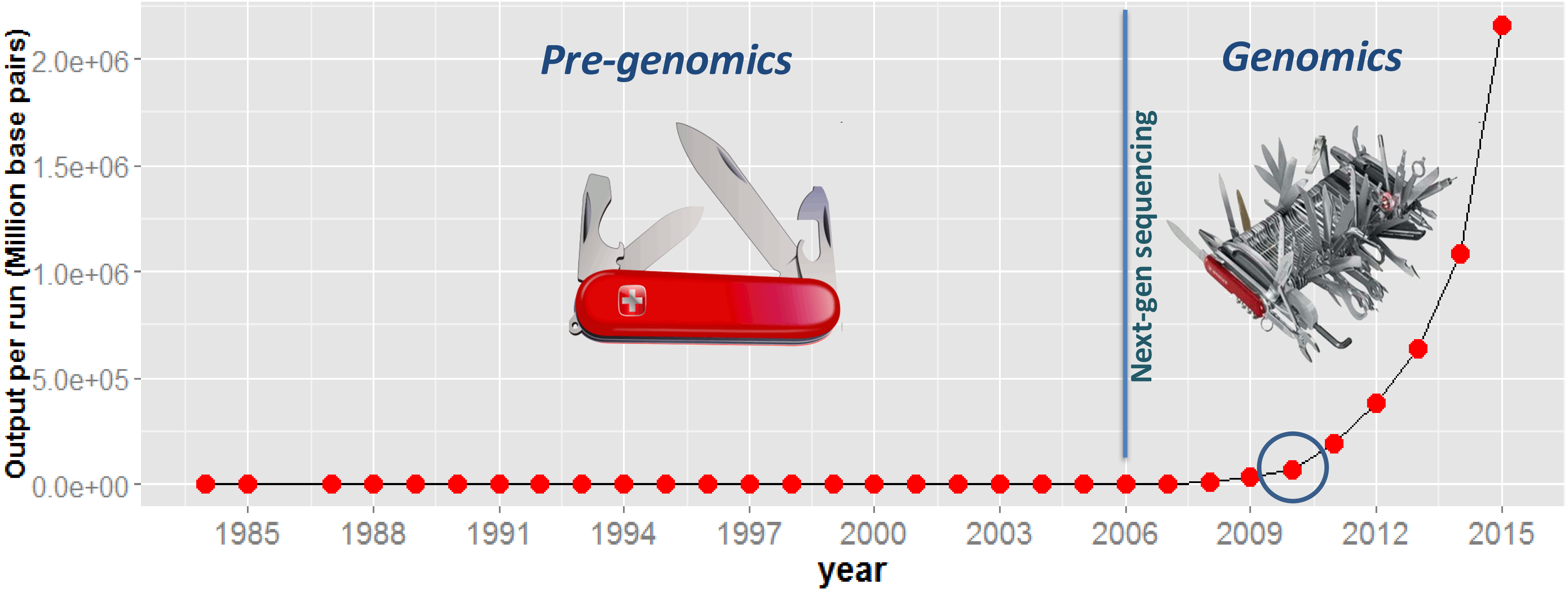


‘Next-generation’ sequencing (2006^{+})



Illumina® NovaSeq

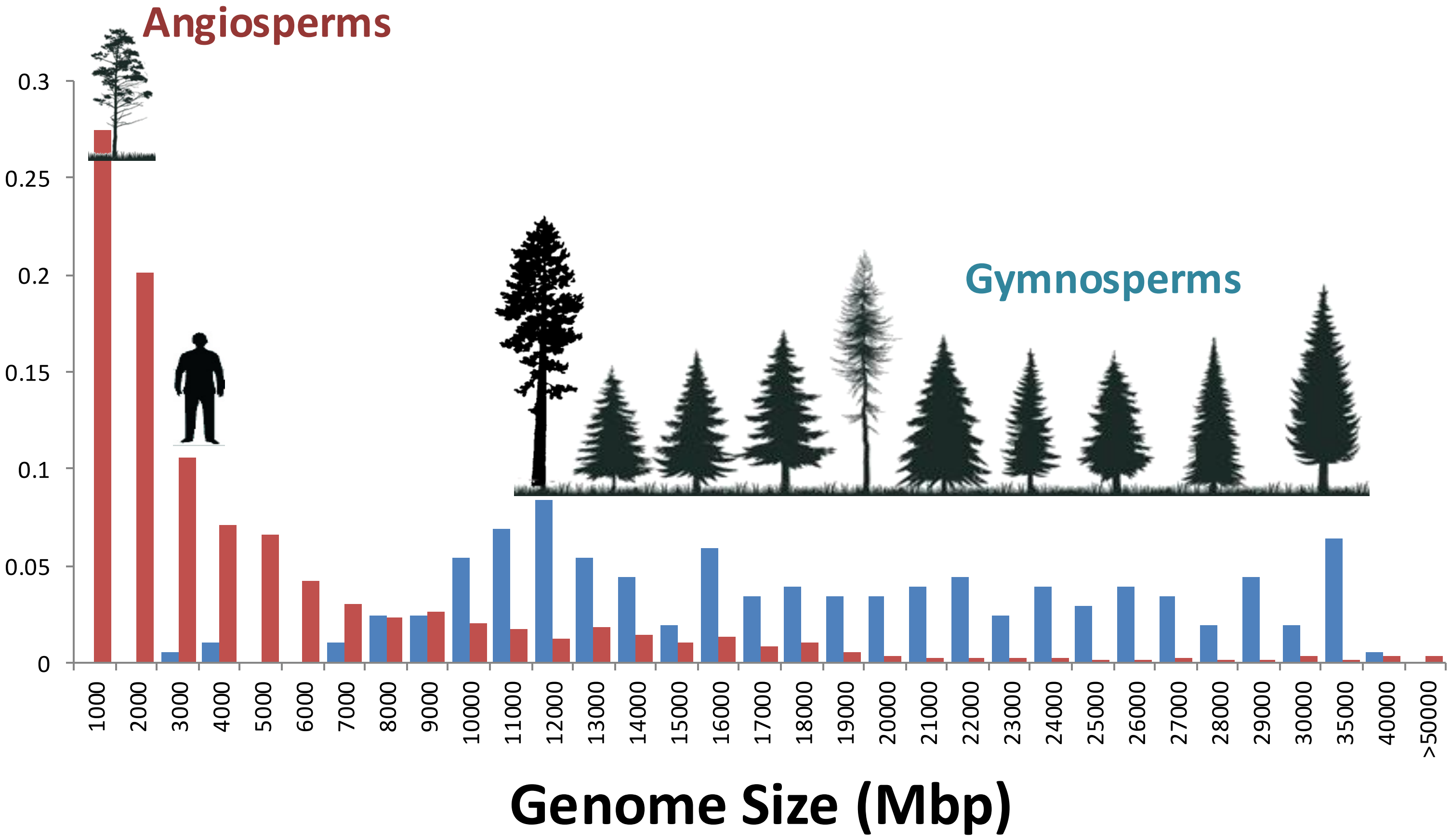
Welcome to the genomics era



Feature	Pre-genomics	Genomics
Variable genetic markers per sample	<< 100	1000s – millions
Cost per data point (*sample*)	high (*Low*)	low (*High*)
Statistical power	***Low***	***High***

Genome size variation in trees

- ○ **Angiosperm average = 0.9 billion bases (400 – 4,000 Mbp)**
- ○ **Gymnosperm average = 20 billion bases (11,000 – 38,000 Mbp)**



Assembling genomes

encing

Dra

○ l

○ l ds

Example questions in molecular genetics

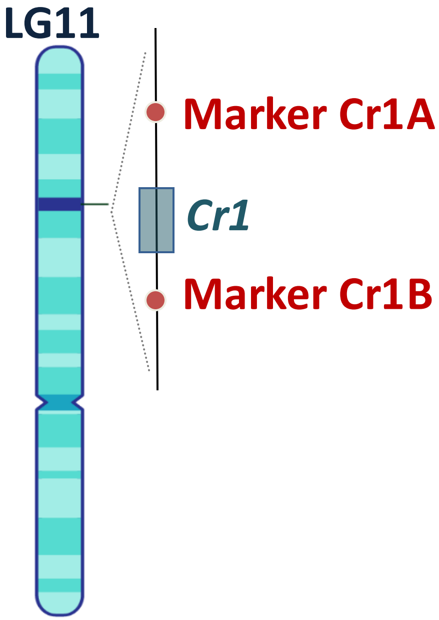
• *Can DNA identify important traits?*	**Association**
• *Is this two trees or one?*	**Individualization**
• *Are trees in this orchard related?*	**Parentage/pedigree**
• *Where did this seed or tree originate?*	**Geolocalization**
• *Can DNA discriminate similar species?*	**Taxonomic ID**
• *Can DNA be used to make seed zones?*	**Genome -- Environment Association**

1. Can DNA identify important traits?

- Seed orchard development for Sugar Pine
- Blister Rust (*Cronartium ribicola*) infects SP, girdles stems
- Resistance gene *Cr1* identified in 1970; rare (0% - 8% of trees)
- Resistance screening established; 2 yr cycle, >$1000 per family

1. Trait Identification

- 2017: Sugar Pine genome sequenced.
- 2 markers strongly associated with *Cr1R* allele*.
  - *Related* trees genotyped at Cr1B match *C.r.* testing 100% of time.
  - *Unrelated* trees genotyped at Cr1B match *C.r.* testing 85% of time.
- Cr1B test can *significantly* streamline parent tree screening

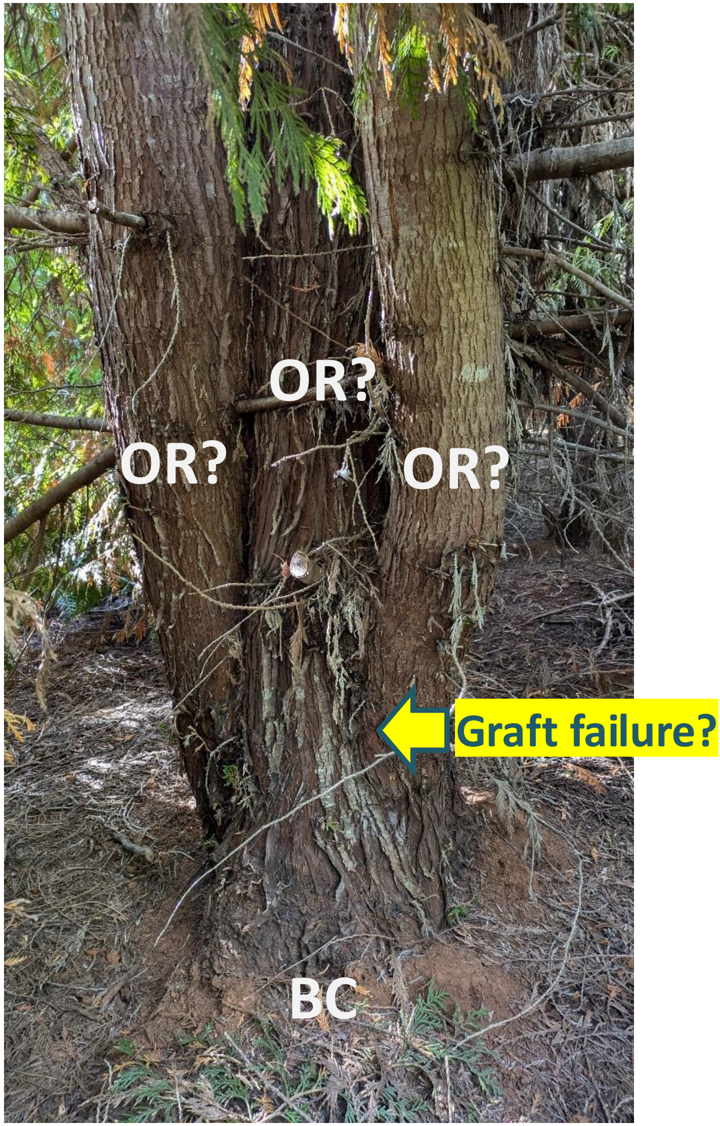


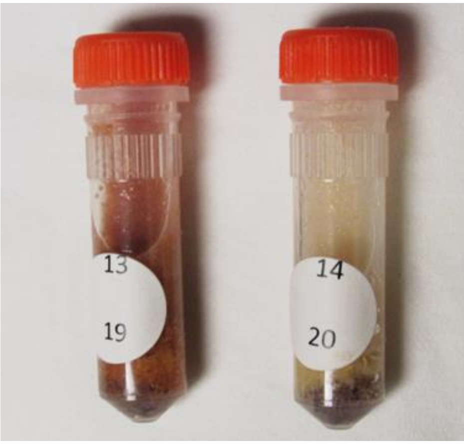
2. Is this two trees or one?

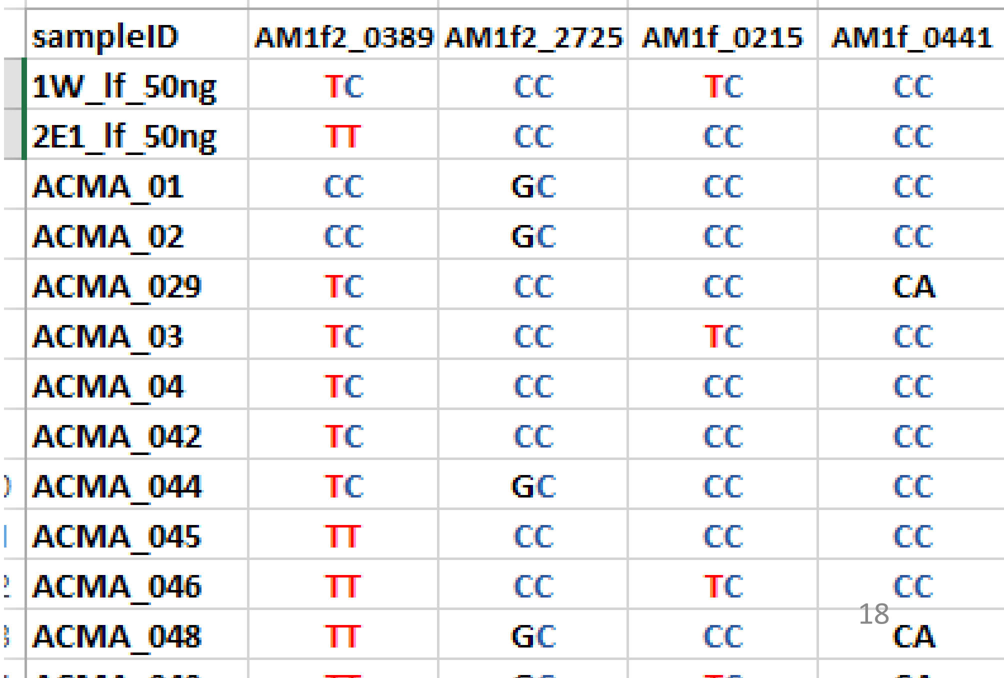
Grafting widely used to preserve, increase desirable selections.

***Thuja plicata* clones, Coastal OR. OR scion on BC rootstock**

2. Individualization







sampleID	AM1f2_0389	AM1f2_2725	AM1f_0215	AM1f_0441
1W_lf_50ng	TC	CC	TC	CC
2E1_lf_50ng	TT	CC	CC	CC
ACMA_01	CC	GC	CC	CC
ACMA_02	CC	GC	CC	CC
ACMA_029	TC	CC	CC	CA
ACMA_03	TC	CC	TC	CC
ACMA_04	TC	CC	CC	CC
ACMA_042	TC	CC	CC	CC
ACMA_044	TC	GC	CC	CC
ACMA_045	TT	CC	CC	CC
ACMA_046	TT	CC	TC	CC
ACMA_048	TT	GC	CC	CA

2. Individualization

- **2018 ‘Maple Fire’, Olympic NF** *Can DNA identify a match?*

2. Individualization

Bigleaf maple analysis

- ○ **Assay: 120 SNPs**
- ○ **Database: 235 trees**
- ○ **Crime evidence: 3 source stumps & 225 blocks**

Results

- ○ **83/225 evidence blocks matched source stumps**
- ○ ***Profile probability* less than 1 in one trillion trillion trillion** (= 5.32×10^{-39})

3. How are trees in an orchard related?

Black Walnut Breeding, 2024*

Breeding for timber yield, wood quality

- 259 open pollinated families *(half-sibs)*
- 56,909 progeny tested at 36 field trials

Open pollination » *half-sibs*

- Half-sibs share 25% of genes (mother).
- Father unknown.
- Model: "half-sibs" (self-pollination, pollen drift, tag errors all introduce error).

» *Errors inflate predicted genetic gain*

*** Breeding without breeding in walnut**

Vikram Chattre, Keith Woeste: Hardwood Tree Improvement and Regeneration Center (USFS – Purdue University)

3. Breeding and pedigree analysis

- **Genetic genealogy:** use DNA to identify relatives
- **Paternity Exclusion**: *exclude pollen donors that share no alleles with tested offspring*

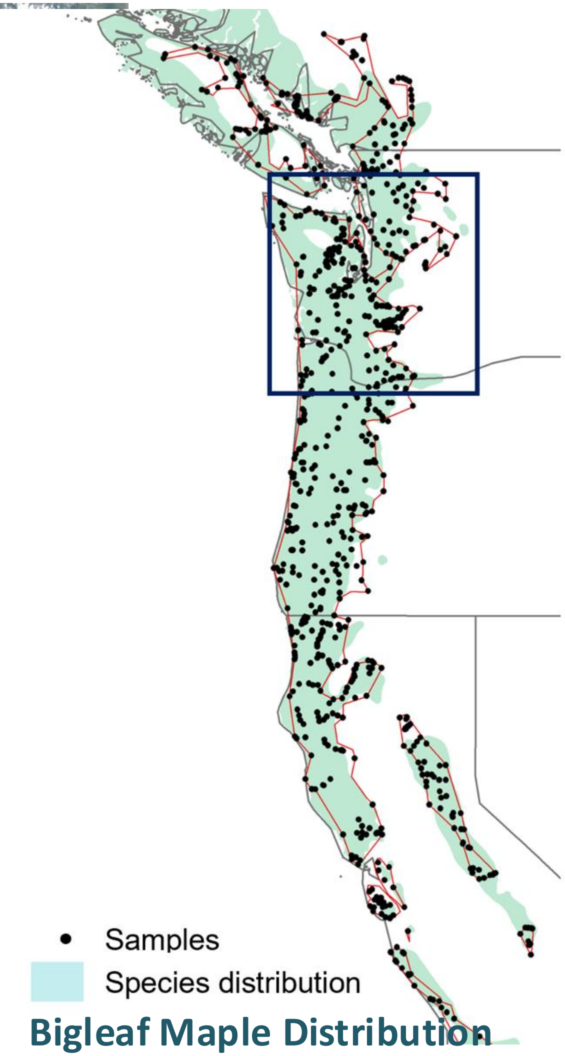
Tree	AM1f2 2725	AM2f 0290
Off.1.A	GG	AA
♀ Seed.1	GG	AT
♂ ~~Pollen.1~~	GT	TT
♂ Pollen.2	GG	AT
♂ ~~Pollen.3~~	GT	TT

*** Breeding without breeding in walnut**

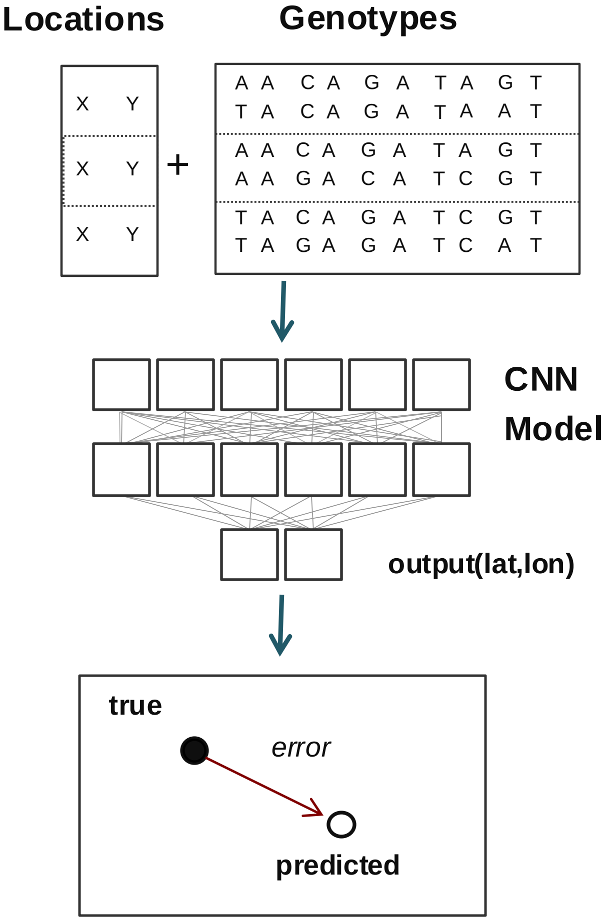
Vikram Chattre, Keith Woeste: Hardwood Tree Improvement and Regeneration Center (USFS – Purdue University)

4. Where did this tree originate?

Mt Hebo, OR



Bigleaf Maple Distribution



Locations **Genotypes**

X Y	A A C A G A T A G T T A C A G A T A A T
X Y	A A C A G A T A G T A A G A C A T C G T
X Y	T A C A G A T C G T T A G A G A T C A T

+

CNN Model

output(lat,lon)

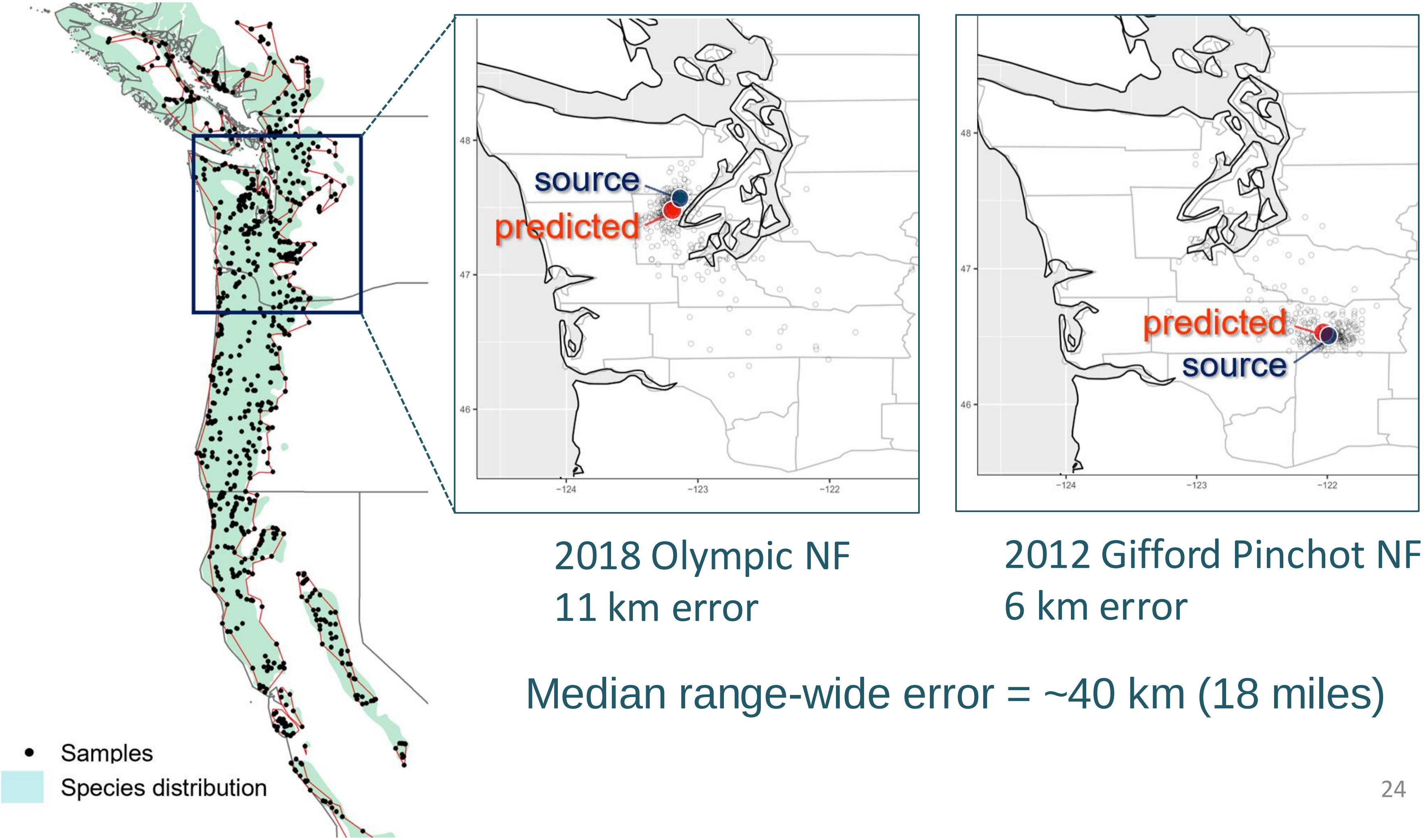
true *error* **predicted**

***'LOCATOR'* -- Battey, Kern & Ralph**

doi: 10.7554/eLife.54507

4. Geolocalization

1,500 Bigleaf maple samples; 120 SNPs



2018 Olympic NF
11 km error

2012 Gifford Pinchot NF
6 km error

Median range-wide error = ~40 km (18 miles)

5. Can DNA discriminate similar species?

Whitebark

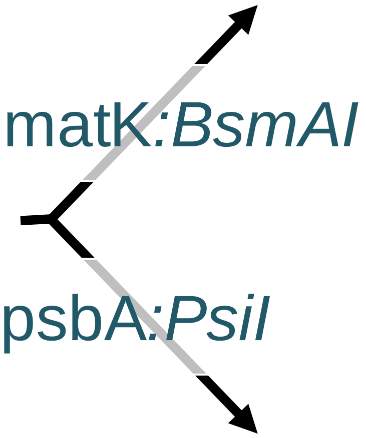
threatened, 2022

Limber

?

5. Genomics and taxon ID

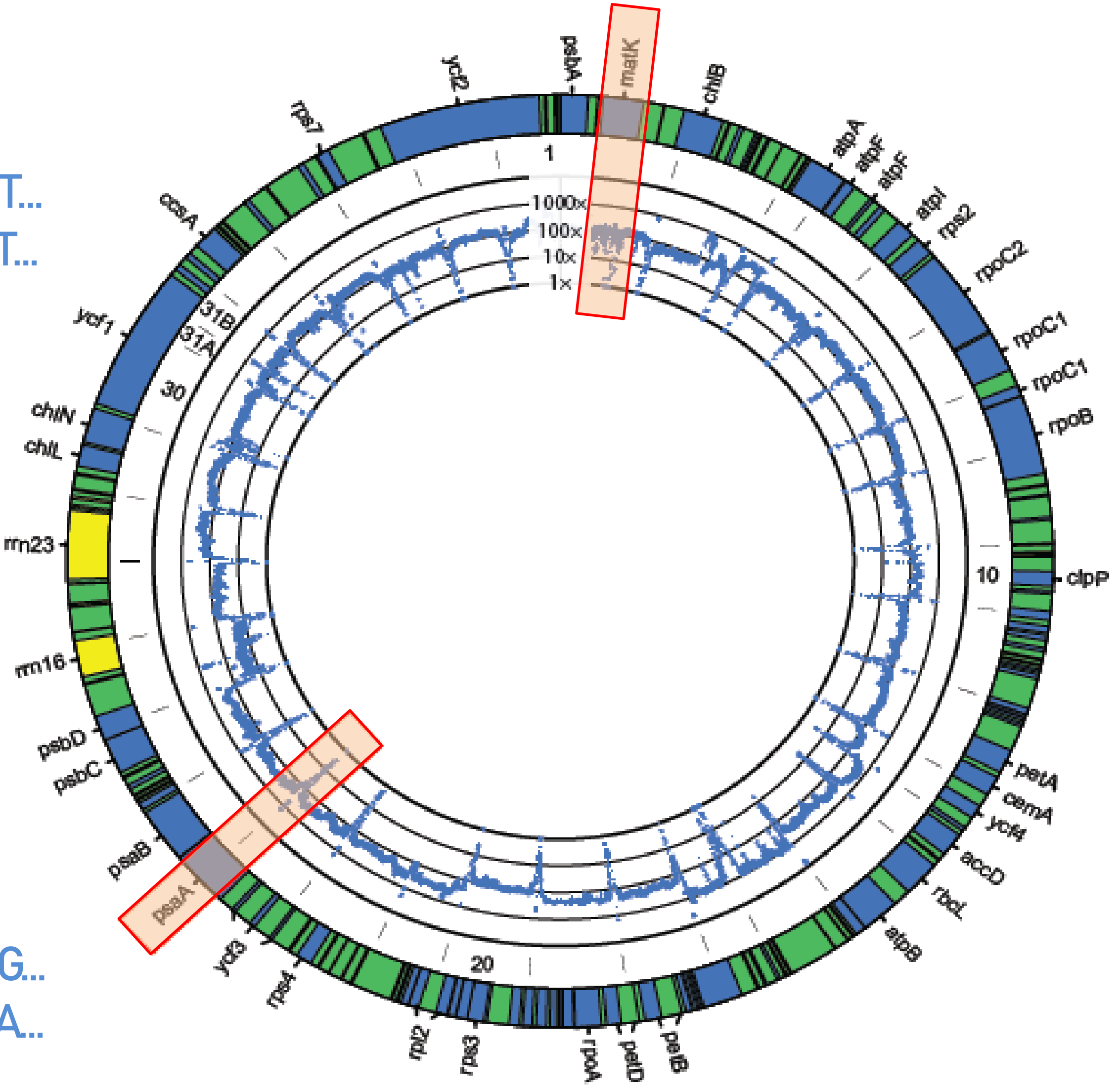
Whitebark	...CTC GTC ACG ATT...
Limber	...CTC GTC TCG ATT...



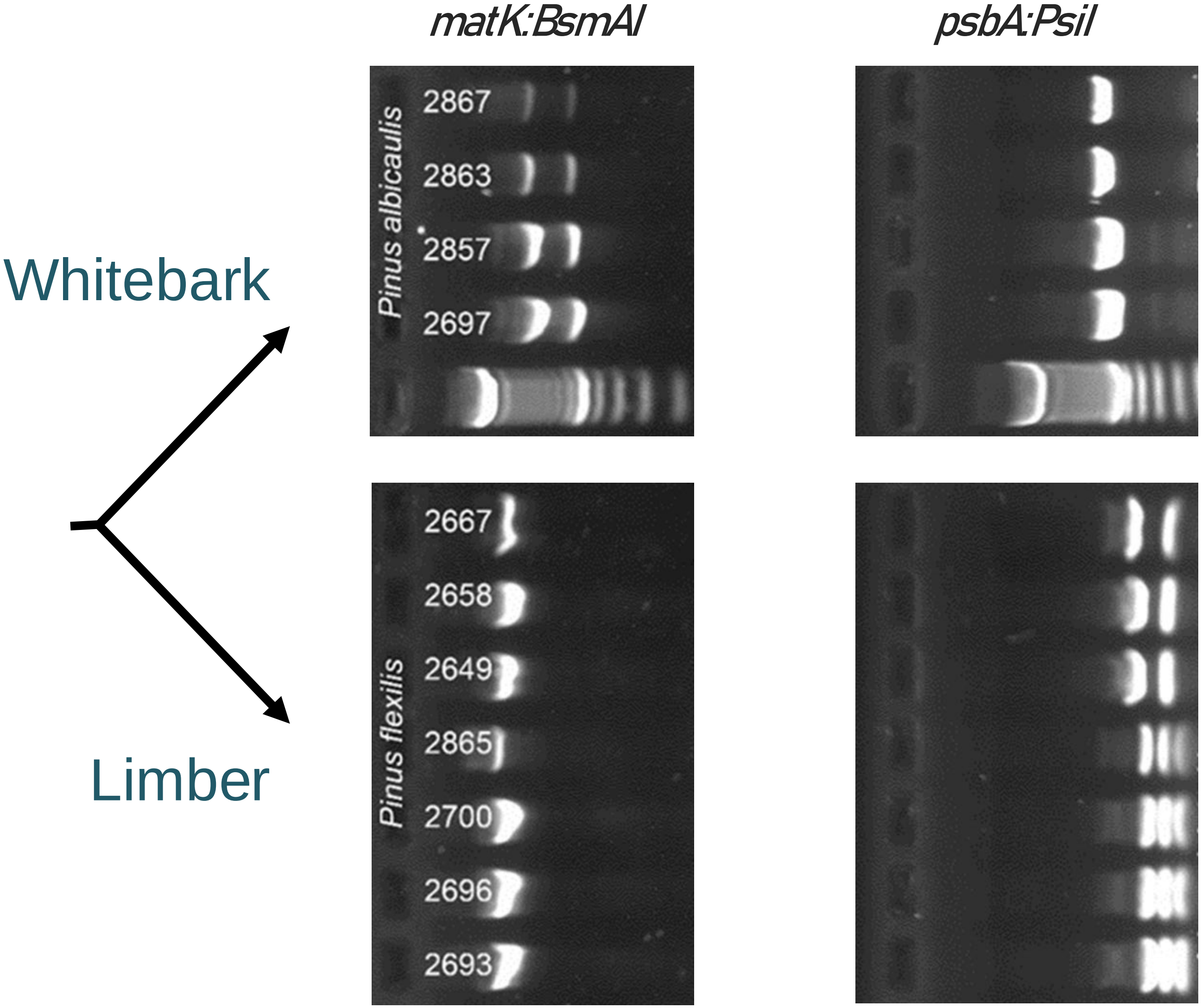
matK:*BsmAI*

psbA:*PsiI*

Whitebark	...CTC TTA TAA AAG...
Limber	...CTC CTA TTA AGA...



6. Genomics and taxon ID



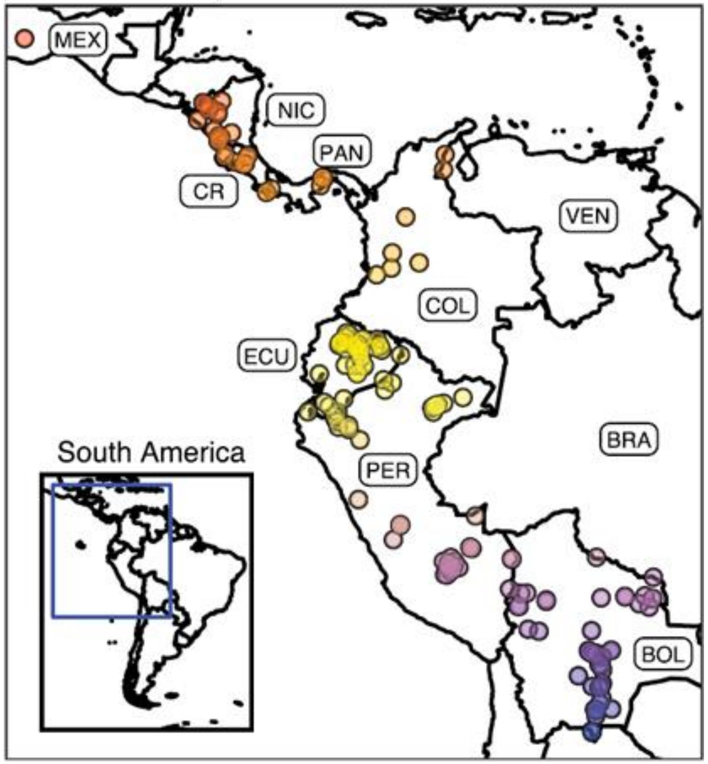
Whitebark

Limber

7. Taxon ID: a cautionary tale

Spanish Cedar (*Cedrela odorata* and relatives)

- Mahogany relatives spanning S. Mexico to Bolivia , 10 – 20(?) species
- Target of illegal timber harvest in Latin America; listed CITES II
- Controversial taxonomy. Can DNA clarify species/wood ID for CITES?



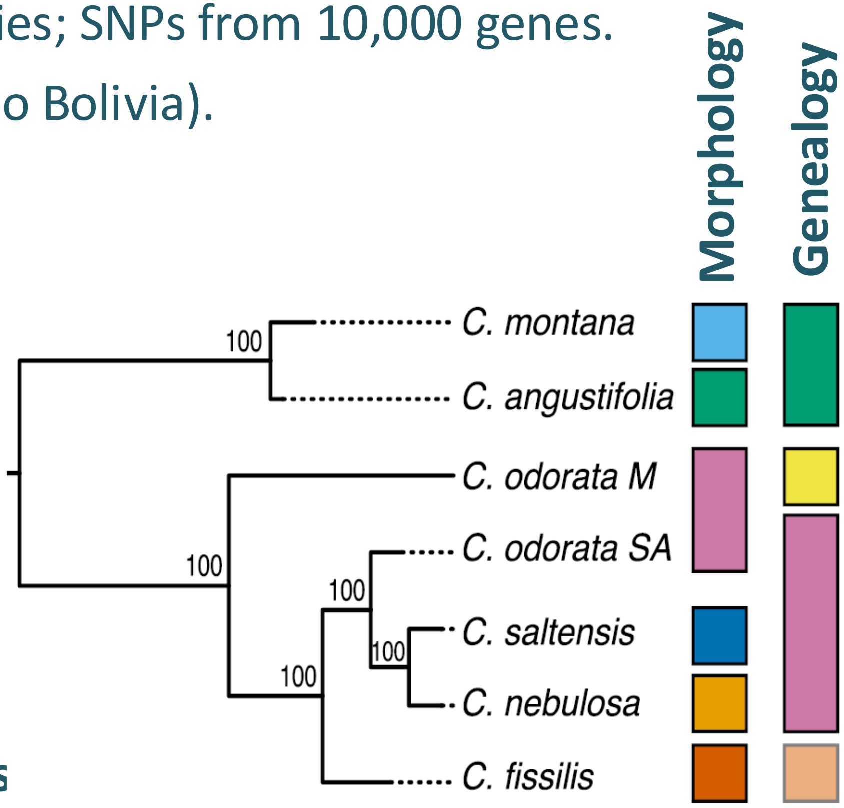
7. Taxon ID: a cautionary tale

Very large genetic analysis

- 170 trees from 6 most abundant species; SNPs from 10,000 genes.
- Samples spanning full range (Mexico to Bolivia).
- Multiple classification models.

Results

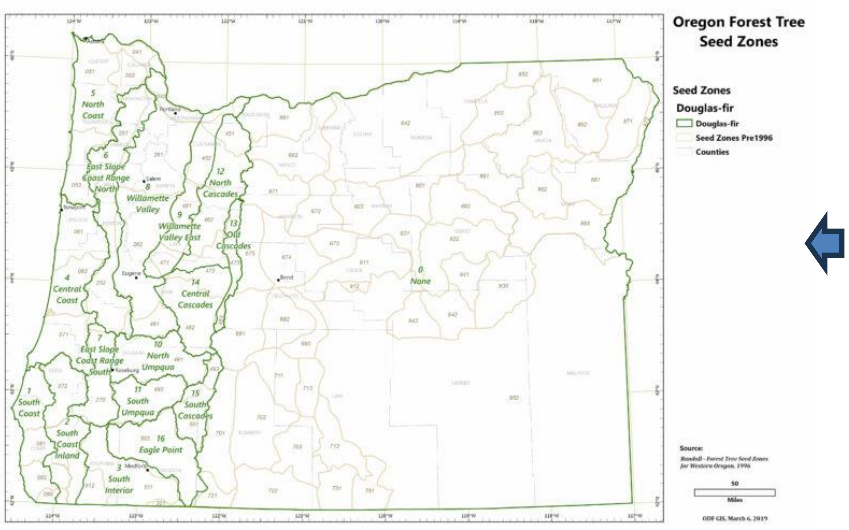
- *C. odorata* in Meso- and South America are different species.
- Best model supports 4 species; narrow endemics included in widespread species (*ssp? var?*).



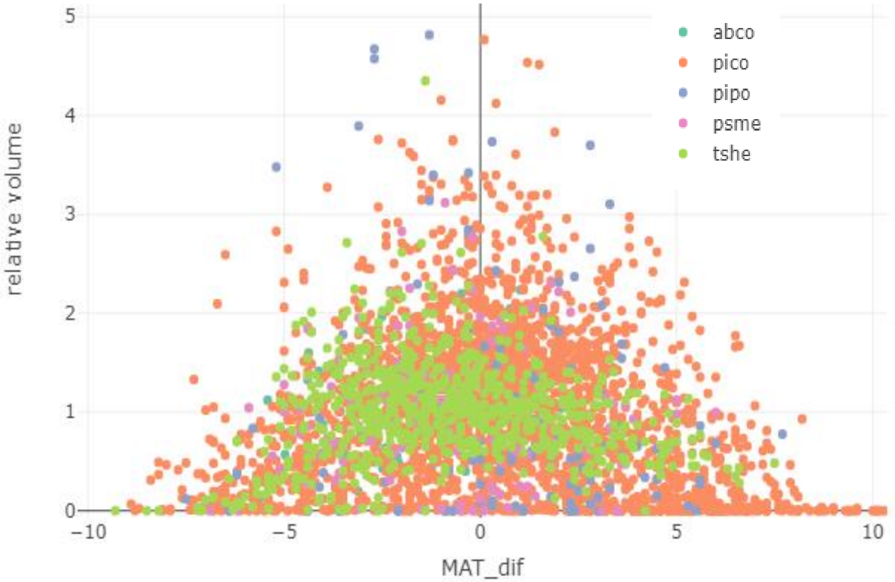
1. **“Species” definitions may be at odds genealogies**
2. **Highly-similar species may not be resolved with DNA**

8. Can DNA be used for seed zones?

Restated… ***can genomic data predict seed transfer zones?***



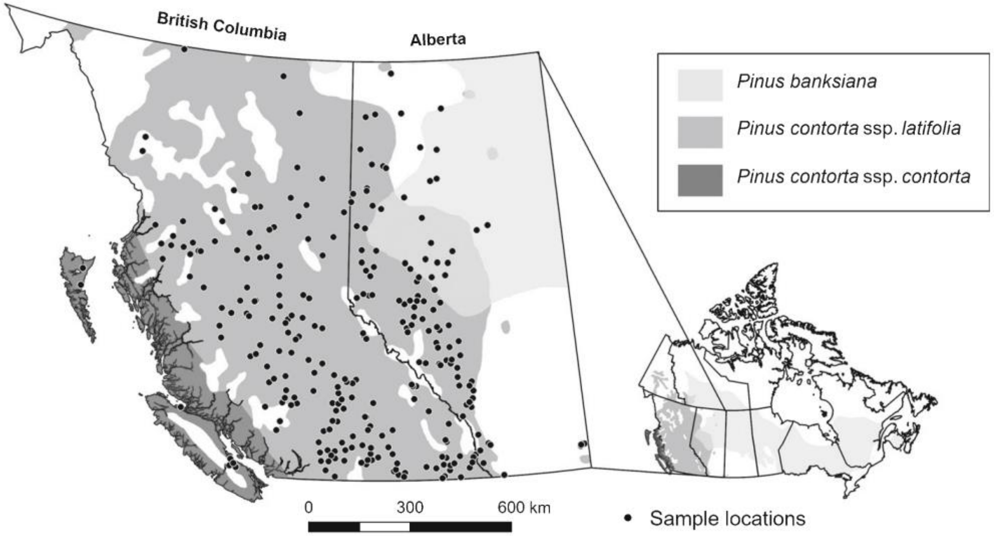
Seed zones to minimize risk of maladaptation



Plant in cooler climate than origin

Plant in cooler climate than origin

8. Can DNA be used for seed zones?



Genome-wide genetic-environmental associations

Pinus contorta

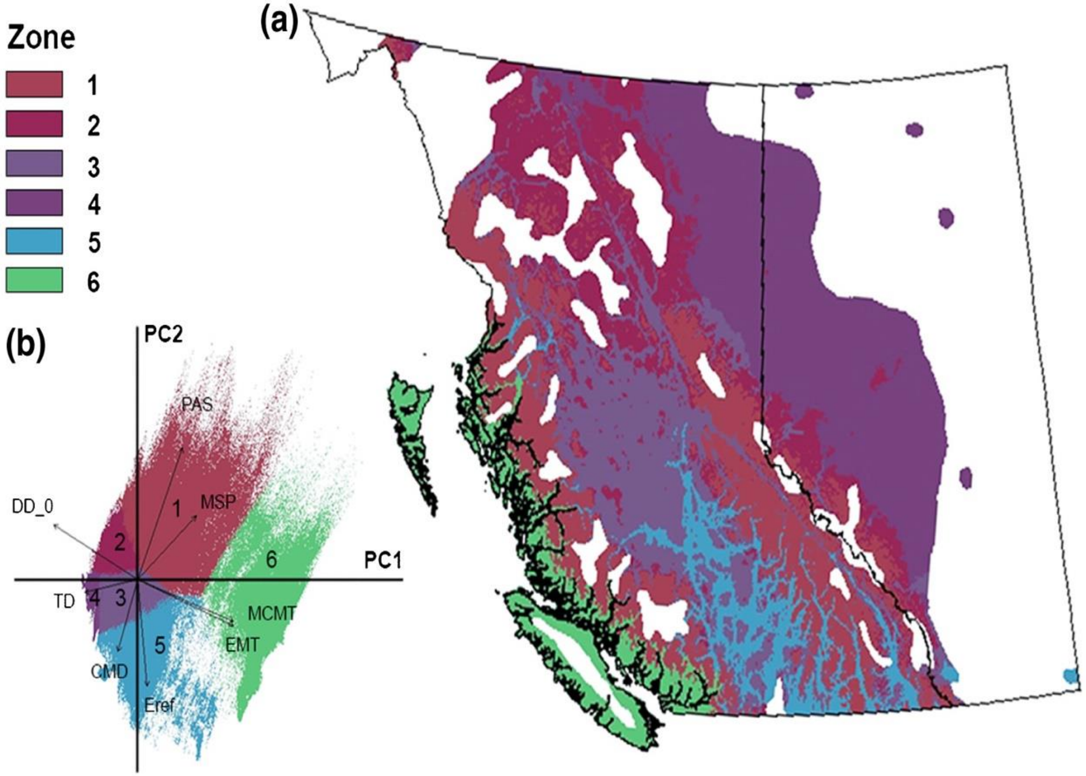
- **1900 trees**
- **36,000 SNPs**

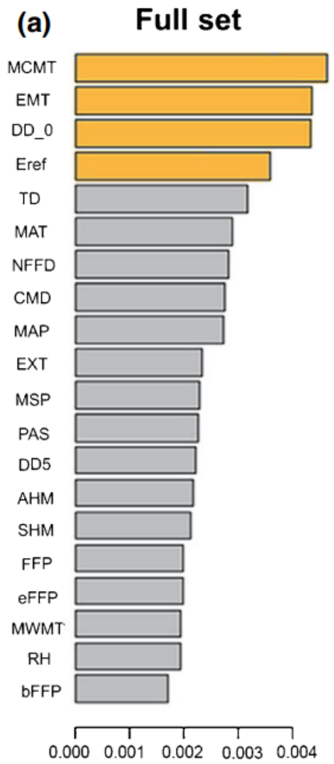
Yu et al., 2022. doi: 10.1111/nph.18223

- ***Landscape genomic*** **methods associate SNP genotypes (instead of phenotypes) with environmental gradients.**
- **Model is built between SNPs and climate using random forests.**
- **Model aggregates relationships across SNPs to predict allelic gradients as function of climate.**
- ***Assumption!*** **Genomic gradients mirror responses of genes responsible for adaptation.**

8. Genomics and local adaptation

Genomic data predict seed transfer zones in Pinus contorta





For more information

1. Review on forest genetics and genomics

- Neale and Kremer, Forest tree genomics: growing resources and applications. doi:10.1038/nrg2931.

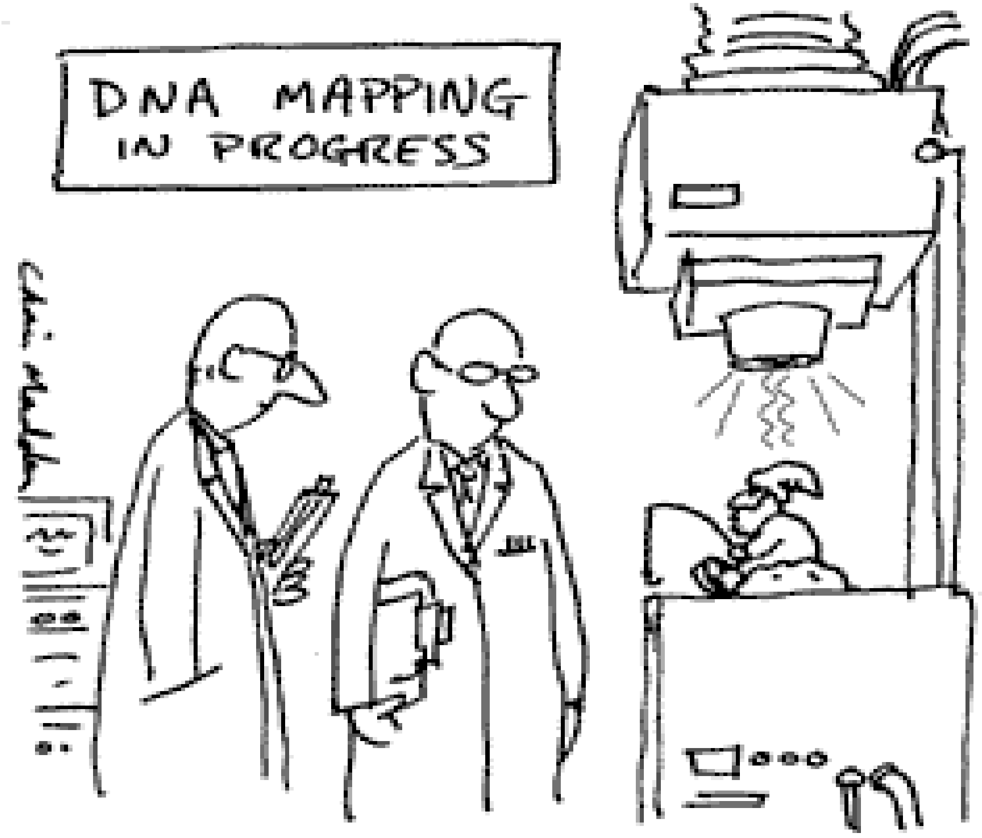
2. Examples of genomics methods in trees

- Wright et al., SNPs in a large genomic scaffold are strongly associated with Cr1^{R}, major gene for resistance to White Pine Blister Rust in range-wide samples of Sugar Pine. doi.org/10.1094/PDIS-08-21-1608-RE. doi:10.3390/f10121060.
- Cronn et al., Range-wide assessment of a SNP panel for individualization and geolocalization of bigleaf maple (*Acer macrophyllum* Pursh). doi.org/10.1016/j.fsiae.2021.100033
- Parks et al., Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. doi.org/10.1186/1741-7007-7-84
- Alongi et al., An economical approach to distinguish genetically needles of Limber from Whitebark pine.
- Finch et al., Cryptic species diversity in a widespread neotropical tree genus: The case of *Cedrela odorata*. DOI: 10.1002/ajb2.16064
- Telfer et al., Parentage reconstruction in Eucalyptus nitens using SNPs and microsatellite markers: A comparative analysis of marker power and robustness. DOI:10.1371/journal.pone.0130601
- Fitzpatrick and Keller, Ecological genomics meets community-level modelling of biodiversity: mapping the genomic landscape of current and future environmental adaptation. doi: 10.1111/ele.12376
- Yu et al., Using landscape genomics to delineate seed and breeding zones for lodgepole pine. doi: 10.1111/nph.18223

3. North American Forest Genetics Society

- https://treegenesdb.org/nafgs

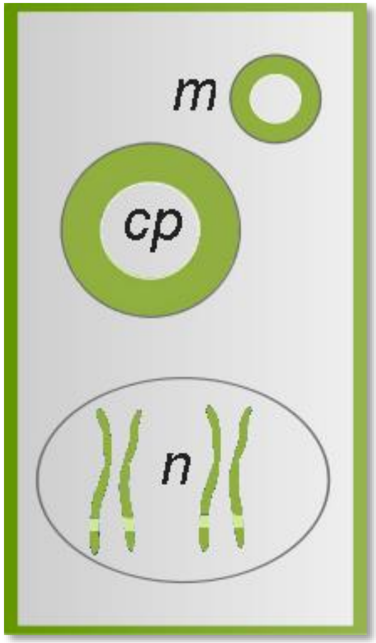
Thank you!



THE G-NOME PROJECT

Genomics Basics

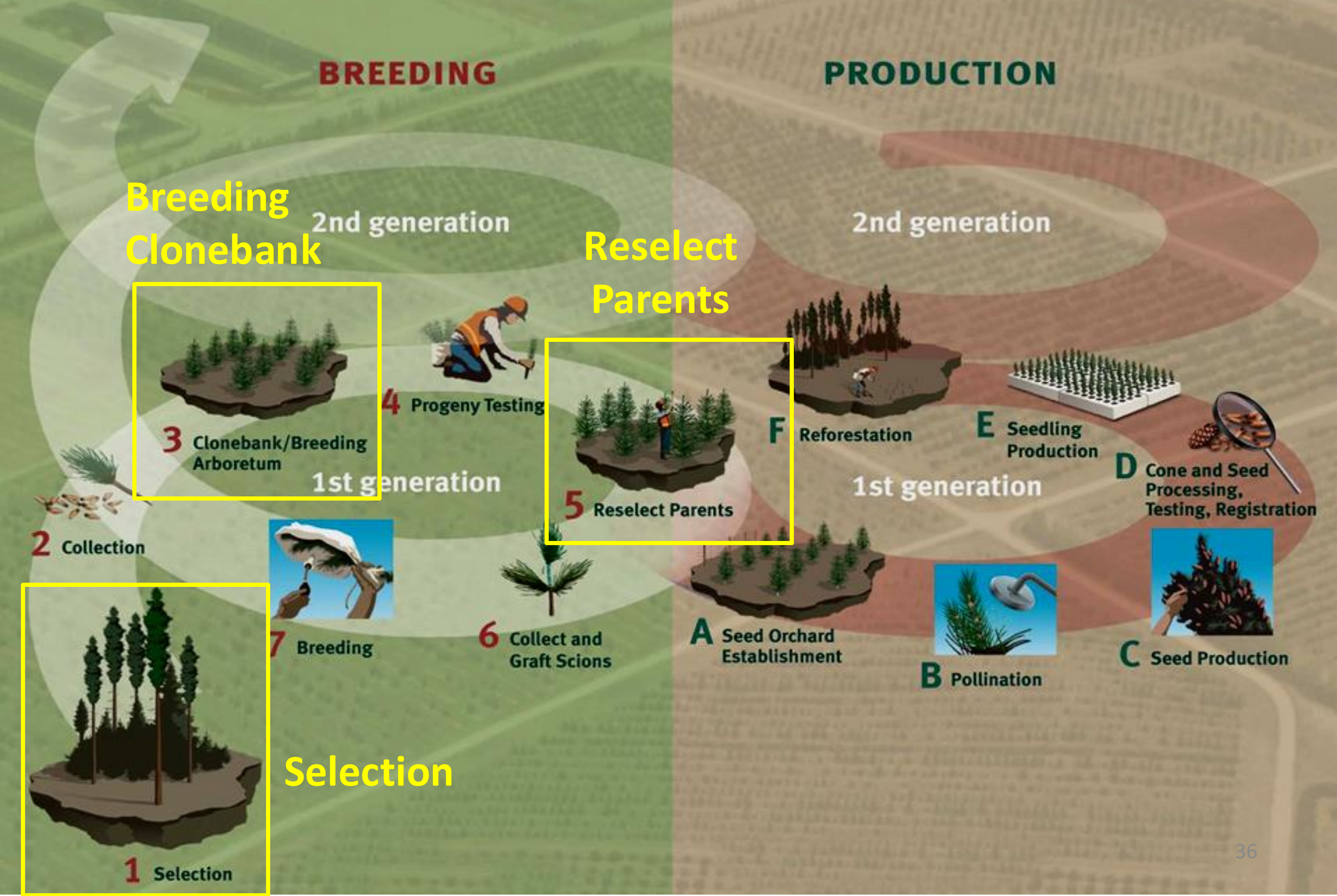
1. **Genome: the total DNA in an individual's *organelles.***
2. **Genomics*:* the study of genome-scale data from a tissue, individual, population, or multiple species.**



Example genomes	Douglas-fir 	You 
nucleus	***>19,000 Mbp****	***3,300 Mbp***
genes	***52,000***	***21,000***
mitochondrion	***4 Mbp***	***0.016 Mbp***
chloroplast	***0.12 Mbp***	**-**

** Mbp = million base pairs*

Tree Improvement Cycle



Breeding Clonebank

Reselect Parents

Selection