

# Tree Improvement 101: Modes of selection in tree improvement programs *MODULE 2*



**Carolyn (Carrie) Pike, PhD**

USDA Forest Service, State, Private and Tribal Forestry Eastern Region  
Reforestation, Nursery, and Genetic Resources (RNGR)



**Kevin Potter, PhD**

USDA Forest Service  
Southern Research Station

---

**RNGR** REFORESTATION, NURSERIES  
& GENETIC RESOURCES

---



<https://rngr.net/>

# Reviewing module 1

- ✓ landscape patterns of natural genetic variation, eastern hemlock, (white pine and jack pine examples)
- ✓ clinal vs ecotypic variation (trees are mostly clinal – genetically-distinct ecotypes are uncommon)
- ✓ life history traits:
  - ✓ most temperate trees species are wind pollinated;
  - ✓ seed size is highly variable among tree species.
- ✓ quantitative traits= polygenic traits- many genes control their expression.
- ✓ progeny tests and provenance trials (different types of common gardens)
- ✓ This is how we begin to parse out Phenotype = Genotype + Environment (P=G+E).

# The progeny test and different modes of selection

In this section, we'll dig into the ways common gardens are used to make selections for tree improvement.

$$\text{Phenotype} = \text{Genotype} + \text{Environment} + \text{G} \times \text{E}$$

$$\text{Physical traits} = \text{(G) Genetic effects} + \text{(E) Environment effects} + \text{G} \times \text{E Interactions}$$

- ✓ Soils
- ✓ Slope
- ✓ Aspect
- ✓ Spacing
- ✓ Silviculture, etc.

When a genotype's rank, relative to other genotypes, varies depending on the environment.

- use common garden studies to tease out these effects.
- use multi-site common gardens (MET: multi-environment trials) to tease out  $g \times e$
- The statistics of a MET trial are more complicated and resemble a repeated measures design.



# Single tree vs 2- or 4-tree row plot or 4-tree plots (randomized)

100	130	131
129	101	132
107	127	134
128	135	138

- We mostly use single tree plots because we can remove spatial variation with row and column effects alone.
- But -- planting in row plots allows you to thin alternate rows without losing any genotypes.

100	130	131
100	130	131
100	130	131
100	130	131
129	101	132
129	101	132
129	101	132
129	101	132
107	127	134
107	127	134
107	127	134
107	127	134
107	127	134
128	135	138
128	135	138
128	135	138
128	135	138

# Single tree vs 2- or 4-tree row plot or 4-tree plots (randomized)

100	130	131
129	101	132
107	127	134
128	135	138

We mostly use single tree plots because we can remove spatial variation with row and column effects alone.

But -- planting in row plots allows you to thin alternate rows without losing any genotypes.

138	135	131
100	130	130
131	131	131
100	130	100
129	101	132
101	101	138
107	101	132
129	129	132
107	134	107
132	127	100
127	129	134
107	127	101
128	135	100
129	130	138
128	129	134
132	135	138

# Randomized complete vs incomplete block design

100	130	131
129	101	132
107	127	134
128	135	138

Each family is represented by one (or more) trees in each block.

It is “complete” because each family occurs in every block. “Incomplete” means that some blocks may be missing a tree from a few families. This is OK.

1	2	3	4	5	6
100	130		100	130	131
129	101	132	129	101	132
107	127	134	107	127	134
128	135	138	128	135	138
100		131	100	130	131
129	101	132		101	132
107	127	134	107	127	134
128	135	138	128	135	138

# Two primary modes of selection:

## Forward selection:

- A progeny test is 'rogued' to remove worst-performing trees from worst-performing families.
- The remaining trees are used for seed collection.
- The spacing becomes uneven.

*By converting the measured trees into a seed orchard we are estimating their genetic worth as parents.*

## Backwards selection:

- A progeny test is measured, but the information from the progeny is used to evaluate the mother trees.
- We're looking backwards a generation.

*A separate clonal seed orchard is developed by grafting mother trees whose progeny have the best performance.*

# Forward selection: Seedling-Seed Orchard

## Typical example:

- 1000 trees
- 10 randomized complete blocks
- 100 families (one tree per family per block)

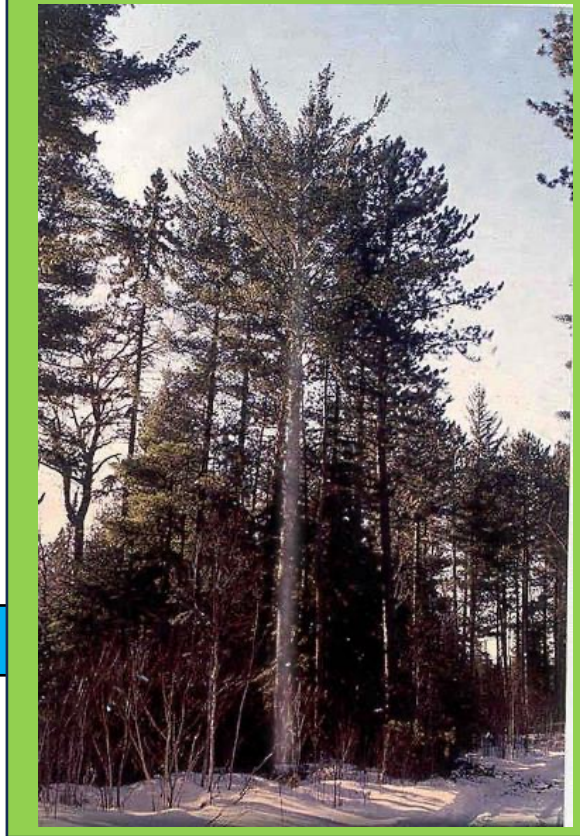


# Forward selection:

1. Select plus trees, collect open-pollinated seed



2. Progeny from hundreds of mother trees planted in common garden



4. Use remaining trees as a seed orchard.



3. Measure and remove ("rogue out") worst trees from worst families





# Forward selection:

We rogue out 95% of 1000 trees in the seed orchard. (This is a heavy roguing: you might keep 10% (instead of 5%))

## Possible scenario:

- 50 trees remain, from 30 families
- 1-5 trees per family

## Benefits:

- Cheap! No separate progeny test is needed. Lots of genetic diversity even after heavy roguing.

## Tradeoffs:

- No more data collection opportunities after roguing!
- We are assuming that best trees produce the best progeny – but we may be retaining some poor performing individuals.
- Uneven spacing after roguing may be a challenge for mowing and managing.

# Backward (family) selection

1. Select plus trees



2. Progeny from hundreds of mother trees planted in a multi-site progeny test



3. Measure progeny test over time



4. Create a separate grafted orchard to use as a seed source.



# Backwards selection: Separate progeny test & clonal seed orchards

## Progeny test:

- 1000 trees
- 10 randomized complete blocks
- 100 families (one tree per family per block)

## Seed orchard\*:

- 300 clones
- 30 genotypes, 10 ramets per clone
- Completely randomized (so no relatives are adjacent)
- Select maternal genotypes based on the progeny test results.

*\*note the wide spacing between ramets*



# Why both creating a grafted seed orchard?

If a progeny test reveals that a tree typically produces excellent offspring, then we want an exact copy of that “mother” tree in our orchard.



# Backwards selection

**We may systematically thin the progeny test, but the seed orchard is maintained as a separate entity.**

- Thin the clonal seed orchard or establish a new one using grafts from the top mother trees.

## **Benefits:**

- Progeny test and data collection can be maintained in perpetuity.
- Most efficient for selecting parents

## **Tradeoffs:**

- You need separate plantings for testing and seed production.

# Three main types of selection

## Mass selection

choosing an individual solely on the basis of its phenotype, without regard to information about performance of ancestors, siblings, offspring or other relatives.

## Family selection

choosing a 'family' on the basis of its average phenotypic value. Useful for traits with low heritability, where individual phenotypes are not a good reflection of genotypes.

## Family plus within family selection or family "index" selection

two-stage method involving selection on families, followed by selection within families. Works well with low heritability traits.

Make sure the final orchard has trees from at least 20 families for genetic diversity!

# Unpacking genetic variance

Our goal is always to improve “G” so that we can have confidence in seed performance when planted on a multitude of different sites.

$$\text{Phenotype} = \text{Genotype} + \text{Environment} + G \times E$$

## Genetic

- Additive genetic variation
- Non-additive

## Environment

- ✓ Soils
- ✓ Slope
- ✓ Aspect
- ✓ Spacing
- ✓ Silviculture, etc.

## $G \times E$ Interactions

**Dominance** = heterozygotes are masked

**Epistasis** = genes interact with other genes in unpredictable ways.

**Epigenetic effects**, etc.

# Tree improvement programs mostly focus on additive genetic variation

*“Additive variance is due to the cumulative effects of genes at all gene loci influencing a trait.” Falconer & Mackay 1996*

Additive genetic variance is the variation that is shared by relatives.

It is the variation that is NOT related to epistasis or dominance or other confounding genetic effects.

If additive variance is present, the heterozygotes can be distinguished from the homozygotes and the families differentiate.



*We'll unpack this in the next few slides....*

# Unpacking additivity: qualitative example

Parent 1: aa



+

Parent 2: AA



=

Progeny: Aa

?

if 'A' is dominant then Aa =



If 'a' is dominant then Aa =



**If neither allele is dominant then Aa =**



If the trait is *additive*, then if you *add* an allele, the phenotype should increase by some measurable amount.

If this is the resulting phenotype then the trait is inherited additively

# Unpacking additivity – quantitative example

- Average weights of mice.
- Midpoint is 10.  
*Dwarfing gene is pg (=pygmy).*

	Genotype		
	+ +	+ pg	pg pg
Weight in grams	14	12	6

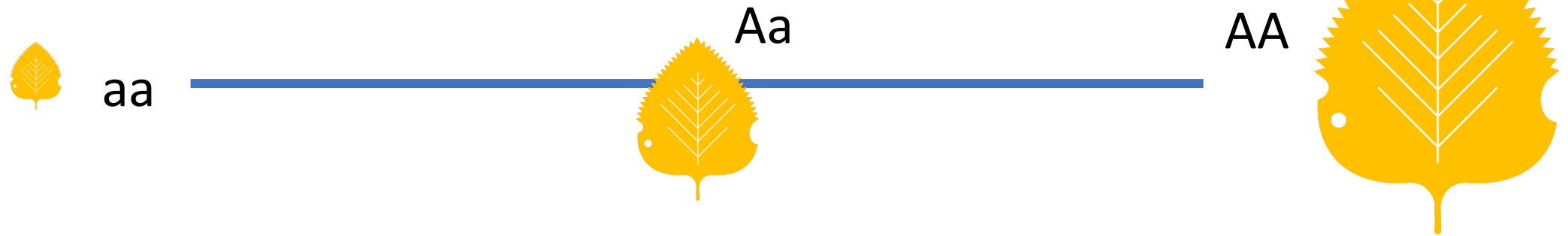
Redrawn from Falconer and Mackay, 1996.

Does this trait appear to be additively inherited?

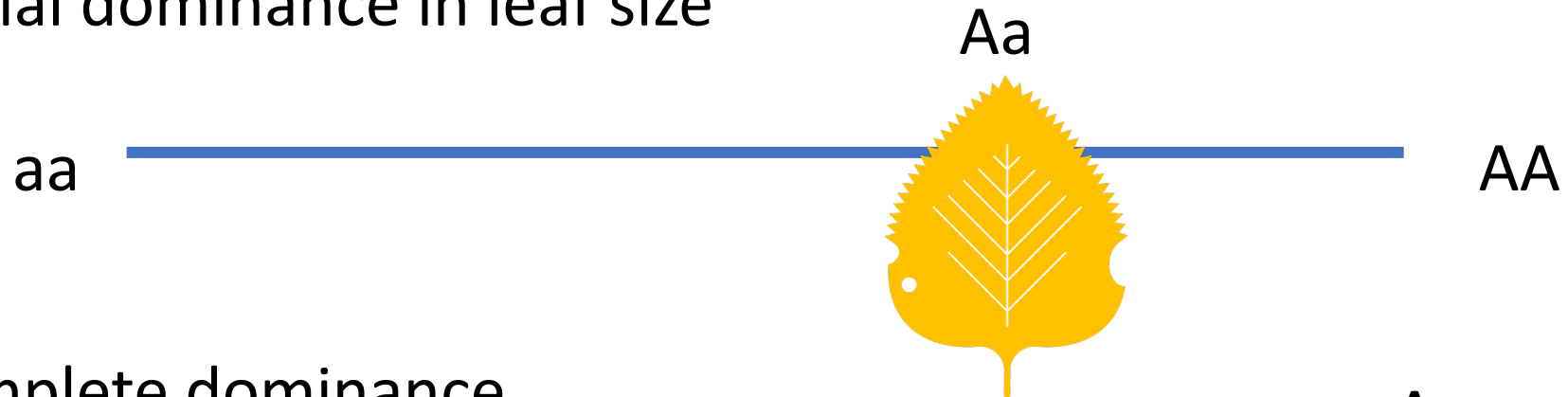
***Sort of.***

The heterozygote is not at the midpoint. The “+” gene is expressing partial dominance.

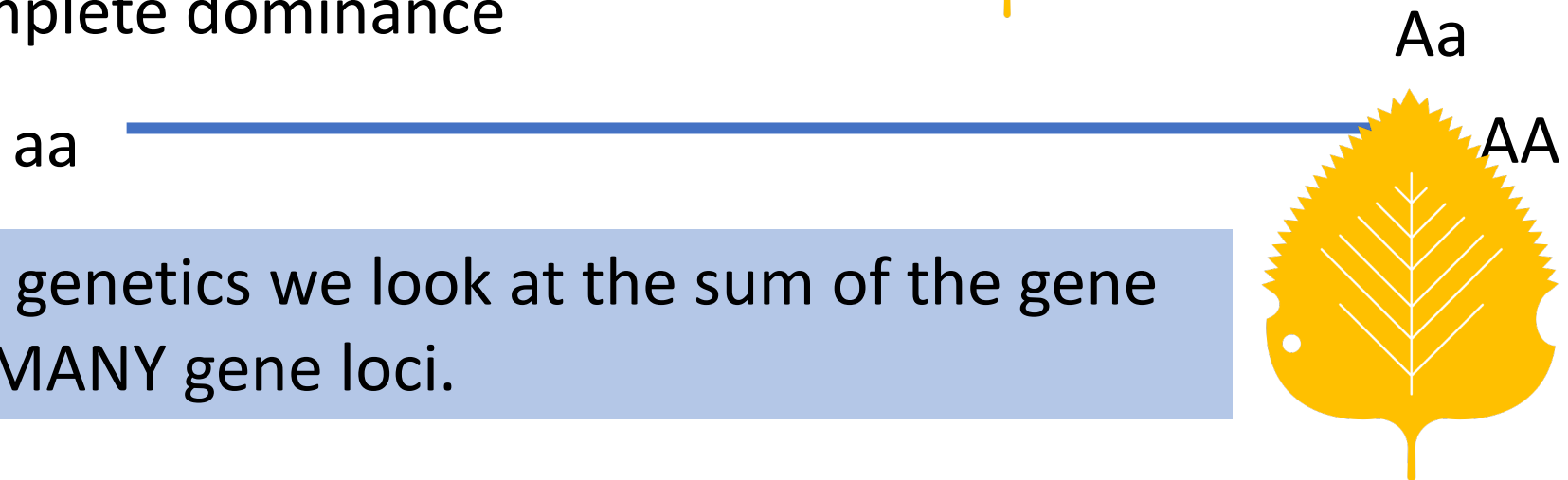
# 1. Additive gene interaction in leaf size



# 2. Partial dominance in leaf size



# 3. Complete dominance



In quantitative genetics we look at the sum of the gene effects across MANY gene loci.

*Redrawn from  
Zobel and  
Talbert, Ch 2,  
p.48,*

# Unpacking additivity – quantitative example

In forest tree systems, we compare the progeny of different families and we usually do NOT know what genes are controlling the traits we are studying!

Is tree height additively inherited in this population?

	Mother genotype			
Family number	101	105	110	112
Mean tree height at year 10 (+SE)	20 (+6)	12 (+3)	30 (+5)	40 (+7)

← Progeny test data

The families have very different means with tight standard errors, so there is likely a high degree of additivity controlling this trait.

# Unpacking additivity for traits in forestry

- Adaptive traits in forest trees are often controlled by many genes, but most traits will express **some level of additivity across the genes that are involved in the trait's expression.**
- When additivity  $> 0$ , then siblings will bear some resemblance to each other.
- The higher the additivity, the greater the resemblance of siblings to each other.
- If a trait is governed by alleles with dominance, or other non-additive gene interaction, then the resemblance among siblings within families may be low, inconsistent, or unpredictable.

# How do we quantify the amount of additive genetic variation?

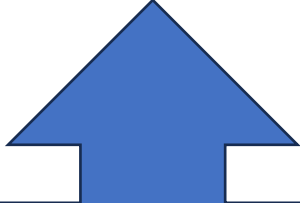
With an estimate of heritability.

*“The most important function of the heritability in metric characters is its predictive role expressing the reliability of the phenotype value as a guide to the breeding value.”*

*Falconer and Mackay, 1996, chapter 10*

Heritability is the proportion of variation attributable to genetic differences among families

**Heritability also measures efficiency of selection.**



What is efficiency? How much improvement can be made at each generation of selection.

# Type main types of heritability: narrow-sense and broad sense

Narrow sense heritability is easily estimated with a progeny test of open-pollinated seedlings

Broad sense heritability can only be estimated with full-sibs or clones.

## Narrow sense heritability ( $h^2_{ns}$ )

can be calculated in open-pollinated progeny tests.

*(The symbol  $h^2_{ns}$  is derived from 1921 terminology)*

$h^2$  in the narrow sense =  $h^2_{ns}$   
Additive variance / Phenotypic variance  
Short version:  $V_A/V_P$

Variances are depicted  
as  $\sigma^2$  while standard  
deviations are depicted  
as  $\sigma$

$V_A = \frac{1}{4}$  of the family variance,  $\sigma^2_{family}$   
 $V_P =$  Genetic + environmental variance

$$4 * \sigma^2_{family} / [\sigma^2_{family} + \sigma^2_{block} + \sigma^2_{residual}]$$

These  
variance  
components  
are derived  
from ANOVA

*Narrow-sense  
heritability is also  
used in evolution to  
estimate shifts over  
time.*

Using full-sib or a clonal progeny test, we can estimate broad-sense heritability for our population

$H^2$  in the broad sense =  $H^2_{bs}$

Genetic variance (additive AND non additive) / Phenotypic variance

Short version:  $V_G/V_P$

$V_G$  = genetic variance =  $\sigma^2_G$

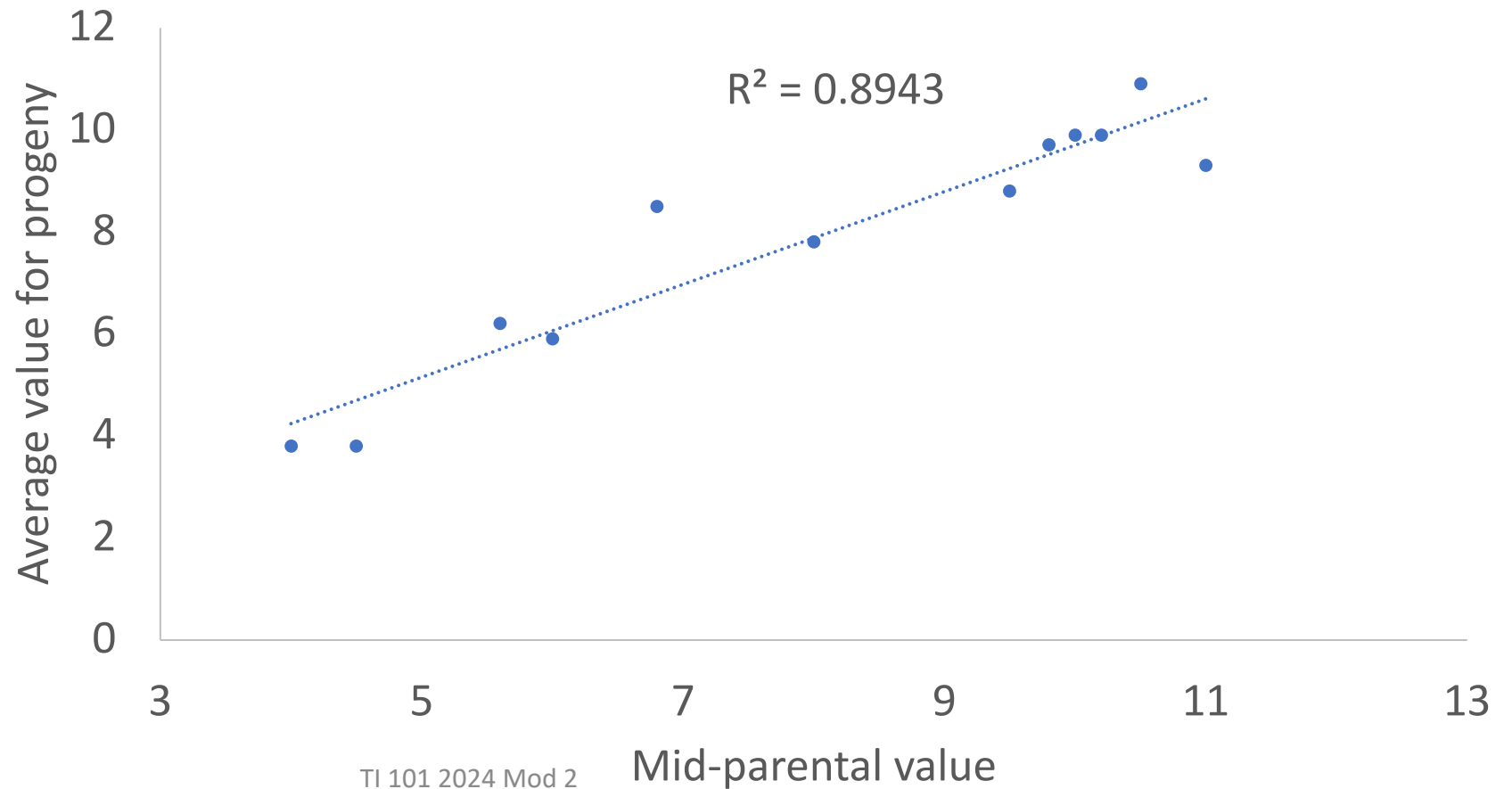
$V_P$  = Genetic + environmental variance

$$2 * \sigma^2_G / \sigma^2_G + \sigma^2_{\text{block}} + \sigma^2_{\text{residual}}$$

These variance components are derived from ANOVA

Heritability can be estimated with parent-offspring regressions. Plot the mid-parent values vs the progeny performance for a single trait. This is a type of narrow-sense heritability.

But trees have a long generation time so we usually rely on resemblance among progeny to measure heritability.



We calculate heritability from a progeny test using a linear mixed model, a type of ANOVA (analysis of variance).

All factors are set as random effects to calculate variance components needed for heritability estimates.

# Steps in calculating heritability from a progeny test

- Common garden study with known half-sibs or full sibs. Ideally will have several hundred families in the study.
- Trees are planted in a Randomized Complete or Incomplete Block design
- Ideally we have more than 1 site, but we start by estimating heritability separately for each site.
- Single-tree or multiple-tree plots
- **Run ANOVA in linear mixed models framework**
- **Set all factors as random effects (block, family)**
- **Extract variance components for all random effects and construct the heritability.**

site	Bk	TreeID	Family	Col	Row	Ht_cm
212	1	212-1983	212	10	12	21
212	1	210-1983	210	10	14	28
212	1	213-1983	213	10	24	37
212	1	215-1983	215	10	5	30
212	1	216-1983	216	11	11	22
212	1	220-1983	220	11	25	32
212	1	221-1983	221	11	5	22
212	1	206-1983	206	11	6	42
212	1	205-1983	205	11	7	30
212	1	202-1983	202	12	23	23
212	1	201-1983	201	13	13	26
212	1	204-1983	204	13	25	25
212	1	207-1983	207	13	9	12
212	1	208-1983	208	15	2	14

# Simplest ANOVA model: one site progeny test

$$y = X\beta + Zu + e \quad \leftarrow \text{This is the model in matrix notation}$$

$$\textit{Trait} = \mu + \textit{Block}^* + \textit{Family}^* + \textit{error}$$

**Trait** may be tree height, diameter, volume, resistance, survival (binary), straightness, etc.

$\mu$  is the overall grand mean, or the y-intercept

**Block** is the randomized complete block

**Family** may be open-pollinated mother tree lines, or full-sib.

**Error** is all the rest of the variance that's unaccounted for by the design.

\*These factors are set as random effects to calculate variance components needed for heritability estimates. The error is automatically designated as a random effect.

# Use the linear mixed model to derive variance components

$$\textit{Trait} = \mu + \textit{Block} + \textit{Family} + \textit{error}$$

Sources of variation	Df	SS, MS, EMS	Variance components
Block	B-1		$\sigma^2_{\text{block}}$
Family	F-1		$\sigma^2_{\text{family}}$
Block x Family	(B-1)(F-1)		$\sigma^2_{\text{blk*fam}}$
Error	BF(T-1)		$\sigma^2_{\text{error}}$

F,B,T refer to the number of families, blocks, and tree per family/block/plot.

# How to construct the “individual” narrow-sense heritability for open-pollinated progeny test



REMEMBER:

If variation from the blocks and residuals are excessively high, then the quotient (heritability) will be LOWER than a site where genetics accounts for most of the observed variation.

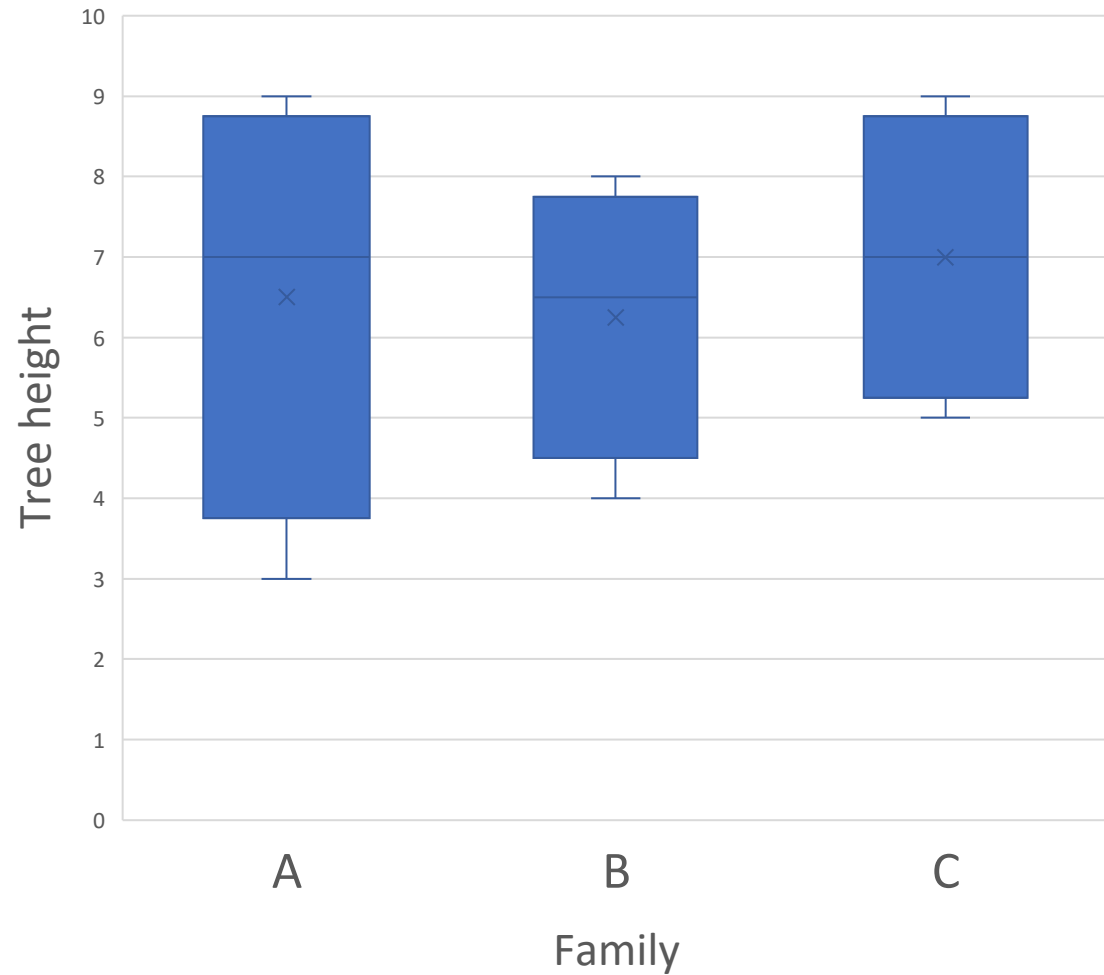
$$h_i^2 = \frac{4 * \sigma^2_{family} = V_A}{\sigma^2_{error} + \sigma^2_{block} + \sigma^2_{family} = V_P}$$

Half sib family

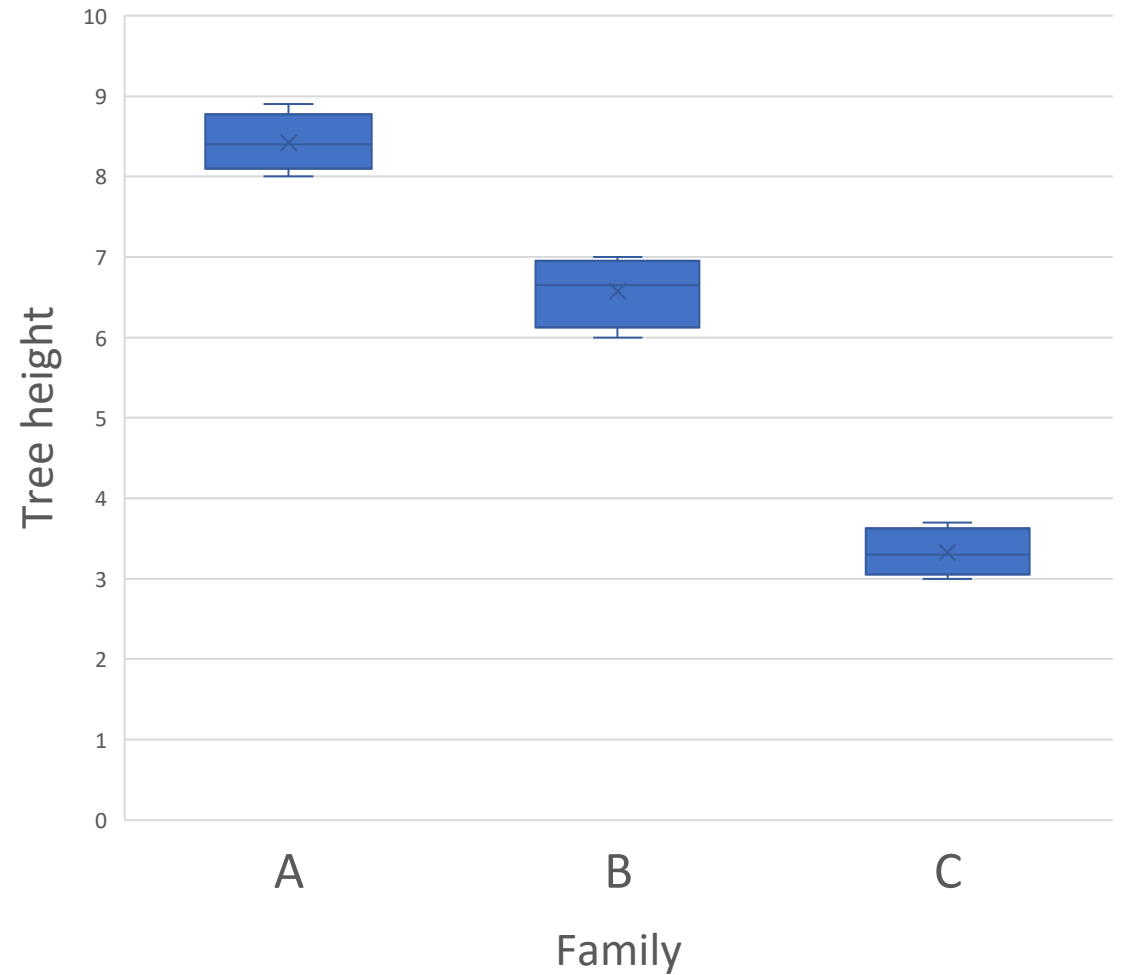
**For open-pollinated tests, the additive genetic variance = ¼ of the  $V_{family}$ . This is based on the half-sib covariance, or the probability that two half sibs share alleles that are identical by descent.**

Derivations can be found in Falconer and Mackay 1996.

Box and whisker plot showing LOW heritability



Box and whisker plot showing HIGH heritability



We can calculate different types of heritability for different modes of selection

We can calculate family heritability, and within family heritability to figure out which mode of selection is more efficient.

Heritability of half-sib family mean:  $h^2_f$  this heritability reflects the efficiency of selecting families vs individual trees

$$h^2_{family} = \frac{V_{family}}{\frac{V_{error}}{nr} + \frac{V_{block}}{b} + V_{family}}$$

Additive variance

Phenotypic variance  
(includes  
environmental noise)

“ $n$ ” is the # of trees per family per rep;  
 $b$  is the # of blocks in the design

Derivations can be found in  
Falconer and Mackay 1996.

Within family heritability :  $h^2_{wf}$  this heritability reflects the efficiency of selecting individuals within families

$$h^2_{family} = \frac{3 * V_{family}}{V_{error}} \quad \begin{array}{l} \frac{3}{4} \text{ of additive variance} \\ \text{Within family variance} \end{array}$$

We can do different calculations with the variance components to figure out efficiency of different selection types.

# Different types of heritabilities are calculated to estimate efficiency of different selection types

Individual

$H^2_i$  individual broad sense\*

$h^2_i$  individual narrow sense

Family

$h^2_{hs}$  half sib, narrow sense

$H^2_{hs}$  half sib, broad sense\*

Within-family

$H^2_{fs}$  full-sib, narrow sense

$H^2_{fs}$  full-sib, broad sense\*

$h^2_{wf}$  within family, narrow sense

$H^2_{wf}$  within family, broad sense\*

\*broad sense heritability can only be estimated if you are testing clones or full-sibs

# Heritability will vary by site and over time

White oak progeny test in Indiana: same families at 3 sites measured at year 2

	$h^2_i$	SE_ $h^2_i$	$h^2_{fam}$	SE_ $h^2_{fam}$
Site 1	0.51	0.134	0.78	0.052
Site 2	0.53	0.114	0.69	0.076
Site 3	0.93	0.145	0.9	0.018

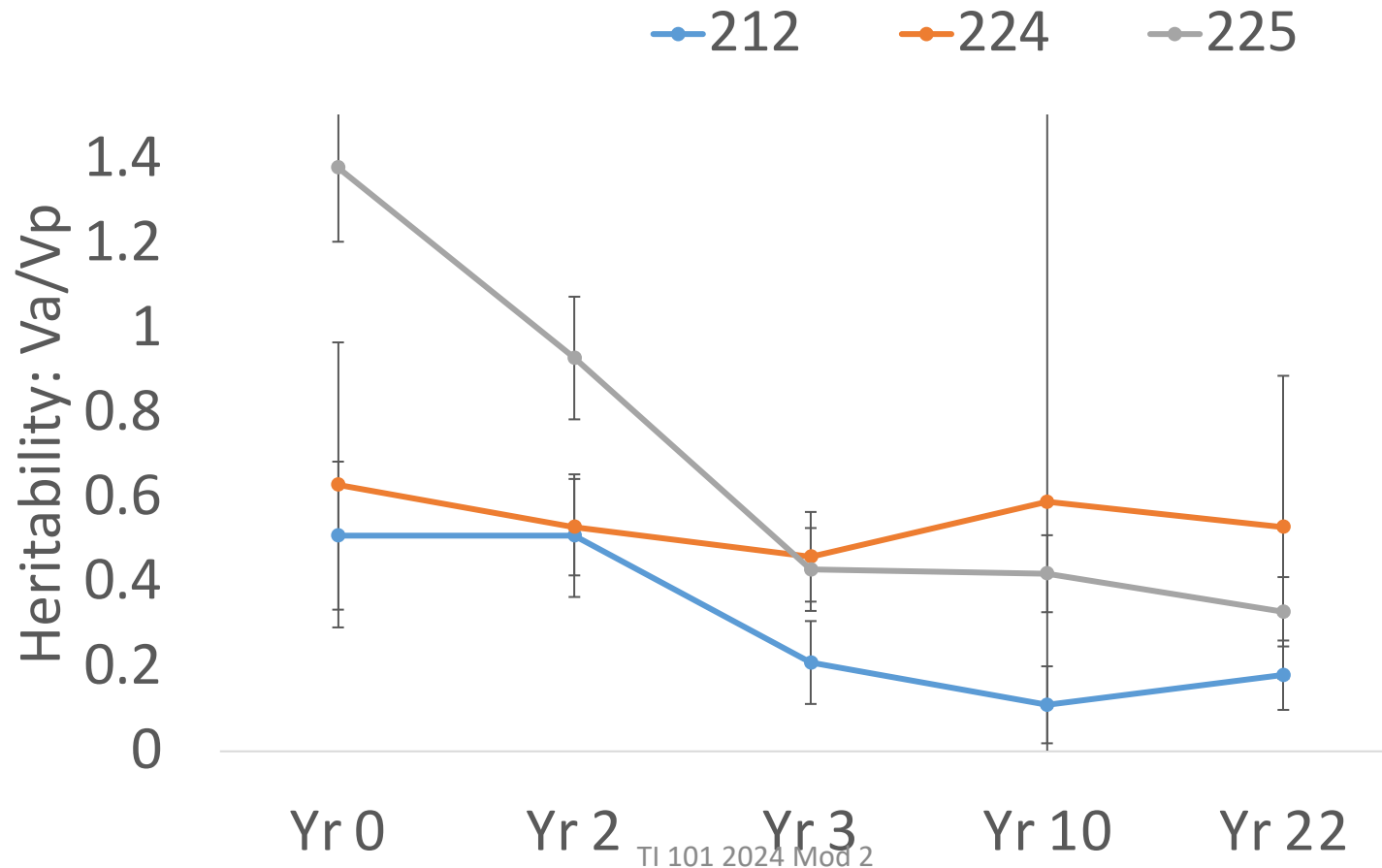
The sites had different overall variances so we calculated heritability separate for each site.

Heritability must have a standard error to be meaningful: must be different from zero!

Family heritability's are usually than individual heritability

# Heritability estimates also vary over time!

$h^2_i$  for tree height in *Quercus alba* at each of 3 sites from year 0-22



# Heritabilities can be calculated on binary variables. Data on *Quercus rubra*.

Site	Type	Pres/abs 12 ft log		Branchiness >=2, <2	
		Value	SE	Value	SE
115	$h^2_i$	0.34	0.378	0.91	0.269
115	$h^2_f$	0.59	0.297	0.82	0.056
117	$h^2_i$	0.00	0.00	0.74	0.284
117	$h^2_f$	0.00	0.00	0.79	0.077
119	$h^2_i$	0.41	0.242	0.73	0.237
119	$h^2_f$	0.65	0.151	0.78	0.067

# Side-notes on heritability



- Heritability ( $h^2$ ) helps assess whether a 'trait' is worth pursuing in a tree improvement program: value be  $> 0$ !
- Heritability is a number between 0 and 1.
- Heritability should always be accompanied by its standard error (which needs to be calculated).
- It can help assess the value of selecting individual trees vs families in a progeny test.
- Heritability for growth and form traits in forestry will range 0.10-0.30 and be higher for wood specific gravity and form (0.3-0.6) (White et al. 2007).

# Side-notes on heritability



- Heritability usually differs by sites and varies over time: the denominator will be HIGHER when the environmental noise is poorly controlled, leading to a smaller heritability.
- Many standard statistical packages can do linear mixed models. SAS: Proc mixed, glimmix (DO NOT USE Proc anova or Proc glm); R: lme, other packages; ASReml software (specific for breeders)
- Heritability is integral to assessing genetic gains (module 3)

# Review of module 2

1. Forward selection results in a seedling seed orchard. Backwards selection usually entails a separate progeny test and grafted orchard.
2. Progeny test are designed as single tree (preferred) or multi-tree plots in randomized complete or incomplete blocks.
3. Breeders are most interested in additive genetic effects; we mostly ignore non-additive effects unless we have full sibs or clone trials.
4. Narrow-sense heritability is the percent of phenotypic variation attributable to additive effects in open-pollinated progeny tests.
5. Heritability can be calculated in different ways to help breeders devise a roguing scheme (family vs within family selection).
6. Heritability represents a snapshot of a trait/site since estimates can vary over space and time.

Please join us for module 3  
next Friday, April 19

### References

**Falconer, D.S. and Mackay, T.F.C. 1996.**  
Introduction to Quantitative Genetics. 4th  
Edition, Longman Group Ltd.

**Isik F., Holland, Maltecca. 2017.** Genetic  
Data Analysis for Plant and Animal  
Breeding, Springer International  
Publishing.

**White, T.L., Adams, W.T., Neale, D.T. 2007.**  
Forest Genetics. CAB Publishing.

**Zobel, B. and J. Talbert. 1984.** Applied  
Forest Tree Improvement, Blackburn Press,  
NJ.

## Breeding designs and genetic gains in tree improvement programs

Thank you!!

[Carolyn.c.pike@usda.gov](mailto:Carolyn.c.pike@usda.gov)

Kevin.potter@usda.gov

Any additional questions?