Tree Improvement 101:
Breeding designs and genetic gain in tree improvement programs
MODULE 3

Carolyn (Carrie) Pike, PhD

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

Trevor Walker, PhD

Co-Director and Assistant Professor

NC State University, Cooperative Tree Improvement Program

I ♥ Baby Trees
RNGR.net

RNGR REFORESTATION, NURSERIES & GENETIC RESOURCES

https://rngr.net/

Reviewing module 2

- ✓We looked at forward vs backwards selection.
- ✓We looked at progeny test designs: single tree and multi-tree row plots in randomized complete or incomplete blocks.
- ✓We looked at additive vs non-additive genetic effects.
- ✓We compared narrow-sense (additive/phenotypic variation) vs broad sense (genetic/phenotypic variation) heritability equations, focusing on open-pollinated progeny tests.
- ✓We looked at various ways to calculate and interpret heritability which describes the similarity of siblings within families

“Gains” from tree improvement are realized in the SEED collected from a seed orchard.
Seed becomes the currency of a tree improvement program

Three main types of selection (module 2)

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 where individual phenotypes are not a good reflection of genotypes.

Make sure the final orchard has sufficient genetic diversity to meet your objectives!

Family plus within family selection or family “index” selection

Two-stage method involving selection on families, followed by selection within families.

Tree improvement programs often focus on quantitative traits

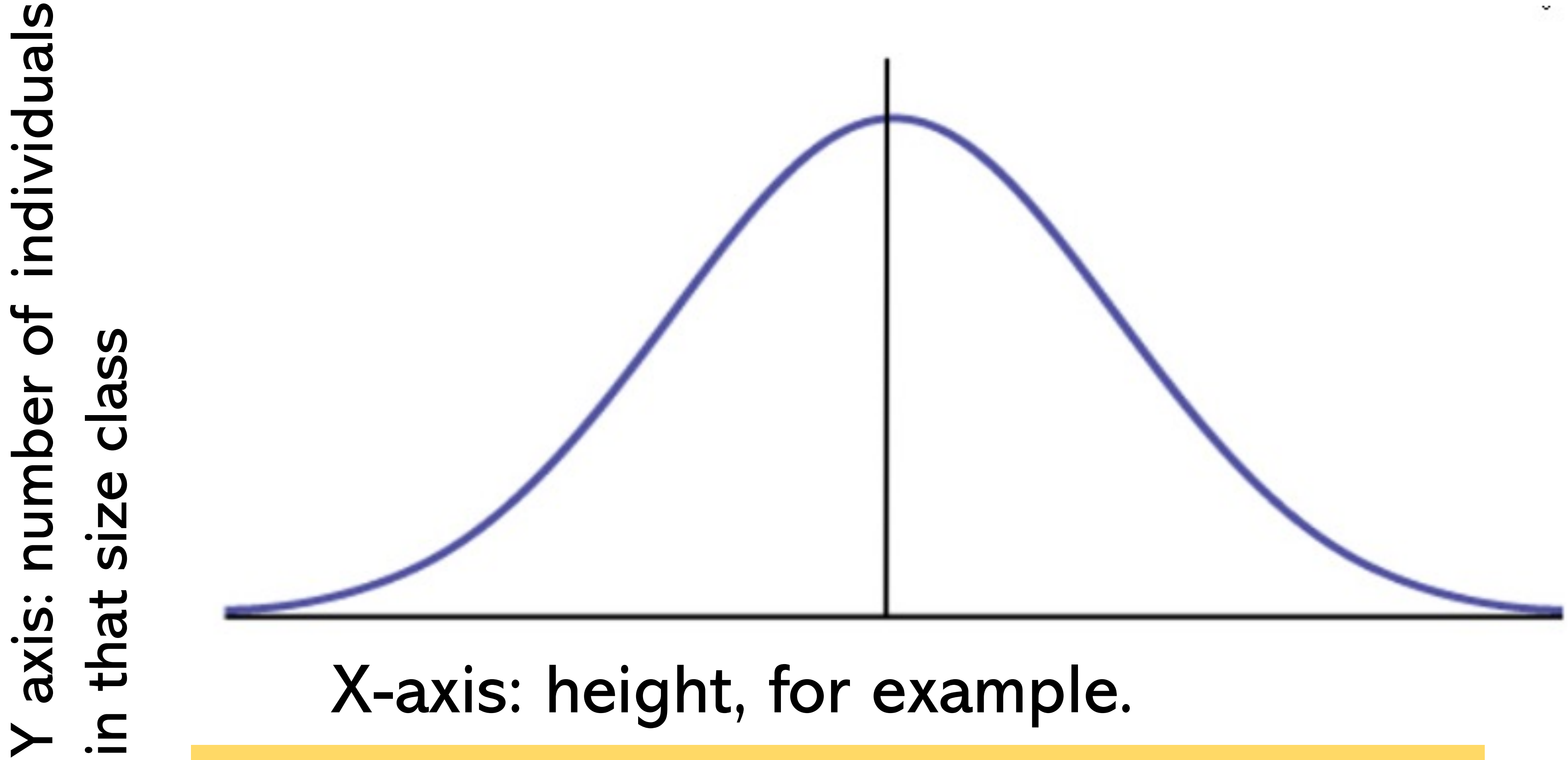
Quantitative traits are those with a continuous distribution (module 1). You measure these traits with a ruler, hypsometer, a scale, etc.

A sample of traits used in tree improvement programs (from module 1)

1. Tree height, diameter
2. Volume
3. Stem straightness (qualitative scale or assess stem sweep quantitatively)
4. Branch angle (measure or develop a categorical scale)
5. Wood density/specific gravity
6. Resistance to insect or disease
7. Many more!

Improvement is incremental from one generation to the next. Some traits can be improved quickly while others might require a few generations to reach the desirable phenotype.

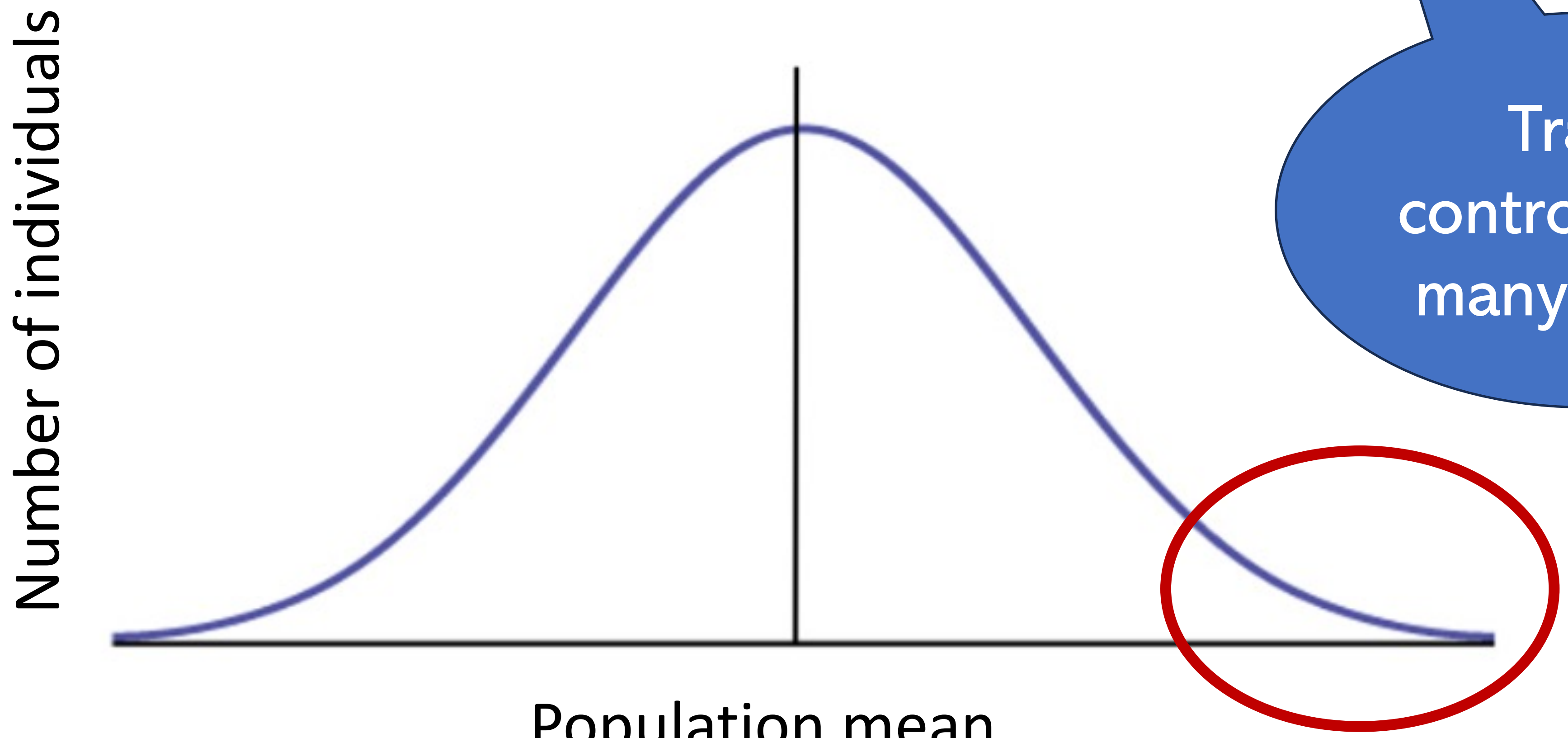
Module 1: Quantitative traits vary on a continuum. This is a histogram:



Quantitative traits are often normally distributed if we have a large enough sample size.

Quantitative traits are often polygenic. We select from one tail of the distribution.

Traits controlled by many genes

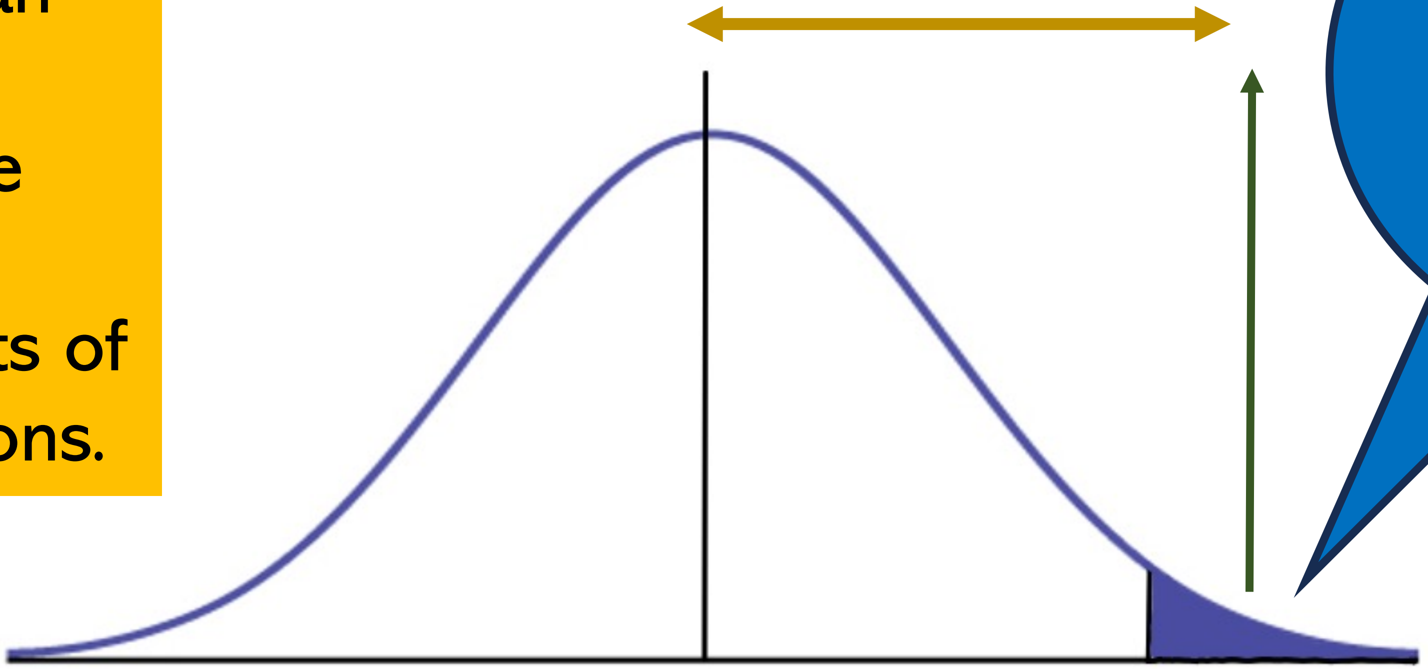


Trait increasing in value →

Figure credit: Jake Warren

Select from the superior tail of the distribution.

Selection differential is the difference between the mean of the entire population vs the sub population, expressed in units of standard deviations.



Select families from this tail and create a new improved first-generation population

Population mean

Mean of selected sub population

Genetic gain Δg : percent change in the mean of the selected population relative to the entire population (or a designated checklot)

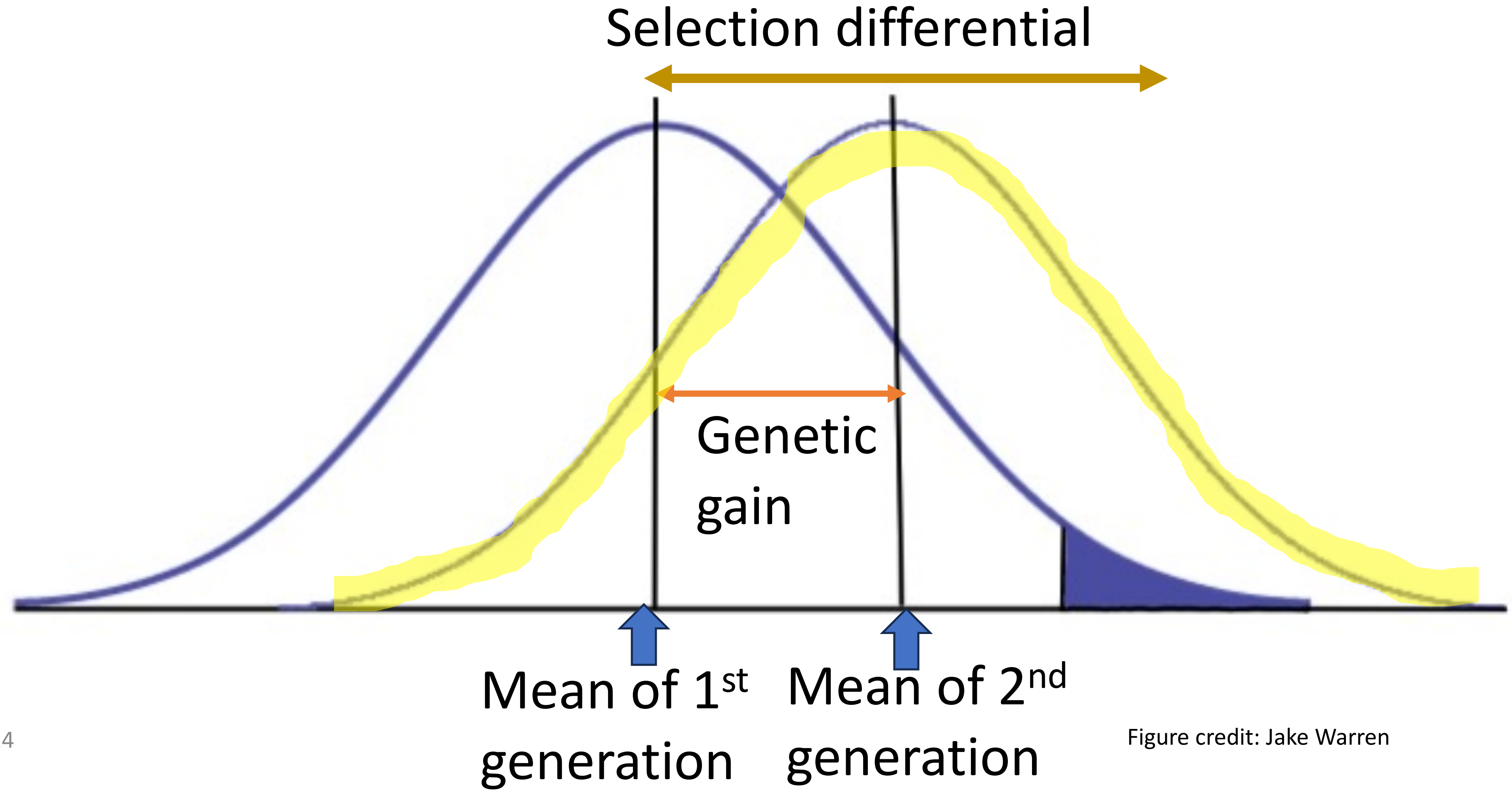


Figure credit: Jake Warren

Genetic gain: breeders’ equation

$$\Delta g = i * h^2 * V_p$$

Δg = gain

Amount of improvement, in the units measured.

i = intensity of selection

How intensively did you select? Units: the number of standard deviations

h^2 = narrow sense heritability

Additive / Phenotypic variation

V_p= phenotypic variation

How much variation is present?

This same equation is used to evaluate shifts caused by natural selection!

Genetic diversity vs genetic gain

Tree improvement programs manage the amount of genetic diversity in their breeding populations.

The manager of the seed orchard determines the amount of gain vs the amount of genetic diversity in their seed orchards!!

Genotype by environment interactions

Performance of a genotype at one site or seed zone may not be consistent across locations!

Genotype by environment interaction

This occurs when the performance of a genotype varies at different sites.

A shift in the ranks among different test sites will trigger a significant g x e interaction:

$$\mathit{Trait} = \mu + \mathit{Site} + \mathit{Replication} + \mathit{Family} + \underline{\mathit{Site}*\mathit{Family}} + \mathit{error}$$

Implication of significant g x e effects? You need separate populations for different sites, or seed zones. Significant g x e interactions complicate tree improvement programs!

Genotype x environment interactions for different traits of *Quercus rubra*: 3 progeny test sites in Indiana at 10 years of age

Type	Trait	Site	Fam	Site*Family
Quantitative	Volume	<.0001	<.0001	0.009
Quantitative	Height	<.0001	<.0001	0.0107
Quantitative	DBH	<.0001	<.0001	0.0211
Binary	Branch angle	0.9994	<.0001	<.0001
Binary	Sweep	0.9998	<.0001	0.9348
Binary	Branchiness	0.9987	<.0001	0.1573

Genotype x environment interactions for different traits of *Quercus alba*: 3 progeny test sites at 10 and 20 years of age.

Tree heights at year 10

Effect	Num DF	Den DF	F Value	Pr > F
Site	2	2570	1113.32	<.0001
Site*Stand	31	2570	2.63	<.0001
family(Stand)	53	2570	4.02	<.0001
Site*family(Stand)	87	2570	3.07	<.0001

Tree height year 22

Effect	Num DF	Den DF	F Value	Pr > F
Site	2	2196	605.77	<.0001
Stand	16	2196	2.42	0.0013
Family(Stand)	53	2196	1.54	0.008
Site*family(Stand)	108	2196	1.14	0.1575

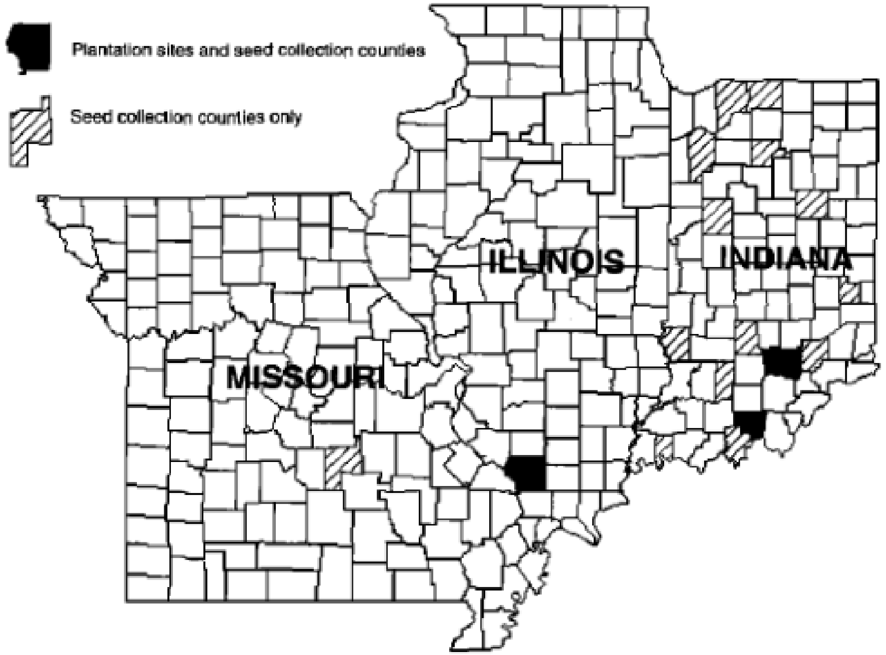


Figure 1. Location of white oak seed collections and outplantin sites.

Rink and Coggeshall 1995

Implications of g x e interactions for a breeding program?

- Different genotypes should be deployed to different geographic regions or seed collection zones.
- In our white oak example, we have g x e over a small geographic area.
  - We should refine our analysis to pull out more within and among site variation.
  - Or we need separate populations

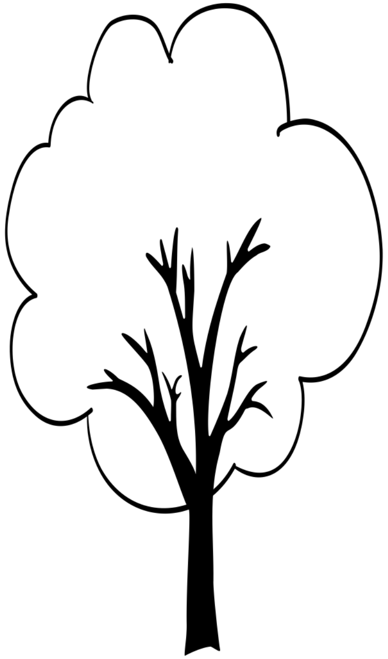
G x E interactions are typically minimal for growth traits in temperate conifers.

Breeding values

This is how we rank our families in a progeny test.

In module 2 we constructed heritability from the variance components. Now, we estimate breeding values (BV) for each family or individual (for within family selection)

> *"The value of an individual, judged by the mean value of its progeny, is called the breeding value of the individual."*
>
> -Falconer and Mackay 1996.



A breeding value is similar to a least-squared mean, but is adjusted for environmental effects (reps, sites).

It reflects how closely the phenotype represents the genotype.

Steps in calculating genetic gains 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.
- Run ANOVA in linear mixed models framework
- Set family as a random effect
- Derive best linear unbiased predictors for the family effect.

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 (module 2)

Trait = μ + *Block**+ *Family** + *error*

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

μ 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.

We generally don't rely on pairwise comparison tests (such as Tukey's) to rank families because family is not a fixed effect!!

General Combining Ability (GCA)

"The portion of the genetic value that is transmitted to the offspring created by mating the individual randomly to all other individuals in the population."

\- White et al. 2007.

- GCA is the average performance of the progeny of a parent (usually mother)
- Statistically GCA is represented by the BLUP, (best linear unbiased predictor).
- The GCA is half the breeding value.
- We look for families that are 'best general combiners,' meaning that the performance of their progeny is consistent across pollen sources.

Breeding values: half-sib (OP)

Is family 3 significantly different than the others?

We don't really care.

We'll look at the range of breeding values to decide which families to select.

Female parent	Progeny means
1	13
2	12
3	14
4	12
Progeny means	**Test mean =** 13 meters

Different parameters for family 2 (subscript 2):

BLUP$_2$ = mean of parent 2 – test mean = 12-13 = -1

Breeding value = BLUP * 2 = -1 * 2 = -2

Adjusted half-sib breeding value = -2+13 = 11 meters

Percent improvement = (11-13)/13 = - 0.15, or -15%

BLUP = the GCA

Breeding value = 2*BLUP

Genetic gains for family selection: sample dataset for white oak.

Trait = Tree heights at year 0

This is a sample output file.

One row per family.

You can change the model to have one value per tree (animal model)

General Combining Ability (GCA)

Half-sib breeding value (BV): GCA * 2

Adjusted half-sib breeding value

Relative to grand mean

fam	BLUP	HS_BV	Grand Mean	Adj_HF_BV	Gain
IL-01-01	-0.251	-0.503	26.6	26.14	-1.9%
IL-01-02	1.075	2.15	26.6	28.79	8.1%
IL-01-03	7.9	15.8	26.6	42.44	59.3%
IL-01-05	6.808	13.616	26.6	40.26	51.1%
IL-01-06	6.184	12.368	26.6	39.01	46.4%
IN-01-01	-3.371	-6.743	26.6	19.9	-25.3%
IN-01-02	-3.449	-6.899	26.6	19.74	-25.9%
IN-01-03	0.412	0.824	26.6	27.47	3.1%
IN-01-04	0.802	1.604	26.6	28.25	6.0%
IN-01-05	-0.446	-0.893	26.6	25.75	-3.4%
IN-02-01	-5.634	-11.267	26.6	15.38	-42.3%
IN-02-02	0.373	0.746	26.6	27.39	2.8%
IN-02-03	2.323	4.646	26.6	31.29	17.4%

Genetic gains for family selection: sample dataset for white oak.

Trait = Tree heights at year 0

Genetic gains for the seed orchard are represented by the average gains of the genotypes.

General Combining Ability (GCA)

Half-sib breeding value (BV): GCA * 2

Adjusted half-sib Breeding value

Relative to grand mean

fam	BLUP	HS_BV	Grand Mean	Adj_HF_BV	Gain
IL-01-01	-0.251	-0.503	26.6	26.14	-1.9%
IL-01-02	1.075	2.15	26.6	28.79	8.1%
IL-01-03	7.9	15.8	26.6	42.44	59.3%
IL-01-05	6.808	13.616	26.6	40.26	51.1%
IL-01-06	6.184	12.368	26.6	39.01	46.4%
IN-01-01	-3.371	-6.743	26.6	19.9	-25.3%
IN-01-02	-3.449	-6.899	26.6	19.74	-25.9%
IN-01-03	0.412	0.824	26.6	27.47	3.1%
IN-01-04	0.802	1.604	26.6	28.25	6.0%
IN-01-05	-0.446	-0.893	26.6	25.75	-3.4%
IN-02-01	-5.634	-11.267	26.6	15.38	-42.3%
IN-02-02	0.373	0.746	26.6	27.39	2.8%
IN-02-03	2.323	4.646	26.6	31.29	17.4%

These three families have the highest genetic gains

Genetic gains for family selection: sample dataset for white oak.

Trait = Tree heights at year 0

General Combining Ability (GCA)

Half-sib breeding value (BV): GCA * 2

Adjusted half-sib Breeding value

Relative to grand mean

fam	BLUP	HS_BV	Grand Mean	Adj_HF_BV	Gain
IL-01-01	-0.251	-0.503	26.6	26.14	-1.9%
IL-01-02	1.075	2.15	26.6	28.79	8.1%
IL-01-03	7.9	15.8	26.6	42.44	59.3%
IL-01-05	6.808	13.616	26.6	40.26	51.1%
IL-01-06	6.184	12.368	26.6	39.01	46.4%
IN-01-01	-3.371	-6.743	26.6	19.9	-25.3%
IN-01-02	-3.449	-6.899	26.6	19.74	-25.9%
IN-01-03	0.412	0.824	26.6	27.47	3.1%
IN-01-04	0.802	1.604	26.6	28.25	6.0%
IN-01-05	-0.446	-0.893	26.6	25.75	-3.4%
IN-02-01	-5.634	-11.267	26.6	15.38	-42.3%
IN-02-02	0.373	0.746	26.6	27.39	2.8%
IN-02-03	2.323	4.646	26.6	31.29	17.4%

These three families have the lowest genetic gains

Genetic gains for the seed orchard are represented by the average gains of the genotypes.

If you have full-sib crosses then you can calculate Specific Combining Ability (SCA), which can inform which parent x parent crosses had the best offspring. This allows you to capitalize on non-additive sources of genetic variation.

SCA is worth pursuing when the genetic gains are high enough to warrant the extra effort required to produce the crosses.

General (GCA) vs Specific Combining Ability (SCA)

- GCA is useful if you plan to collect **open-pollinated seed** from your seed orchard for planting projects.

- SCA is useful if you plan to collect **seed from specific crosses** in your seed orchards, then SCA is needed to figure out which combination works best.

Mass controlled pollination of loblolly pine

Selection for more than one trait

Be careful not to select too many traits with each generation!

Considerations for selecting multiple traits

Breeding to improve one quantitative trait in forest trees requires a lot of individuals and families!

Tandem selection.

1st generation: improve for one trait

2nd generation: improve for the second trait.

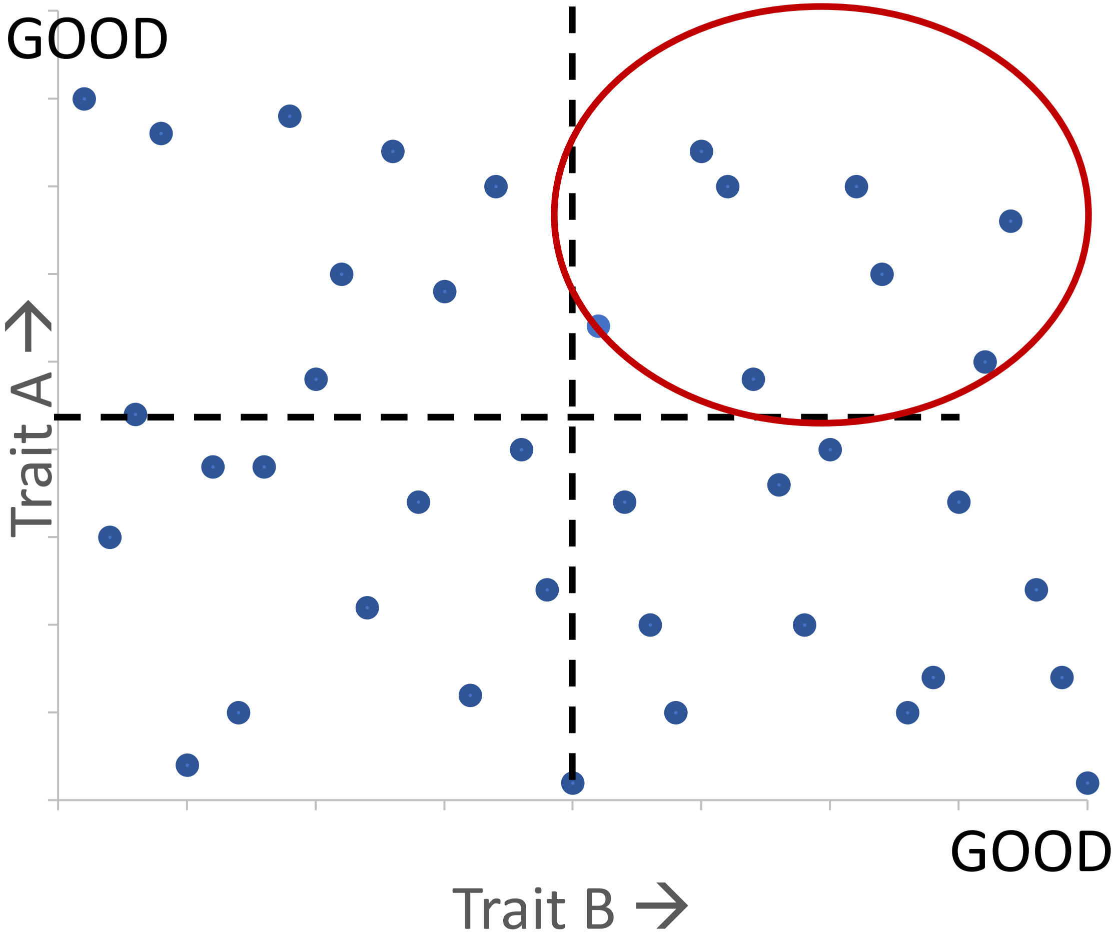
1st generation: Improve for trait 1

2nd generation: improve for trait 2

Considerations for selecting multiple traits

Independent culling. Set minimum values for each trait of interest.

- If the traits are positively correlated then this method can work.
- If the traits are uncorrelated, and you have enough families, then this method also works.

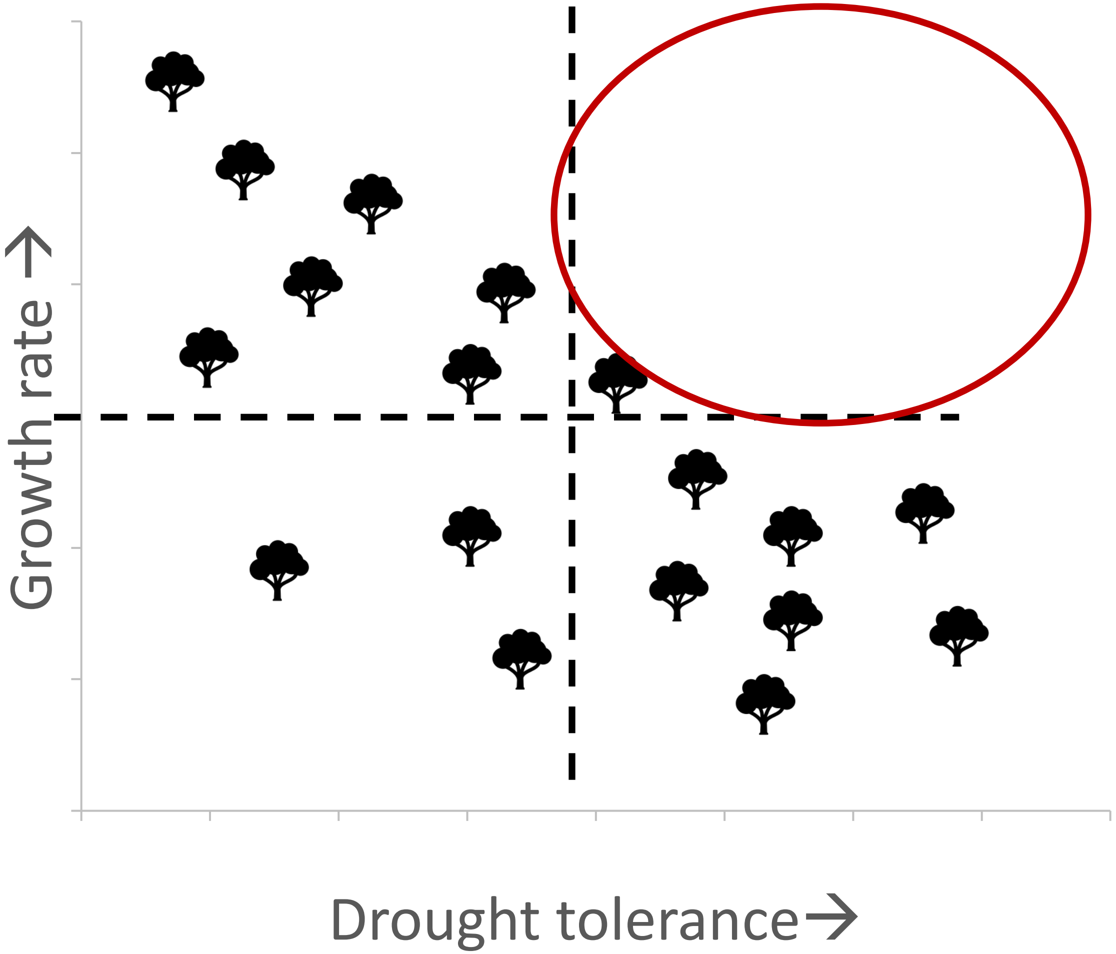


Considerations for selecting multiple traits

Independent culling. Set minimum values for each trait of interest.

BEWARE:

- If traits are negatively correlated, you may never develop a population that is strong for both traits!
- Example: fast growth and drought tolerance may tradeoff.



(not real data)

Tree breeding: many different mating designs are available. Choose one that meets your objectives and your budget!

Why we breed: the rationale for controlled pollinations

- We can improve the genetics of our breeding population by crossing selected individuals or families to “stack” good genes.

- This increases genetic gains, but also lowers the effective population size (the total number of unrelated genotypes in a breeding program) with each generation!

Mating designs:
1. Factorial crosses

These are ‘dioecious friendly’ designs. Each genotype is tested as a male or female depending on its flower type.

Ash (*Fraxinus* spp.) are all dioecious. Most conifers are monoecious.

Single Pair Matings
“Positive Assortative”

Female (rank)

Male (rank)

	1	3	5	7	9	11
2	X					
4		X				
6			X			
8				X		
10					X	
12						X

Nested or hierarchical matings

Female

Male

	1	2	3
4	X		
5	X		
6	X		
7		X	
8		X	
9		X	
10			X
11			X
12			X

Disconnected factorial

		Female					
		1	2	3	4	5	6
Male	7	X	X	X			
	8	X	X	X			
	9	X	X	X			
	10				X	X	X
	11				X	X	X
	12				X	X	X

Tester mating design

		Female			
		1	2	3	4
Male	5	X	X	X	X
	6	X	X	X	X
	7	X	X	X	X
	8	X	X	X	X
	9	X	X	X	X
	10	X	X	X	X
	11	X	X	X	X
	12	X	X	X	X

2. Diallels – not for the dioecious! Monoecious tree species only!

Each genotype is used as both a male and a female

Disconnected half-diallels

Female

Male

	1	2	3	4	5	6	7	8	9	10	11	12
1		X	X	X	X	X						
2			X	X	X	X						
3				X	X	X						
4					X	X						
5						X						
6							X	X	X	X	X	X
7								X	X	X	X	X
8									X	X	X	X
9										X	X	X
10											X	X
11												X

Modified half-diallel

Female

Male

	1	2	3	4	5	6	7	8	9	10	11	12
1		X	X	X	X	X	X	X	X	X	X	X
2			X	X	X	X	X	X	X	X	X	X
3				X	X	X	X	X	X	X	X	X
4					X	X	X	X	X	X	X	X
5						X	X	X	X	X	X	X
6							X	X	X	X	X	X
7								X	X	X	X	X
8									X	X	X	X
9										X	X	X
10											X	X
11												X

Full diallel:
includes all self crosses and reciprocal crosses (1x2 and 2x1)

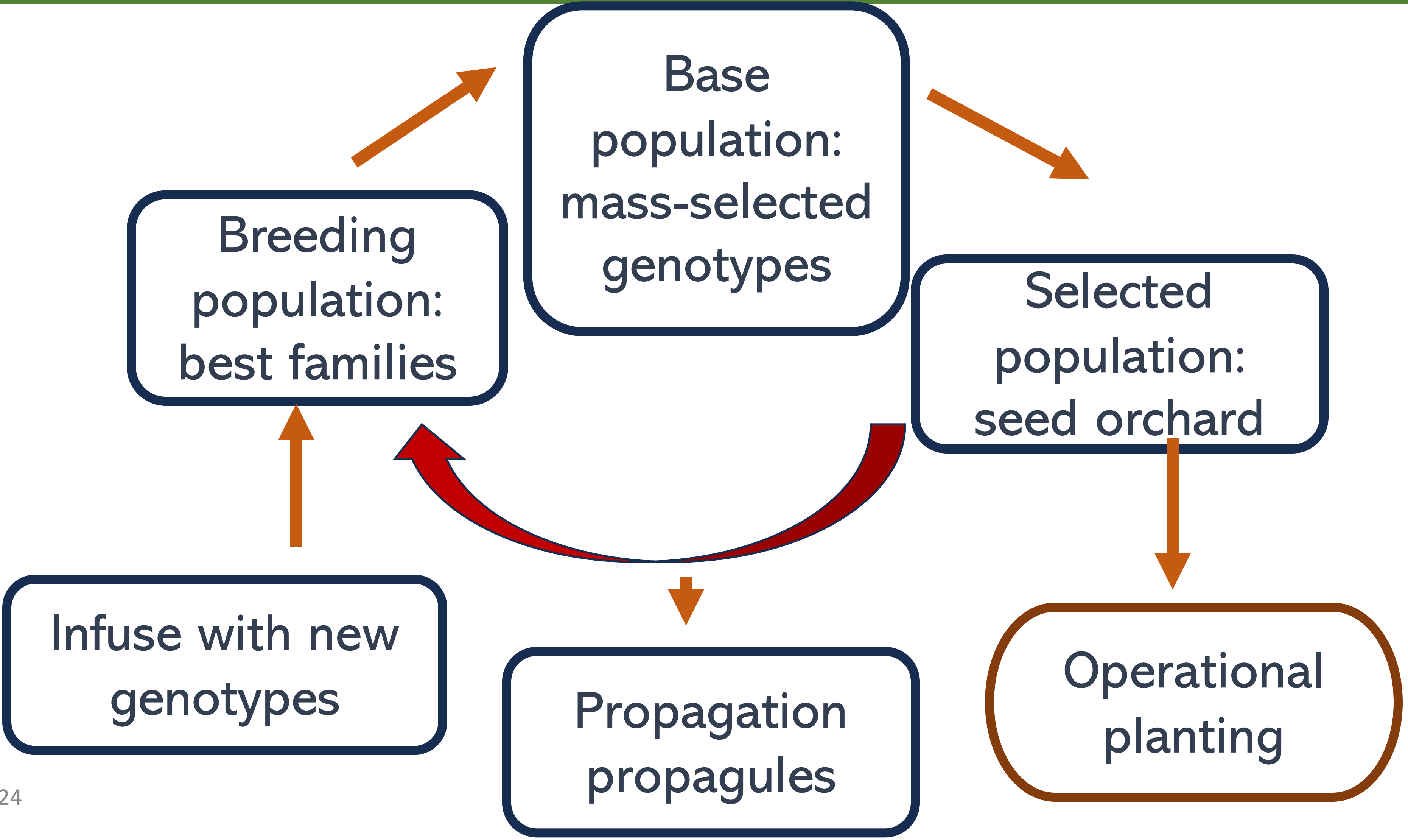
Male

Female

	1	2	3	4	5	6	7	8	9	10	11	12
1	X	X	X	X	X	X	X	X	X	X	X	X
2	X	X	X	X	X	X	X	X	X	X	X	X
3	X	X	X	X	X	X	X	X	X	X	X	X
4	X	X	X	X	X	X	X	X	X	X	X	X
5	X	X	X	X	X	X	X	X	X	X	X	X
6	X	X	X	X	X	X	X	X	X	X	X	X
7	X	X	X	X	X	X	X	X	X	X	X	X
8	X	X	X	X	X	X	X	X	X	X	X	X
9	X	X	X	X	X	X	X	X	X	X	X	X
10	X	X	X	X	X	X	X	X	X	X	X	X
11	X	X	X	X	X	X	X	X	X	X	X	X
12	X	X	X	X	X	X	X	X	X	X	X	X

These are extremely labor intensive!

Tree improvement programs maintain both BASE populations and production population



Base population: mass-selected genotypes

Selected population: seed orchard

Operational planting

Propagation propagules

Breeding population: best families

Infuse with new genotypes

Summary from module 3

- When family ranks are inconsistent across sites then the g x e is significant and implies that we may need separate orchards for separate zones.
- Genetic gain, or the breeders' equation is the product of the narrow sense heritability, selection differential, and phenotypic diversity.
- Breeding values are an estimate of how closely the phenotype resembles the genotype.
- General combining ability (GCA) is the best linear unbiased predictor. This is easily calculated from open-pollinated progeny tests.
- Multiple traits can be incorporated with tandem selection or independent culling.
- Different breeding designs can be incorporated depending on your objectives. These include variations of factorials and diallel designs.
- Tree improvement programs often maintain separate base and breeding populations. Genotypes are often conserved in a program.

Tree improvement cooperatives require technical expertise to support seed orchards, progeny tests, data management and analysis of common garden data

- We are fortunate to have several tree improvement cooperatives in the US: North Carolina State University, University of Florida, Texas A&M, Oregon State University, University of Idaho, University of Minnesota, Michigan State U.
- Consider joining a cooperative if you are interested in developing improved seed orchards for a forest tree species.

Please join us next Friday for module 4 on resistance breeding!

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.

Thank you!!

Carolyn “Carrie” Pike
Carolyn.c.pike@usda.gov

Trevor walker
trevor_walker@ncsu.edu