

Quantitative Genetics

Helene Wagner, University of Toronto

Goal: Test for adaptation in Western white pine

- Estimate trait from common garden experiment
- Assess trait heritability
- Compare trait variation to neutral markers (SNPs)
- Correlate trait variation with environment

Methodological Challenges

1. Design of Western white pine study
2. Introducing SNP data
3. Specifying linear models
 - What type of linear model to fit?
 - Model formulas in R
4. Video 2:
 - Under the hood
 - Interpreting results

Trait: d13C
(related to water use efficiency)

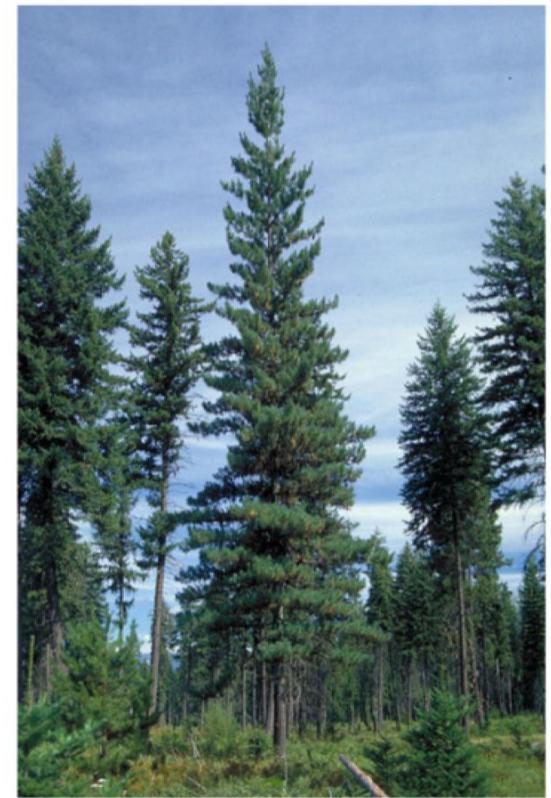
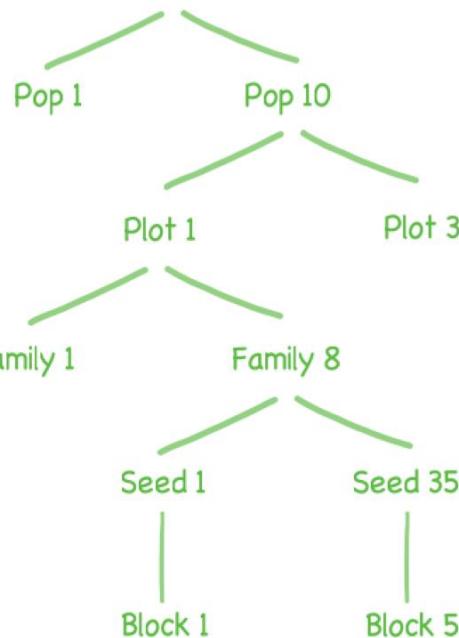


Image source: wikipedia.com

Study Design

10 Populations



3 Plots per Population

- Spatial coordinates
- 7 bioclimatic variables

2 - 8 Trees per Plot

- 164 SNP markers

5 - 35 Seeds per Family

5 Blocks in Common Garden

- $\delta^{13}\text{C}$ of each seedling

Seedling trait:

- 3 - 6 seedlings per family
- common garden
- trait: $\delta^{13}\text{C}$
- no molecular markers

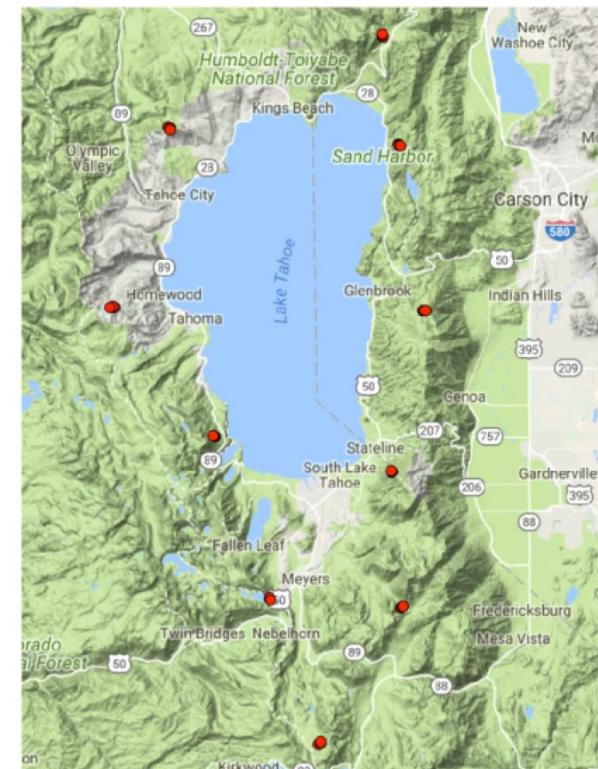


Estimate heritable trait:

- Trait = $G + E + G:E$
- no trait measurement for mother
- estimate from offspring trait
- account for hierarchical sampling
- account for blocks



Trait - Environment Association: Variation in heritable trait explained by Climate or Geography?



Genetic differentiation:

- Trait: calculate Q_{ST}
- SNPs: calculate F_{ST}
- Compare Q_{ST} to F_{ST}

Introducing SNP Data

family	population	snp102.Plmn	snp106.Plmn
<int>	<fctr>	<fctr>	<fctr>
59	blk cyn	CC	AC
60	blk cyn	AC	AA
61	blk cyn	CC	AA
63	blk cyn	CC	AA
64	blk cyn	AC	AA
65	blk cyn	AC	AA
67	blk cyn	AC	AA
69	blk cyn	CC	AA
72	blk cyn	AC	AA
73	blk cyn	AC	NA

1–10 of 157 rows... Previous 1 2 3 4 5 6 ... 16

SNP markers:

- co-dominant
- genotype or haplotype?
- wildtype vs. mutant?
- monomorphic?
- many markers!
- neutral or not?

Here:

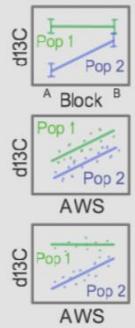
- 164 SNP markers
- monomorphic: 4
- 0 – 20% missing values
- unrelated to trait of interest

What Type of Linear Model to Fit?

1. Variable types?

		Response Y:	
		Quantitative	Binary
Predictors X:	Quantitative	Regression (dummy vars)	Logistic Regression
	Binary (single)	t - Test	Odds Ratio
	Categorical	ANOVA (covariates)	Log-linear Model

5. Interactions?



Two crossed factors: Pop, Block

- Effect of block (mean) depends on population (level of second factor)

Factor and covariate: Pop, AWS

- AWS has same effect for each pop (same slope, different intercepts)
 - AWS has different effect (slope, intercept) for different pops.

2. LM, LMM, or GLMM?

Method:	Conditions:	
	Normality	Balanced
LM	✓	✓
LMM	✓	✗
GLMM	✗	✗

4. Fixed or random factors?

Warning: Different philosophies!

	Replicate study	# Levels
Random	Different levels	Many (> 5?)
Fixed	Same levels	Often few

- Fixed effect: fits one parameter per level
 - Random effect: fits 2 parameters

Example

- Y: seedling d13C
 - Hierarchical sampling: Families = trees within plots within populations
 - Blocking: common garden with 5 blocks
 - Covariate (plot level): AWS
(soil available water supply)

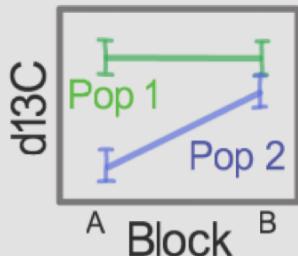
3. Nested or crossed?



Crossed factors: randomized experiment

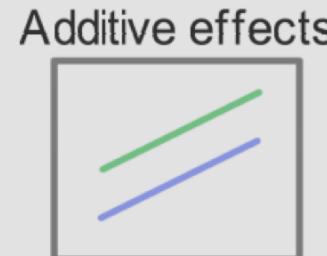
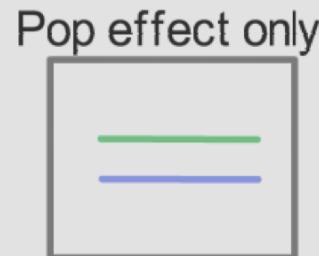
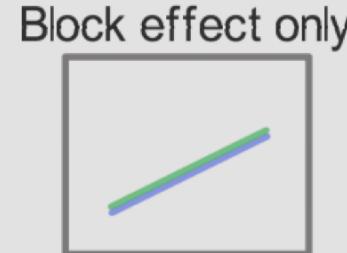
	Block 1	Block 2	Block 3
Family 1	1	1	1
Family 2	1	1	2
Family 3	1	1	0

5. Interactions?



Two crossed factors: Pop, Block

- Effect of block (mean) depends on population (level of second factor)



Specify Model Formula in R

Factors only:

- Global mean:
- Population mean:
- Additive effects:
- Interaction:
- Hierarchical model:

Fixed effects only:

$d13C \sim 1$
 $d13C \sim \text{Pop}$
 $d13C \sim \text{Pop} + \text{Block}$
 $d13C \sim \text{Pop} * \text{Block}$
 $d13C \sim \text{Pop/Family}$

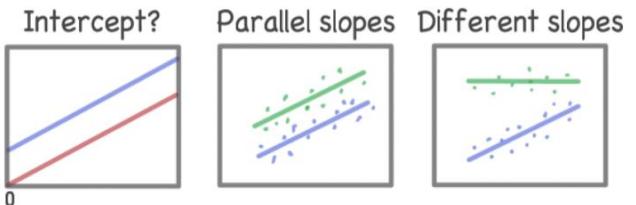
With random effects:

$d13C \sim (1 | \text{Pop})$
 $d13C \sim (1 | \text{Pop}) + (1 | \text{Block})$
 $d13C \sim (1 | \text{Pop/Family})$

With covariate (AWS, standardized):

- Simple regression:
- No intercept:
- Parallel slopes:
- Different slopes:

$d13C \sim \text{AWS}$
 $d13C \sim -1 + \text{AWS}$
 $d13C \sim \text{AWS} + \text{Block}$
 $d13C \sim \text{AWS} * \text{Block}$



$d13C \sim \text{AWS} + (1 | \text{Block})$
 $d13C \sim \text{AWS} + (\text{AWS} | \text{Block})$

Interaction

$$A * B = A + B + A : B$$

Nested

$$A / B = A + A : B$$

Intercept

$$A = 1 + A$$

Sum

$$I(A + B)$$

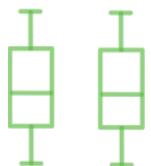
Quadratic

$$I(A^2)$$

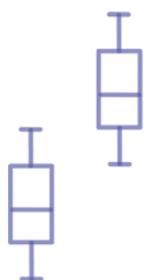
Degrees of Freedom

ANOVA F-test (LM)

F follows F-distribution with df(model), df(residual)



$$F = \frac{\text{Variability between groups}}{\text{Variability within groups}} = 1$$



$$F = \frac{\text{Variability between groups}}{\text{Variability within groups}} \gg 1$$

Assumptions

- Same variability in all groups
- Same group size (balanced)
- Residuals normally distributed
- Observations are independent

Degrees of freedom

$$\text{df(total)} = (n - 1)$$

- n = # independent observations
- one df used for global mean

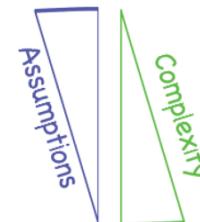
$$\text{df(model)} = (k - 1)$$

- k = number of parameters fitted

$$\text{df(residual)} = n - k$$

Alternative tests?

- Wald chi-square test
- Likelihood ratio test
- Conditional F-tests (balanced)
- Cond. F-tests with df correction
- MCMC or bootstrap tests



Non-independence?

Study design:

- Hierarchical sampling
- Blocking
- Paired samples
- Repeated measures

Autocorrelation:

- Spatial
- Temporal
- Phylogenetic

Co-ancestry:

- Kinship
- Population history
- Phylogeography

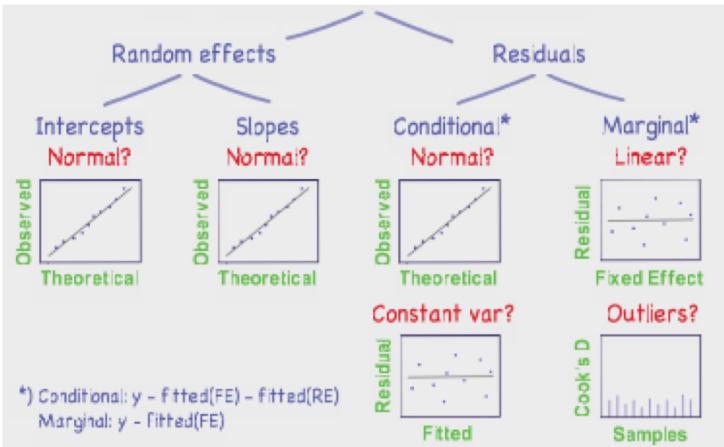
The algorithm matters!

	Model Comparison	Model Fitting
LM	LS	LS
LMM	ML	REML
	AIC	R-squared
	Fixed effects	Random effects

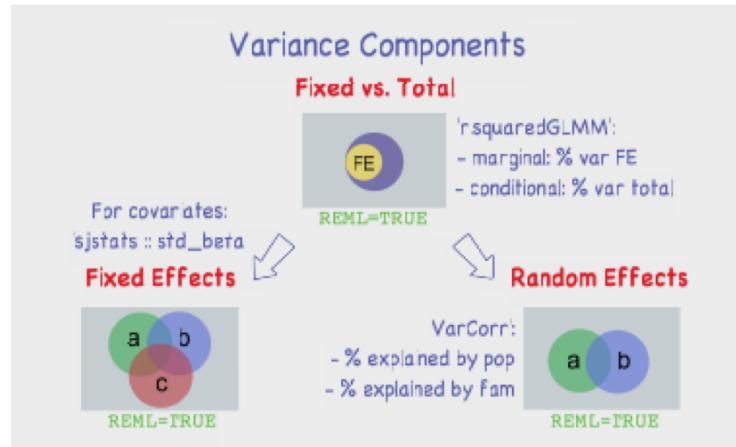
Interpreting LMM Results

Take-away: there's 2 (or more) of everything

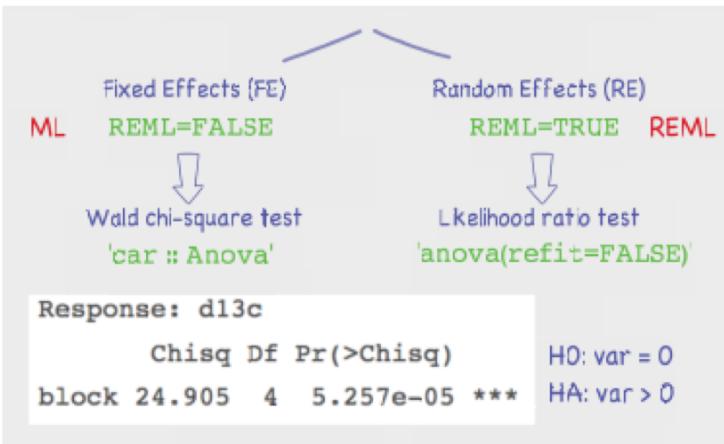
1. Residual Analysis



3. Size of Effects



2. Statistical Significance



4. Direction of Effects

Fitted = Intercept + Block + Population + Family

$$-30.37 = -30.63 + -0.39 + 0.56 + 0.09$$

'fixef' 'ranef'

Fixed effects:	
Estimate	(Intercept)
Block 1 (Intercept)	-30.62635
block2	-0.13833
block3	-0.35071
block4	-0.10060
block5	-0.39443
armstrong	-0.16372453
blk cyn	0.56424793
echo lk	-0.10480710
flume	0.37270688
hvn	-0.25055199

LMM Model Summary

Method?

Linear mixed model fit by REML ['lmerMod']

Formula: d13c ~ 1 + (1 | population) + (1 | family) + block Model?

Data: phen

REML criterion at convergence: 2050.4

Scaled residuals:

Min	1Q	Median	3Q	Max
-2.5436	-0.7485	0.0151	0.6028	3.7867

Random effects:

Groups	Name	Variance	Std.Dev.
family	(Intercept)	0.08164	0.2857
population	(Intercept)	0.10859	0.3295
Residual		0.71429	0.8452

Variance components?

Sample sizes?

Number of obs: 779, groups: family, 157; population, 10

Fixed effects:

	Estimate	Std. Error	t value
Effects? (Intercept)	-30.62635	0.12666	-241.79
block2	-0.13833	0.09667	-1.43
block3	-0.35071	0.09520	-3.68
block4	-0.10060	0.09538	-1.05
block5	-0.39443	0.09651	-4.09

Quantitative Genetics

Helene Wagner, University of Toronto

Goal: Test for adaptation in Western white pine

- Estimate trait from common garden experiment
- Assess trait heritability
- Compare trait variation to neutral markers (SNPs)
- Correlate trait variation with environment

Trait: d13C

(related to water use efficiency)



Methodological Challenges

1. Introducing SNP data
2. Specifying linear models
 - What type of linear model to fit?
 - Model formulas in R
3. Video 2: Why is model fitting so complicated?
 - Under the hood: degrees of freedom
 - Interpreting LMM results

Image source: wikipedia.com

Residual analysis (Week 4)

Spatial linear models (Week 7)

Model selection (Week 12)