

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)

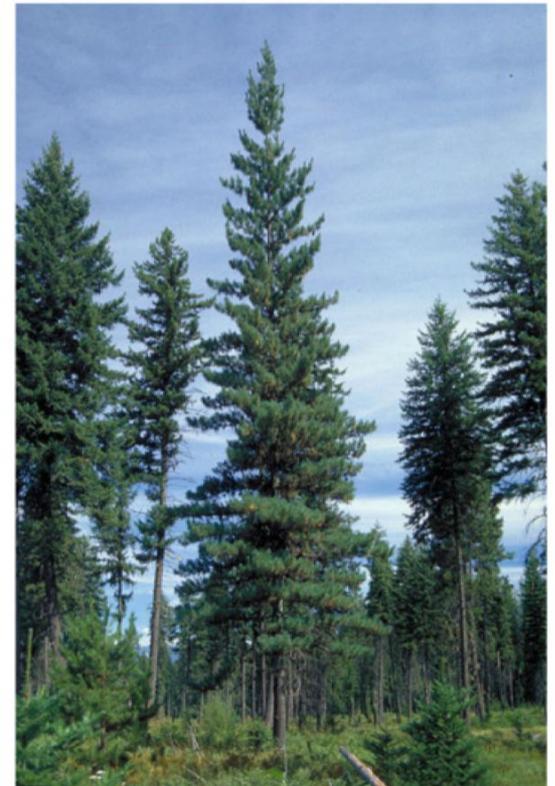


Image source: wikipedia.com

Study Design

10 Populations

3 Plots per Population

- Spatial coordinates
- 7 bioclimatic variables

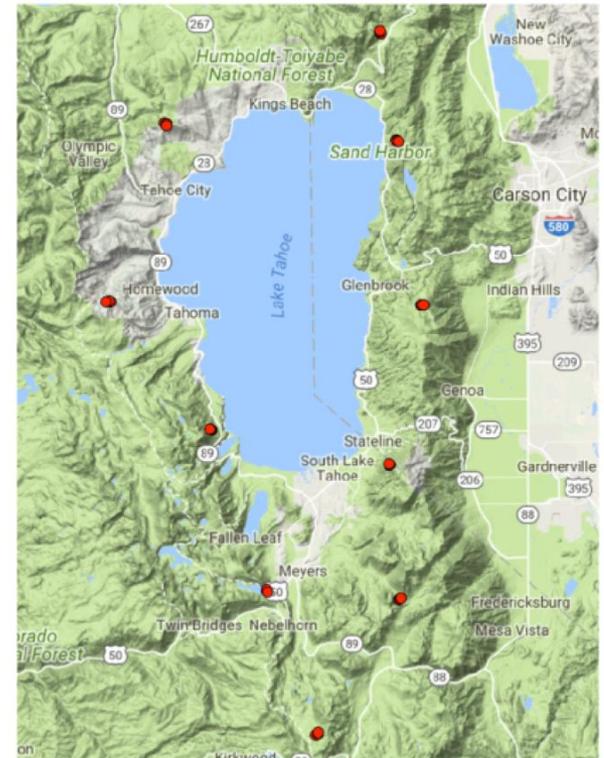
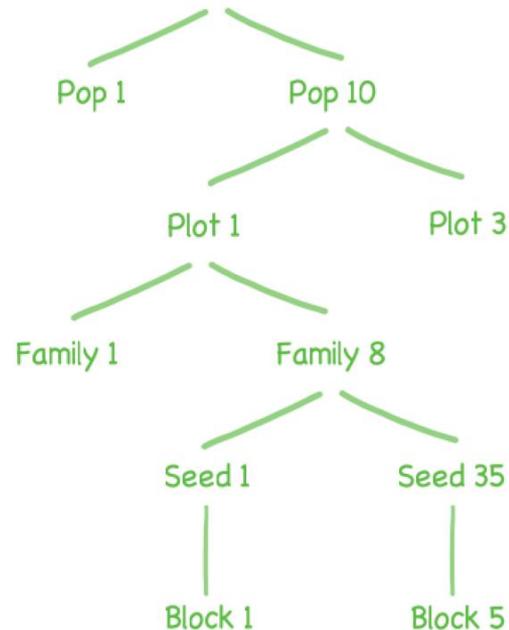
2 - 8 Trees per Plot

- 164 SNP markers

5 - 35 Seeds per Family

5 Blocks in Common Garden

- d13C of each seedling



Seedling trait:

- 3 - 6 seedlings per family
- common garden
- trait: d13C
- 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



Genetic differentiation:

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

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

Introducing SNP Data

family <int>	population <fctr>	snp102.Plmn <fctr>	snp106.Plmn <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

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

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

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

2. LM, LMM, or GLMM?

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

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)

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.

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

3. Nested or crossed?

Nested factors: hierarchical sampling

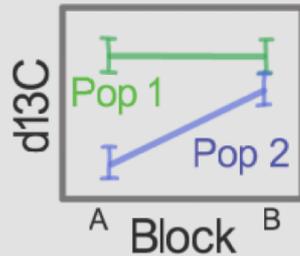


Crossed factors: randomized experiment

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

balanced unbalanced

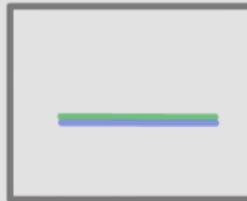
5. Interactions?



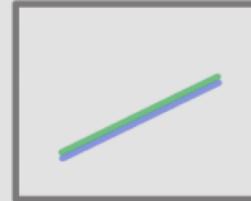
Two crossed factors: Pop, Block

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

No effect



Block effect only



Pop effect only



Additive effects



Specify Model Formula in R

Factors only:

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

Fixed effects only:

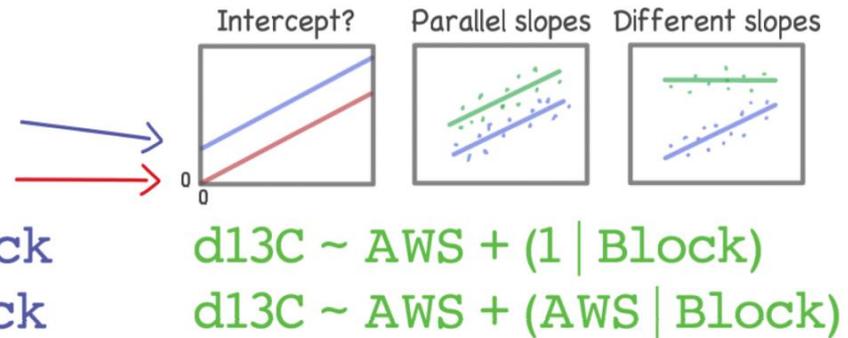
d13C ~ 1
 d13C ~ Pop
 d13C ~ Pop + Block
 d13C ~ Pop * Block
 d13C ~ Pop/Family

With random effects:

d13C ~ (1 | Pop)
 d13C ~ (1 | Pop) + (1 | Block)
 d13C ~ (1 | Pop/Family)

With covariate (AWS, standardized):

- Simple regression: d13C ~ AWS
- No intercept: d13C ~ -1 + AWS
- Parallel slopes: d13C ~ AWS + Block
- Different slopes: d13C ~ AWS * Block



Interaction

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

Nested

$A / B = A + A : B$

Intercept

$A = 1 + A$

Sum

$I(A + B)$

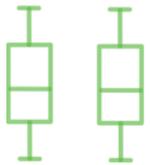
Quadratic

$I(A^2)$

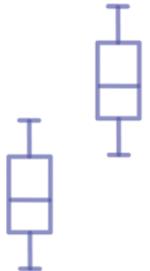
Degrees of Freedom

ANOVA F-test (LM)

F follows F-distribution with $df(\text{model})$, $df(\text{residual})$



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



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

Assumptions

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

Degrees of freedom

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

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

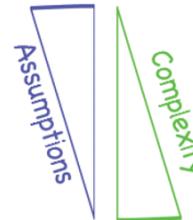
$$df(\text{model}) = (k - 1)$$

- k = number of parameters fitted

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

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

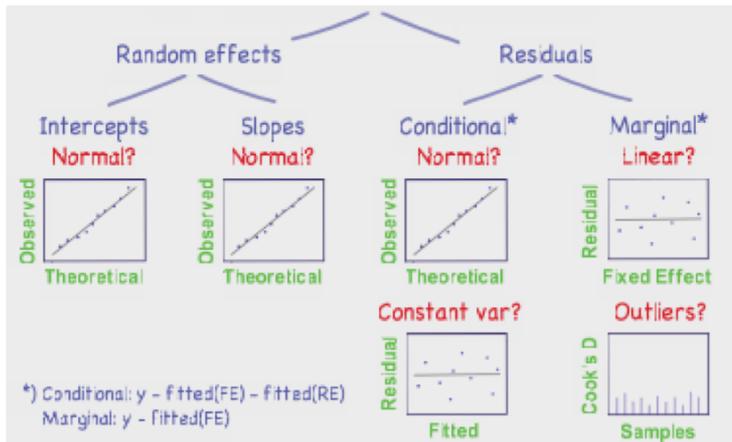
AIC R-squared

Fixed effects Random effects

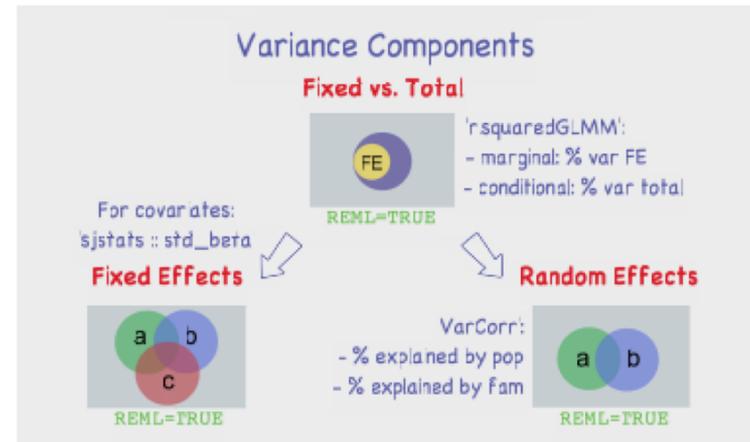
Interpreting LMM Results

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

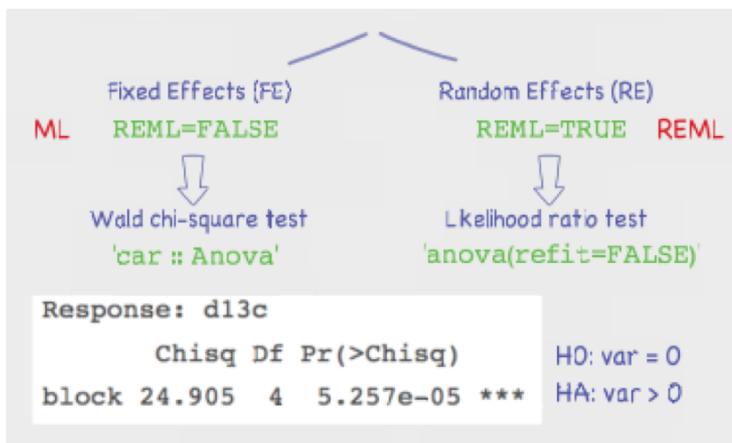
1. Residual Analysis



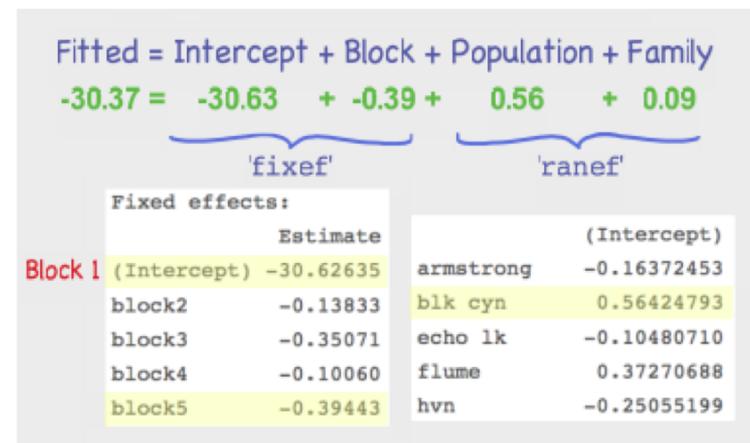
3. Size of Effects



2. Statistical Significance



4. Direction of Effects



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:

Effects?

	Estimate	Std. Error	t value
(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

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

Trait: $\delta^{13}C$

(related to water use efficiency)



Image source: wikipedia.com

Residual analysis (Week 4)

Spatial linear models (Week 7)

Model selection (Week 12)