Heritability of the Microbiome of a Proto-Social Beetle using NGS Data (Take Nothing for Granted)

Janet Sinsheimer

David Geffen School of Medicine at UCLA

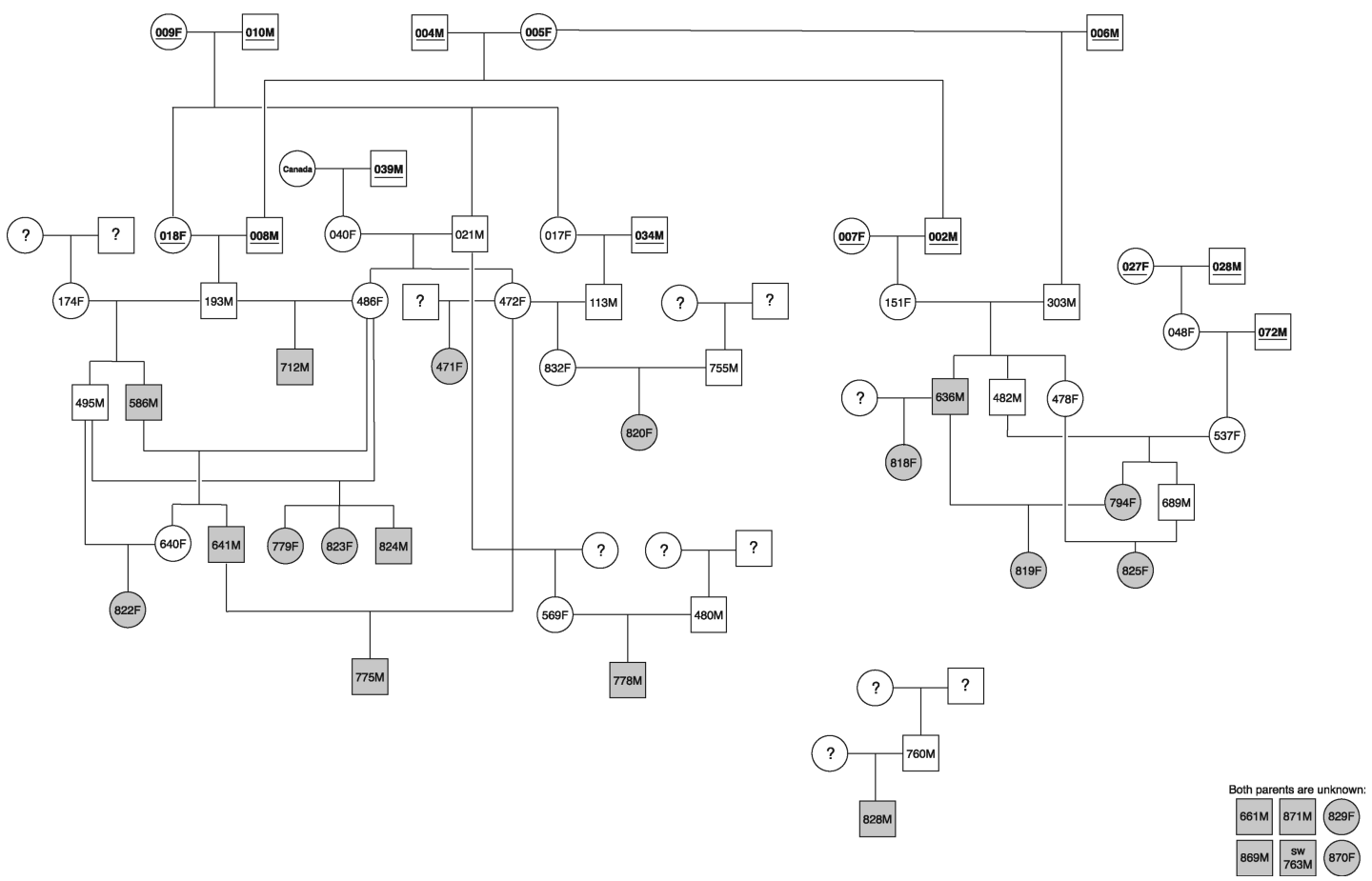
Maria Rivera and Alex Waldrop

Virginia Commonwealth University

The Request

- Calculate the heritability of several quantitative traits in a non-model organism

I imagined something like:



From: Genome Biol Evol. 2018;10(6):1546-1553. doi:10.1093/gbe/evy112

The Actual Goal

- Estimate the heritability of several quantitative traits related to nitrogen fixing microbes in a beetle that lives in family units
- No pedigree data but likely have near relatives (e.g. cousins)

The Organism = a Beetle

- Odontotaenius disjunctus (the Patent-leather, Bess (Betsy), horned passalus or Jerusalem beetle)
- O. disjunctus live in colonies in decaying oak and elm logs throughout forests in Eastern USA
- The young cannot feed themselves. There is delayed dispersal (a year to leave). Both male and female adults feed masticated wood and feces to the young
- When mature, beetles fly away to find mate and colonize a new log (or new part of a used log).
- Once adults, they live ~1 year more
- Microbes (identity unclear) are needed for wood digestion. May be needed to compensate for low nitrogen wood food source

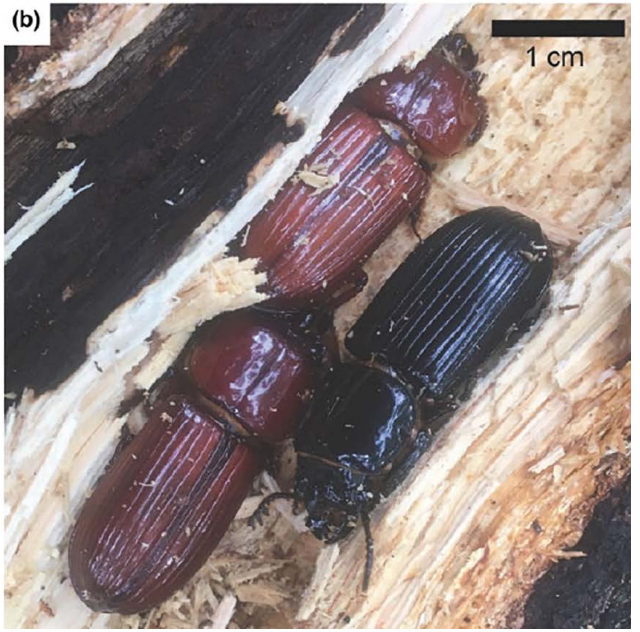
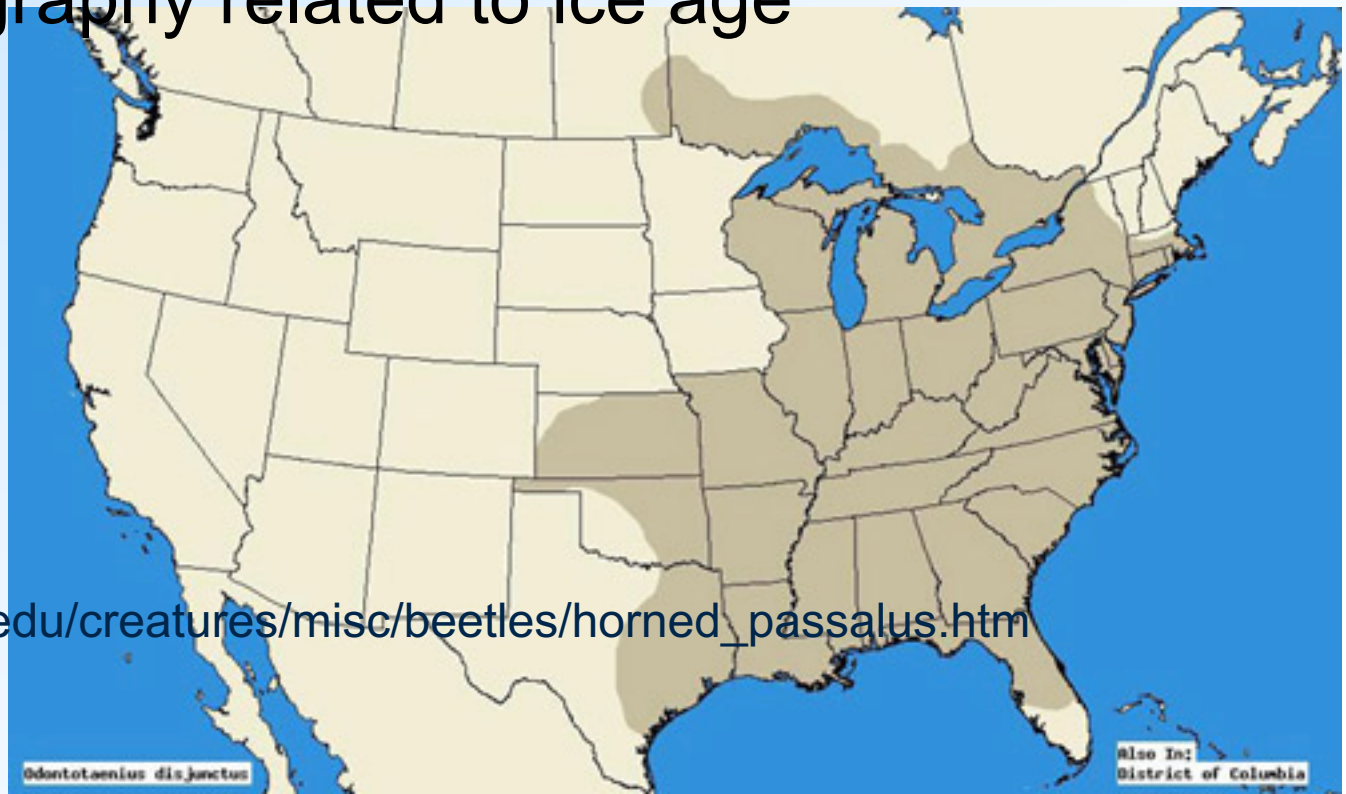


Figure 1b of Dillard and Maigret 2017 J Evol Bio

Why Care?

- Quite a bit of controversy about the heritability of the microbiome in general and for these beetles in particular
- Interesting population history and phylogeography related to ice age

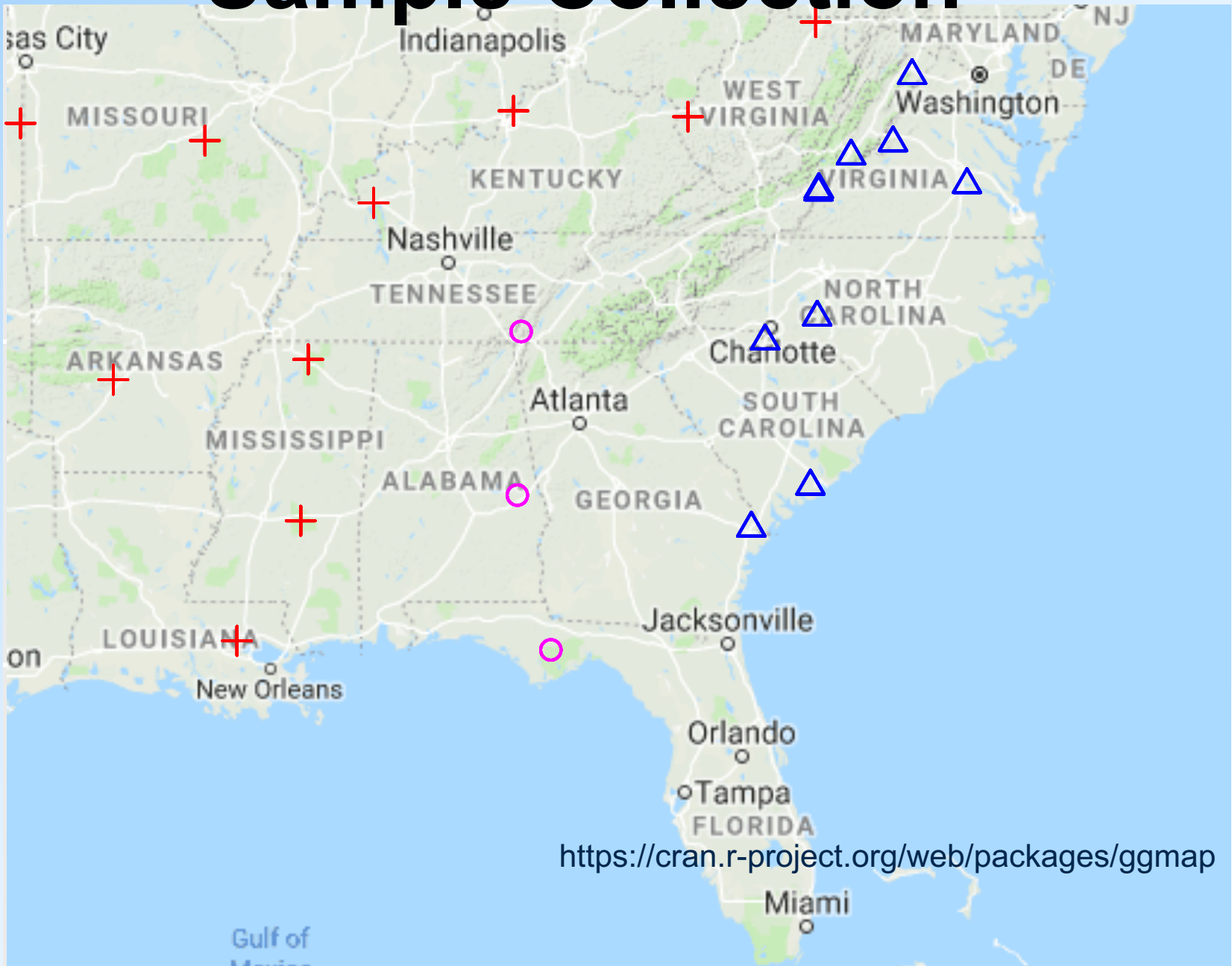


http://entnemdept.ufl.edu/creatures/misc/beetles/horned_passalus.htm

Study Design

- 23 forest locations in south-eastern US
- ~10 beetles per forest
- 230 beetles total
- Beetle data: sex, weight, date collected, location collected
- Climate and forest data collected (metadata)
- Microbiome from beetle gut segments
- DNA from muscle – whole genome sequencing

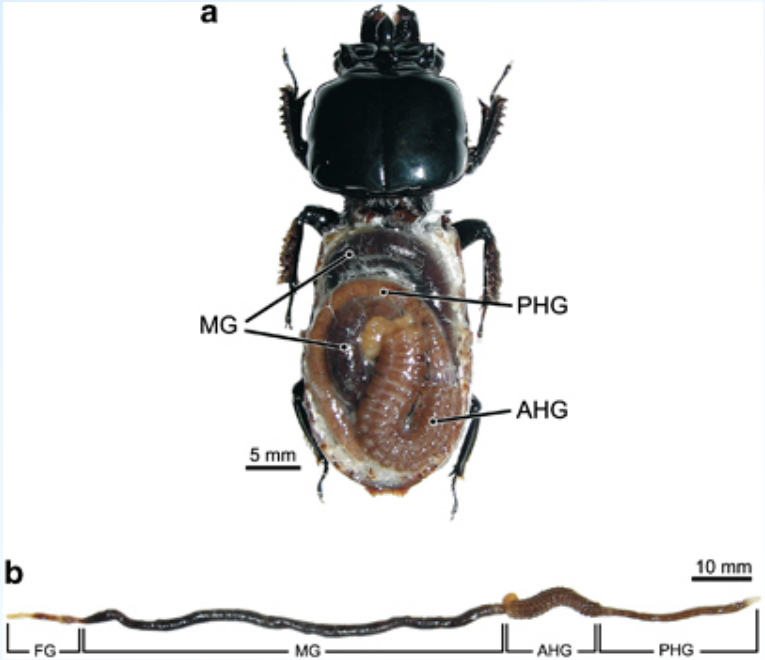
Sample Collection



https://cran.r-project.org/web/packages/ggmap

Outcomes

- Weight (as a reality check)
- Microbiome: abundance of OTUs generated from beetle fecal matter from the colon – three segments used: The midgut (MG), posterior hindgut (PHG), and the anterior hindgut (AHG)



Picture from Cela-Navarro et al. 2014 ISMEJ.

Study Design - Genotypes

- Whole genome sequencing on the cheap without a reference genome
- ddRAD seq
  - ddRAD seq = Double digest restriction-site associated DNA sequencing

dd RAD Sequencing:

- Cut genome with two restriction enzymes and adapters specific to each enzyme
- Select fragments by size
- Use next-generation sequencing to generate sequence data adjacent to a large number of restriction cut sites
- Targets only a fraction of the genome but at high coverage depth per locus
- Don’t have to have a reference genome

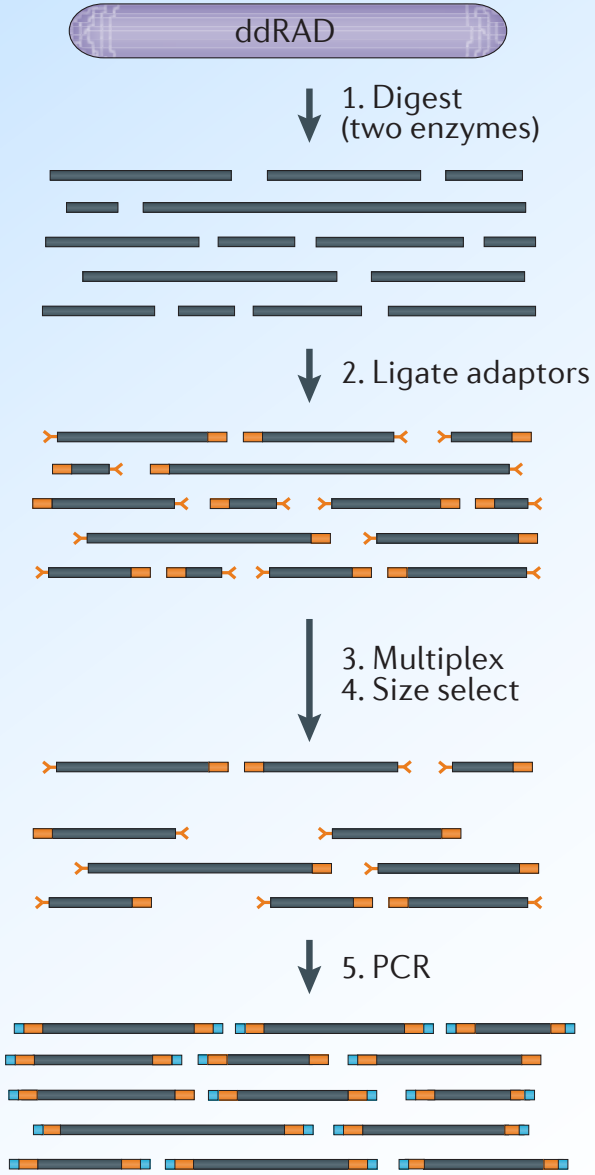


Diagram from Andrews et al 2016 Nat. Rev. Gen.

Inclusion Criteria

- Markers
  - QC protocols followed (Andrews et al. 2016 Nat. Rev. Genet.)
  - Minimum genotyping success >90% samples
  - At or near beetle gene coding regions
  - Correlation, r^2, less than 0.8 between Markers
  - MAF >0.05
  - Autosomal (allele frequencies not sex specific)
  - HWE holds within clade
  - ~2000 SNPs meeting these criteria

Inclusion Criteria

- Samples
  - Sex was determined (can be difficult)
  - Weight was determined
  - Microbiome DNA and host DNA extraction successful for all three segments
  - typed at >90% DNA Markers
  - 2 outliers in the overall microbiome abundance
  - 158 Beetles included

Microbiome Data

- 16s rRNA sequence used
- Clustered sequences by similarity
- >97% sequence identity to define OTUs
- Raw data: Number of reads of a particular OTU
- Which OTUs to use? Diversity measures?
- Transformations?

(see Goodrich et al. Cell 2014 for overview)

No Consensus on How to Analyze Microbiome Data

- Alpha Diversity – e.g. number of microbes, number of types of microbes
- Beta Diversity - measures of difference in composition between beetles
- Analysis on individual OTUs
  - Need to standardize for differences in abundance – use proportions
  - Transformations – e.g. CLR, inverse normal, others?

OTU Distribution

- 12163 OTUs found in at least two samples
- Most OTUs are found in only a few samples
  - 50% of OTUs found in only 7.5% or fewer samples
  - Only 97 OTUs found in >90% of samples

Choice of OTUs to Analyze

- For preliminary analyses chose:
  - Total abundance and number of types of microbes observed for each of the three segments
  - 10 most abundant OTUs overall that are found in all three segments
  - Also examined the 4 most abundant OTUs exclusively in one gut segment

Estimating Heritability

- SNP heritability: Used LMM with two variance components
- Reminder of the model:
  - Fixed effects: $Y_i = \mu + \beta^T X_i + A_i + e_i$
  - Random effects:

$Var(e_i) = \sigma_e^2, Cov(e_i, e_j) = 0$

$Var(A_i) = 2\Phi_{ii}^* \sigma_A^2, Cov(A_i, A_j) = 2\Phi_{ij}^* \sigma_A^2$

Φ^*_{ij} = kinship estimated using correlation of the 1952 SNPs:

$$\Phi_{ij}^* = \frac{1}{2S} \sum_{k=1}^{S} \frac{\left(X_{ik} - 2p_k\right)\left(X_{jk} - 2p_k\right)}{2p_k\left(1 - p_k\right)}$$

- "SNP" Heritability: $h^2 = \sigma^2_A / (\sigma^2_A + \sigma^2_e)$
- Used open Mendel, VarianceComponentModels.jl for analysis

Potential Covariates and Transformations

- For weight and microbiome measure used fixed effects of sex, forest elevation and mean annual temperature as potential covariates
- No transformation needed for weight
- Need some form of transformation for OTUs as it is highly skewed data
- Used log for total abundance, square root for number of OTUs
- Individual OTUs standardized by sample's total abundance (OTU proportion) and then used inverse normal transformation

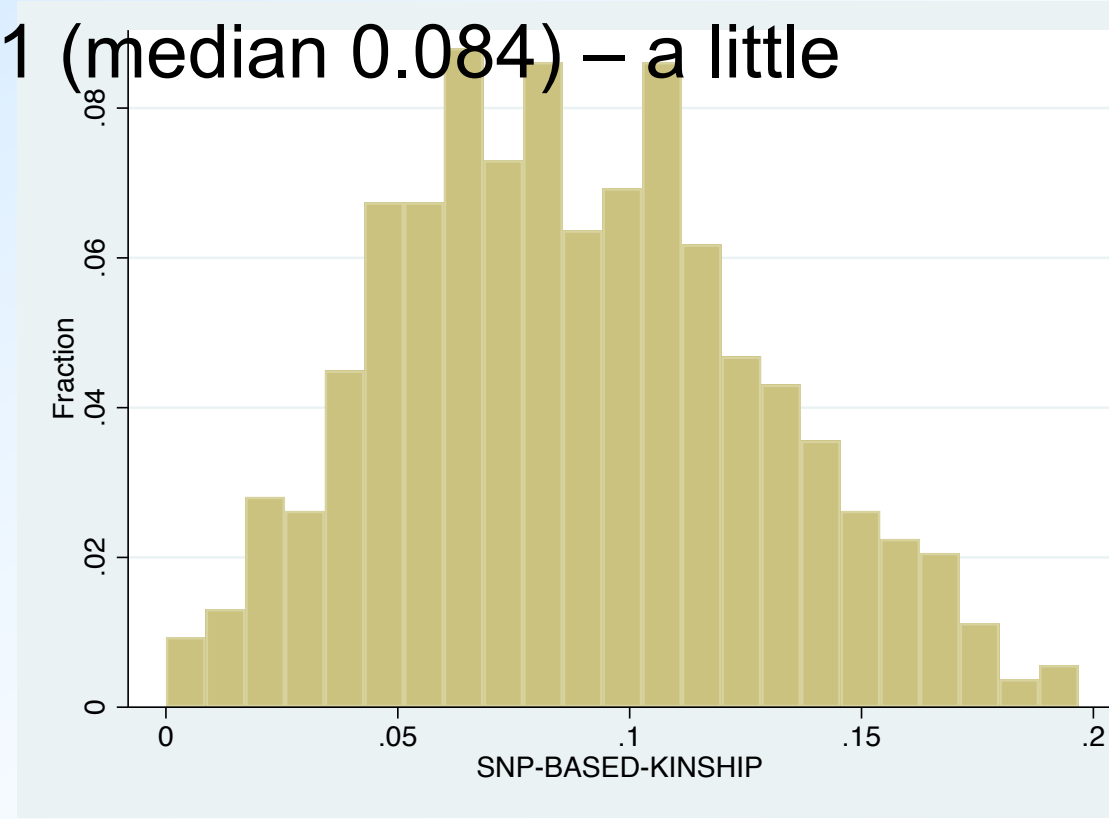
Results – GRM estimates of Kinship

- Within Forest:
- Inbreeding coefficient, 0.011 relatively low
  - Mating habits may reduce inbreeding
- Mean Kinship 0.0881 (median 0.084) – a little more than cousins

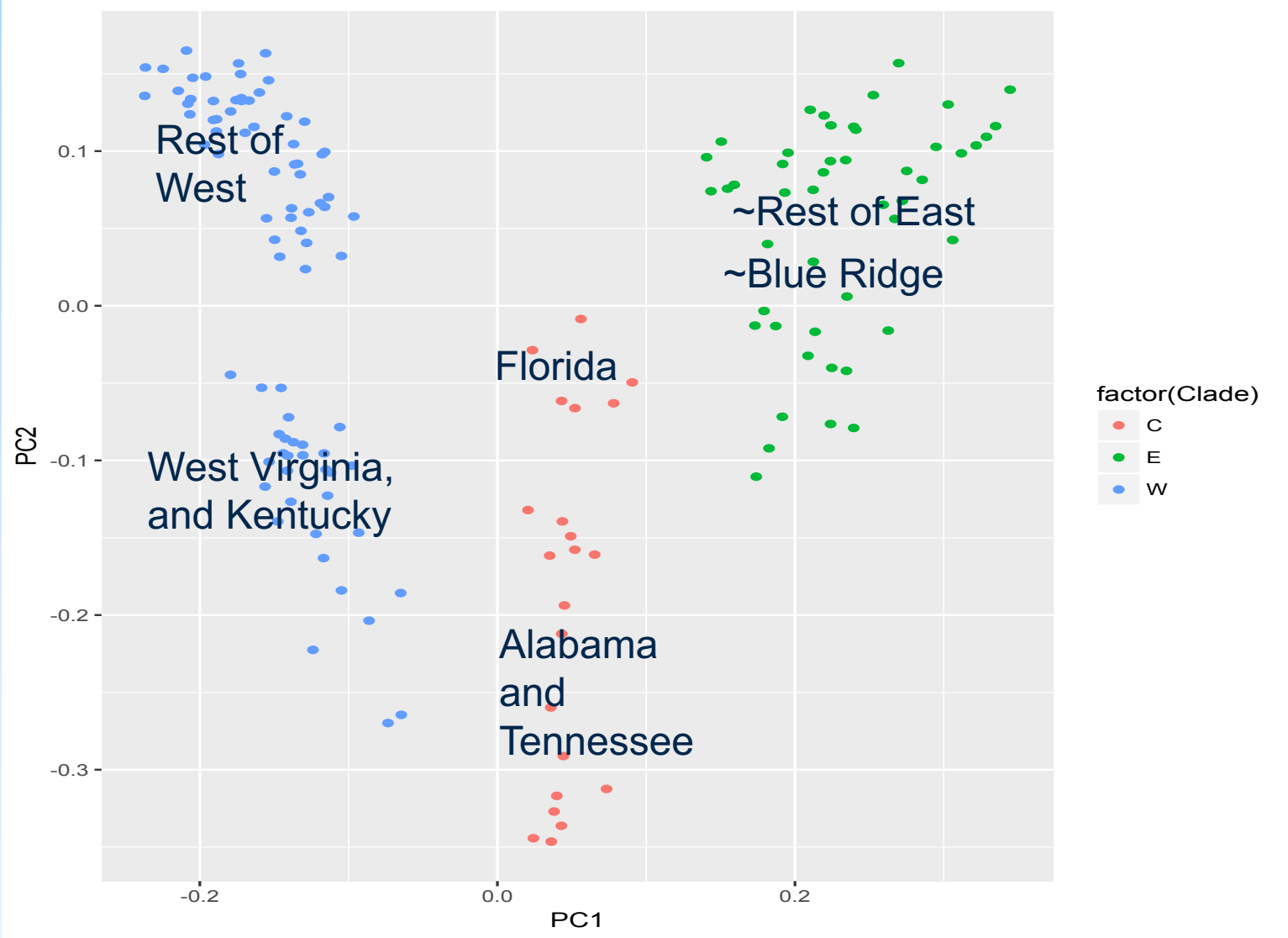
Results consistent with Dillard and Maigret 2017

- Between Forest:
- Mean Kinship 0.014

(median 0.001)



Principal Components



Results – Weight Heritability

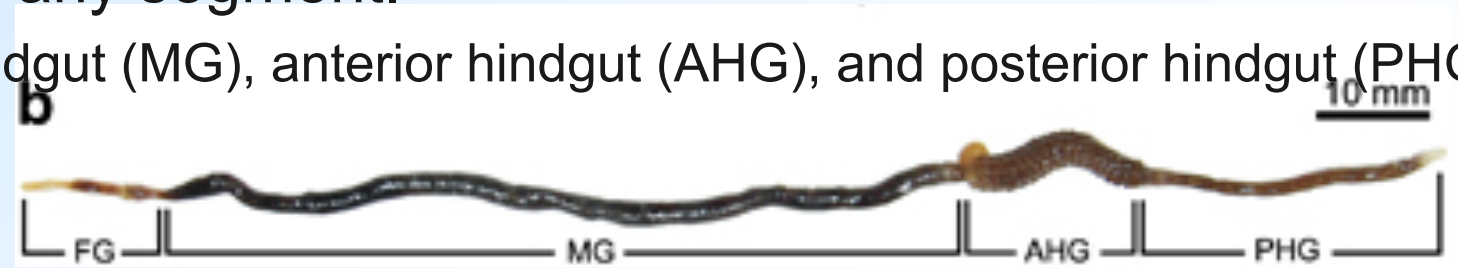
Parameter	estimate	SE	comment
μ	1.6998g	0.0131	Male at mean elevation and temp
β_{female}	0.2142g	0.0356	Females weight ~0.2g more than males
β_{elev}	0.0013g/ft	0.0004	Change in grams per foot
β_{MT}	0.0590g/deg	0.0094	Change in grams per degree Fahrenheit
σ_A^2	0.0205	0.0103	Additive genetic variance
σ_e^2	0.0305	0.0101	Environmental variance
h^2	0.4017	0.1381	Narrow sense heritability

Microbiome Heritability – Total Abundance

- No fixed effect values were significantly different from zero

Parameter	Estimate	SE	Mean-Counts
μ_{MG}	11.46	0.037	95,200
μ_{PHG}	11.49	0.051	97,600
μ_{AHG}	11.93	0.014	151,800
h^2_{MG}	0.3154	0.2791	
h^2_{PHG}	0.1959	0.3349	
h^2_{AHG}	0.4205	0.3574	

- Abundance heritability not significantly different from zero at any segment.
- Midgut (MG), anterior hindgut (AHG), and posterior hindgut (PHG)



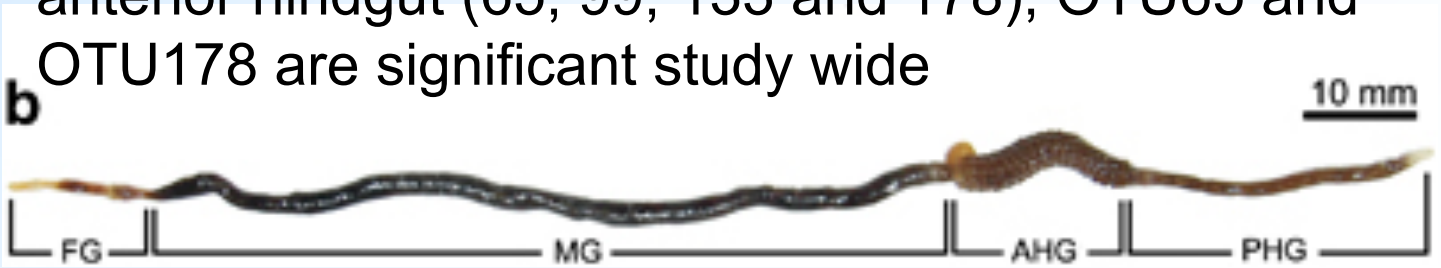
Correlations

Additive / Environment	MG	PHG	AHG
MG	1	-0.727	0.292
PHG	0.458	1	-0.823
AHG	-0.091	-0.050	1

Similar results with number of microbial species as the outcome (α-diversity measure)

Individual OTU results

- Of the 10 most abundant OTUs, (1, 5, 10, 11, 15, 16, 19, 24, 30 and 50) OTU1 and OTU24 show significant additive variance study wide. OTU5, OTU11, OTU30 and OTU50 are nominally significant (individual test p-value < 0.01)
- None of the 4 most abundant OTUs observed only in the midgut (2, 3, 4 and 9) have significant additive variance
- None of the 4 most abundant OTUs observed only in the posterior hindgut (6, 13, 29 and 36) have significant additive variance
- Of the 4 most abundant OTUs observed only in the anterior hindgut (65, 99, 133 and 178), OTU65 and OTU178 are significant study wide



OTU Classification

OTU	Phylum	Class	Order	Family
1	**Firmicutes**	**Clostridia**	**Clostridiales**	**Lachnospiraceae***
5	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae
10	Firmicutes	unclassified	unclassified	unclassified
19	Bacteroidetes	unclassified	unclassified	unclassified
24	**Proteobacteria**	**Deltaproteo-bacteria**	**Desulfo-vibrionales**	**unclassified**
30	Firmicutes	Clostridia	Clostridiales	unclassified
50	Actinobacteria	Actinobacteria	unclassified	unclassified
65	**unclassified**	**unclassified**	**unclassified**	**unclassified**
178	**Proteobacteria**	**Deltaproteo-bacteria**	**Desulfo-vibrionales**	**unclassified**

*Lachnospiraceae abundant in the termite microbiota - important for wood digestion (e.g. Dietrich et al. (2014) Applied and Environmental Microbiology.

Conclusions

- Too early to say much - more and better data needed
- Preliminary results show potential snp heritability of important bacteria groups
- With simple model used, we can’t rule out common household random effects
- Suggestions?