

How to GWAS - ConGen 2022

Eryn McFarlane

2022-09-11

What is GWAS?

Genome Wide Association Study

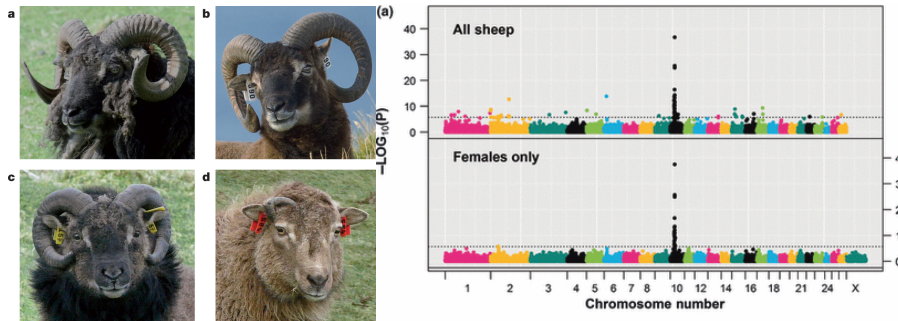
- ▶ $y = a + bx + g + e$
- ▶ y: phenotype
- ▶ b: fixed effect
- ▶ x: the SNP predictor coded as 0, 1, 2
- ▶ g: the polygenic random effect (e.g. Kinship matrix)
- ▶ e: the error term

Why do GWAS?

- ▶ Understand the genetic architecture of phenotypic traits
- ▶ link genotype to phenotype of individuals, extending quantitative genetics
- ▶ be able to measure selection directly on genomic regions, using fitness data

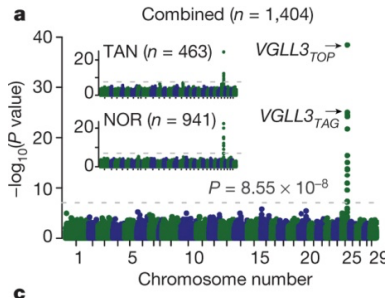
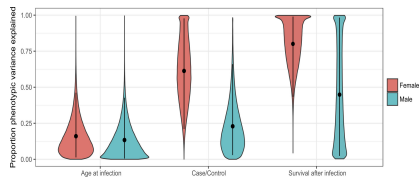
Soay Sheep Horns RXFP2 genotype

Johnson et al. 2011 Mol Ecol, Johnson et al. 2013 Nature



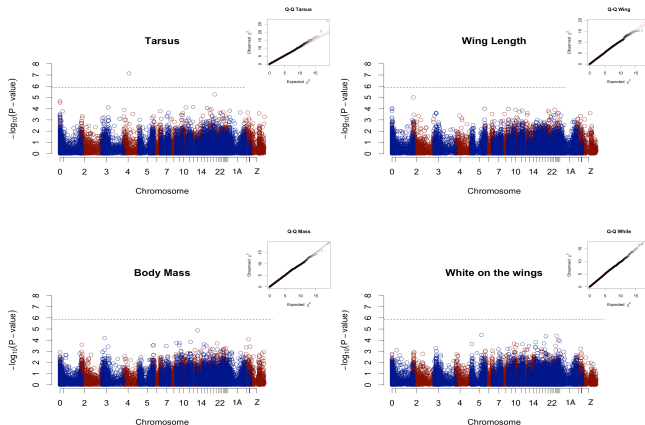
- ▶ 486 sheep
- ▶ 35 831 SNPs
- ▶ one huge effect locus, explains 76% of the variation in horn size

How often do people find huge effect loci using GWAS?



- ▶ Tasmanian devils face cancer 5 SNPs explain ~61% of variation in female survival (Margres et al. 2018 Mol Ecol)
- ▶ atlantic salmon - VGLL3 locus for delaying age of maturation (Barson et al. 2015 Nature)

What are the expectations for a GWAS?



- ▶ No significant SNPs for morphological traits
- ▶ 3 SNPs explaining $\sim 3\%$ of the variation in clutch size

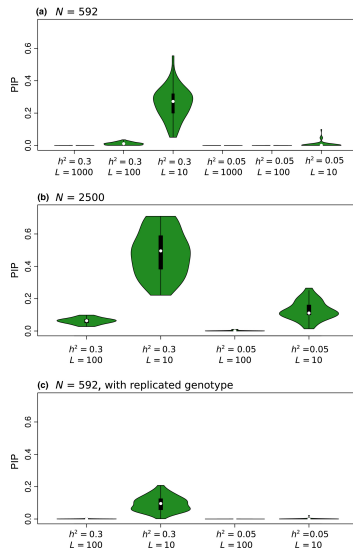
Silva et al. 2017 Heredity, Husby et al. 2014 Proc B

What are the expectations for a GWAS?

- ▶ Even if statistically significant loci are identified, they often explain a small proportion of the variation.
- ▶ 'Missing Heritability'
- ▶ not great for prediction of the trait

Some considerations for GWAS

- heritability of the trait
- genetic architecture of the trait (often unknown!)
- Linkage Disequilibrium
- number of individuals sampled
- allele frequency of causal loci (also unknown!)



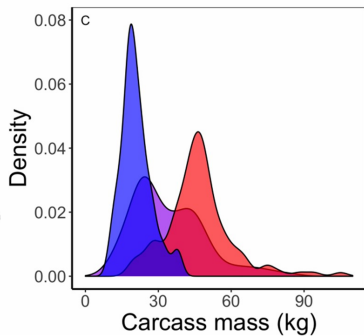
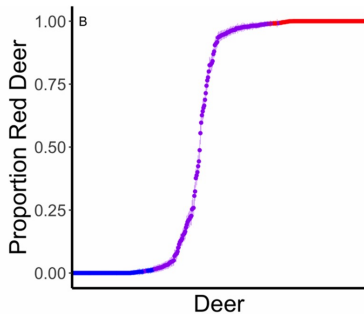
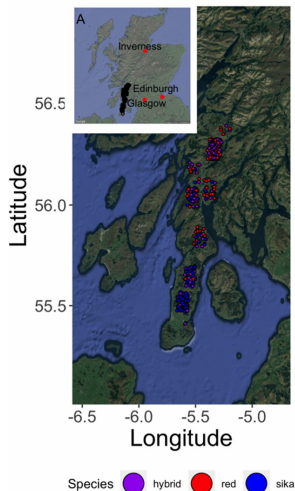
Some solutions

- ▶ Sample more individuals!
- ▶ replicated samples of individuals (RepeatABEL, RIP)
- ▶ Admixture mapping instead of within population GWAS
- ▶ sample in closed populations (but know this limits generalizability)

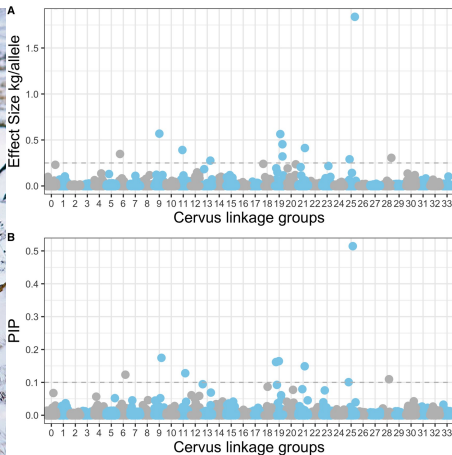
What is Admixture mapping?

- ▶ special case of GWAS
- ▶ GWAS on individuals with recombinant genotypes (from diverged species/populations interbreeding)
- ▶ Allows for QTL mapping with fewer individuals and fewer markers because of long tracts of LD

Deer Case study



Deer Case study



► McFarlane & Pemberton 2021 G3

So what are we doing today?

- ▶ Simulated GWAS, using GEMMA
- ▶ BSLMM in GEMMA (GEMMA can do LMM as well)
- ▶ Simulated 10,000 SNPs, 2000 individuals
- ▶ ' Choose your own adventure'

What is BSLMM, and how does it work?

-Bayesian Sparse Linear Mixed Model

- ▶ Models two distributions for the genetic effects:
 - ▶ polygenic distribution where we estimate the overall genetic variation
 - ▶ sparse (or beta) distribution where we estimate the effect sizes for specific SNPs
 - ▶ assuming a negligible effect size for all the other SNPs
- ▶ This sparse framework allows us to estimate effect sizes for only a few parameters, which helps with the $p \gg N$ problem that we have.
 - ▶ ask me about sparsity in general, it's a current research avenue of mine!
- ▶ Gives $n.\gamma$, or number of expected SNPs explaining the trait.
- ▶ Accounts for kinship before the two genetic effect distributions

Learning outcomes of GWAS tutorial

- 1) Understand what a GWAS is, and how to run it in gemma
- 2) Think about specific predictions for GWAS, and what we really expect to see for most of them
- 3) Think critically about when GWAS are robust, when they should be taken skeptically, and why they're interesting to do anyway