Genomic Prediction: from animals to plants and back

Luc Janss
GenSAP SFA2 leader “genomic prediction”



GenSAP

Ministeriet for Fødevarer, Landbrug og Fiskeri
NaturErhvervstyrelsen
gudp

Outline

- **Use of genotyping by sequencing**
- **Disentangling genomic covariance**
- **Biases in breeding values**

Sequencing

- **Whole genome sequencing at high depth**
  - Limited samples (10's – 100's)
  - Other samples are imputed to full sequence
  - Rare allele imputation poor
- **Whole genome sequencing at low depth (<1x)**
  - Larger samples
  - Good haplotype reference needed (HapMap in humans)
- **Reduced genome sequencing**
  - Larger samples
  - Rare alleles available
  - Low depth causes missing data and bias

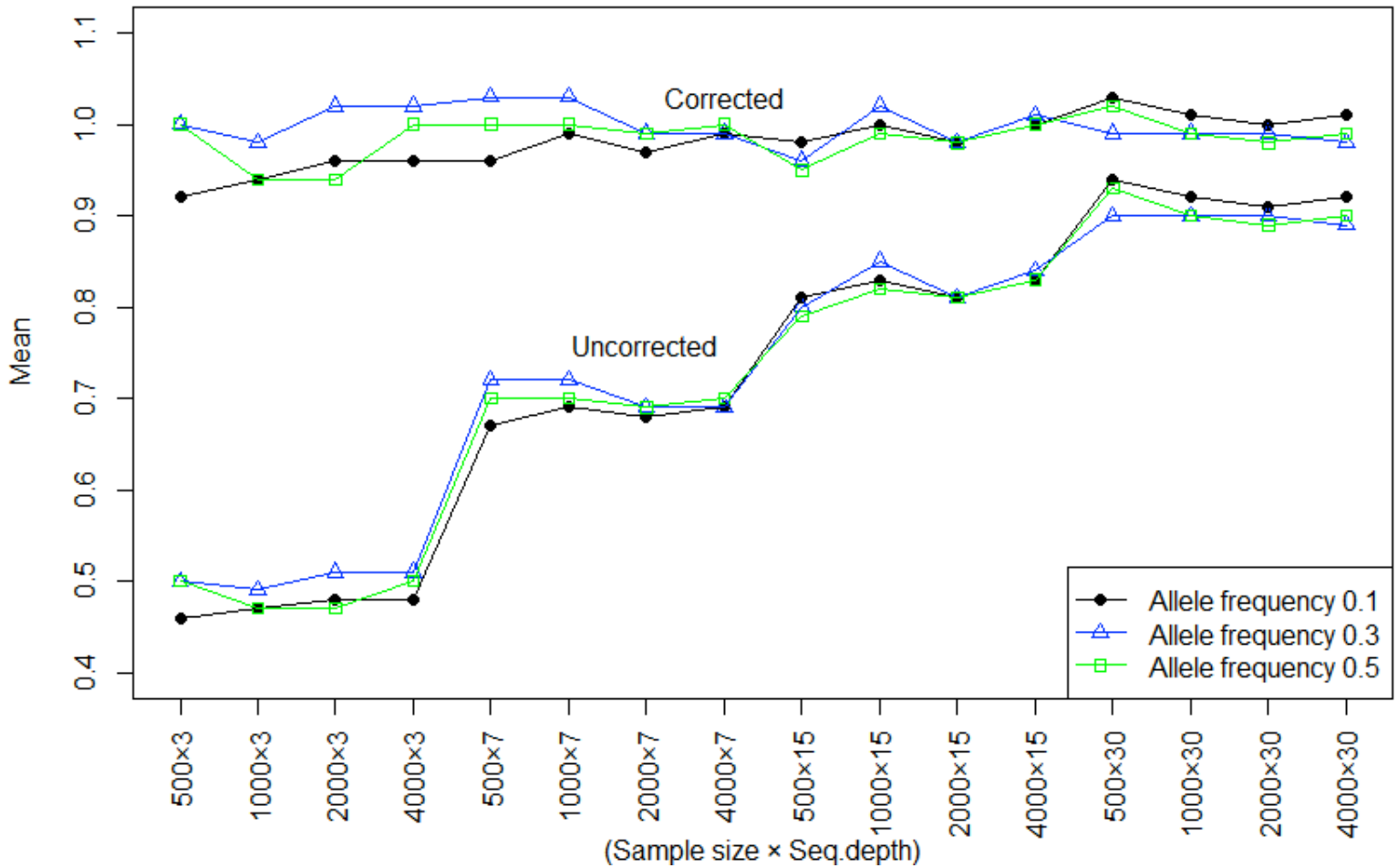
Determining a heterozygote genotype (AB) by sequencing

Reading	1 Read	2 reads	3 reads	4 reads	5 reads
AA	50%	25%	12.5%	6.3%	3.1%
AB	0%	50%	75%	87.5%	93.8%
BB	50%	25%	12.5%	6.3%	3.1%

Using low-depth genotyping by sequencing

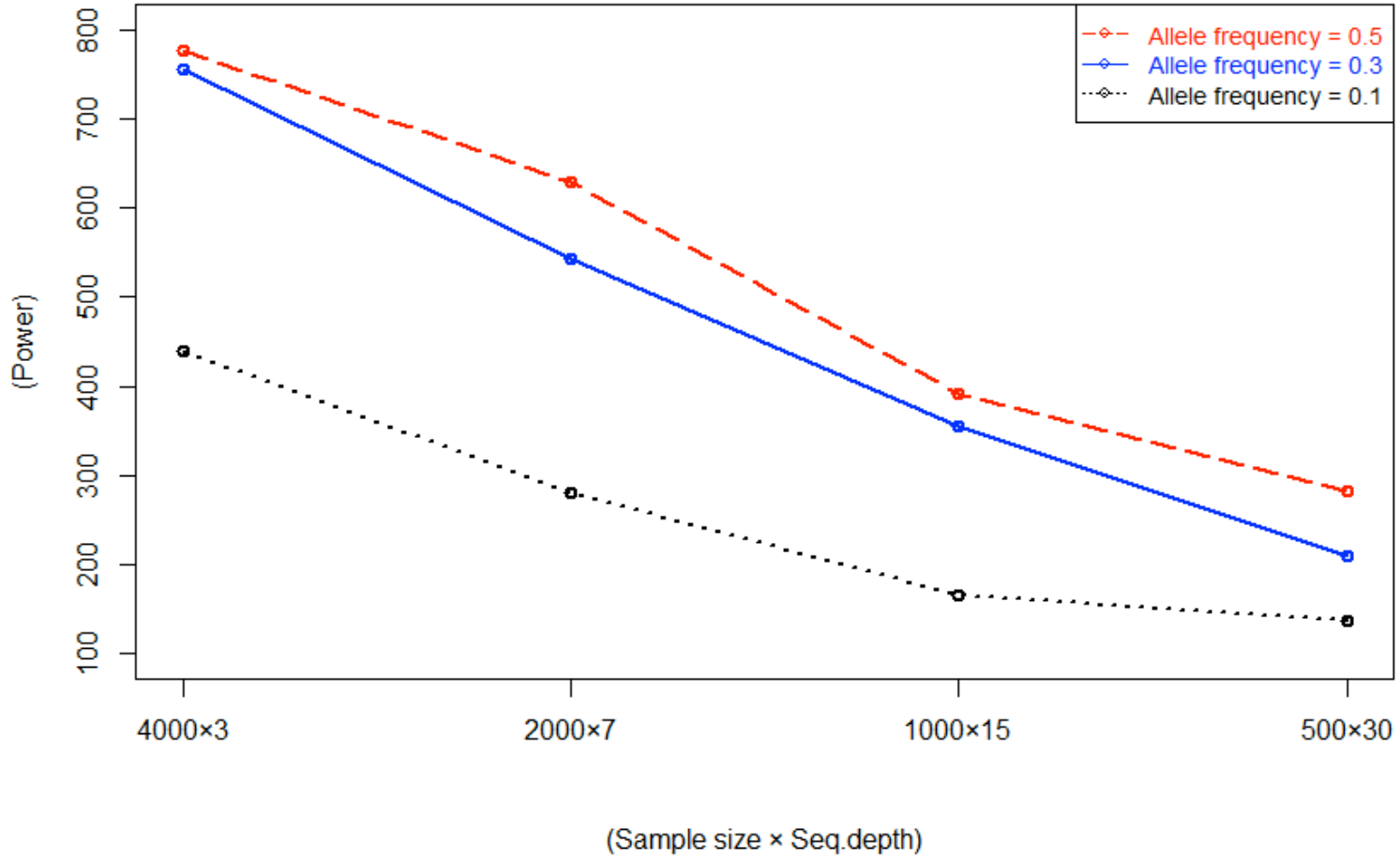
- **Using as low as 1 read works!**
  - On average the genotype is OK
  - A phenotype has more value than the 'damage' of inaccurate genotype
  - Multiplexing and doing more samples at lower depth attractive
- **But**
  - Inaccuracy in genotype causes bias
    - Under-estimated allele effects in GWAS;
    - over-estimation of genomic variances
  - No reads = missing genotype!

Under-estimation of allele effect using GBS data for GWAS
From: Ashraf et al., Theor. Appl. Genet., 2014



At constant sequencing, highest power to detect a significant association at low depth

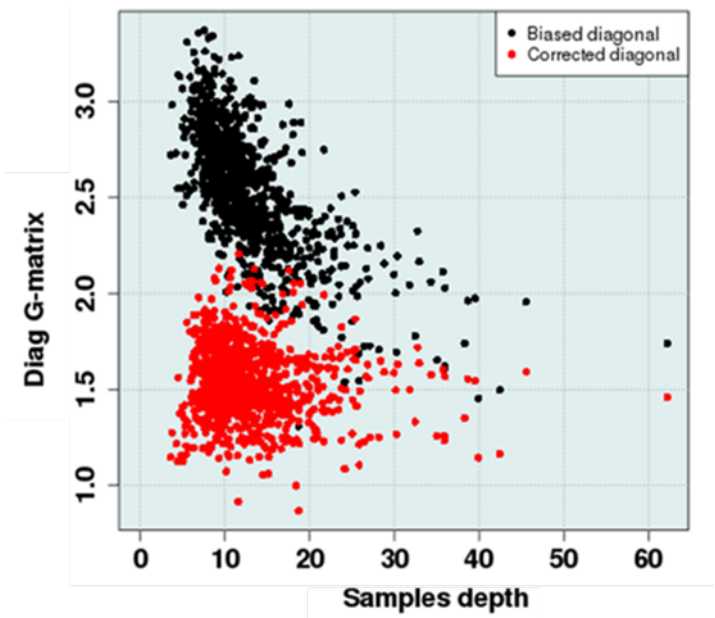
From: Ashraf et al., Theor. Appl. Genet., 2014



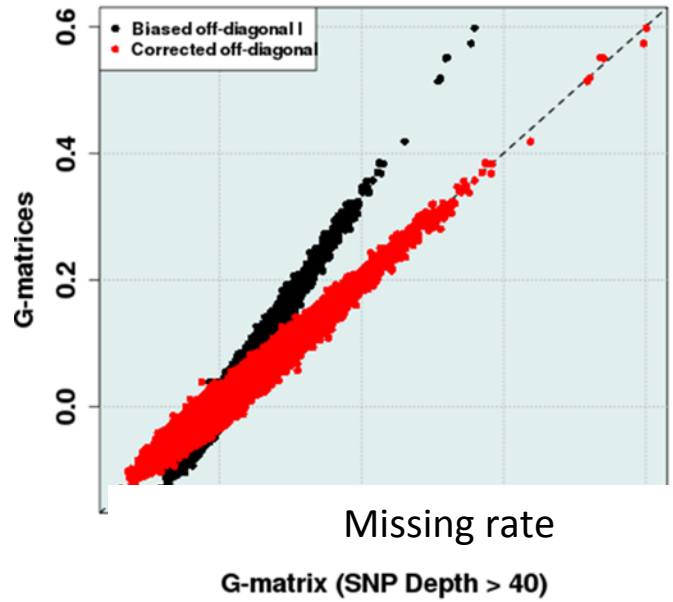
Bias from low-sequencing depth in computed genomic relationships

Cericola et al., PAG conference 2015

Diagonals are biased with sample sequencing depth



Off-diagonals are biased with sample missing rate when using standard VanRaden adjustment



Lessons learnt from genotyping by sequencing

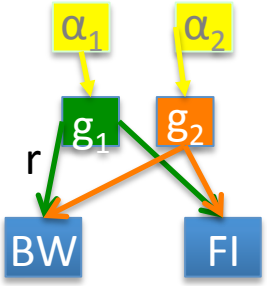
- Great tool for genotyping
  - Cost effective, competitive with arrays
  - Directly applicable to tetraploids and pools
    - Use allele-frequency estimates instead of genotypes
- Biases due to genotype inaccuracy, but we learn to deal with it
- Implications (in animals)
  - For species without genome sequence and arrays
  - Similar biases for imputed genotyped?
  - Better missing rate adjustment

Disentangling covariance between traits: mapping SNPs to different correlation structures

Janss, WCGALP Conference 2014

- 748 F2 Mice
- Body Weight, Feed Intake and Feed Efficiency

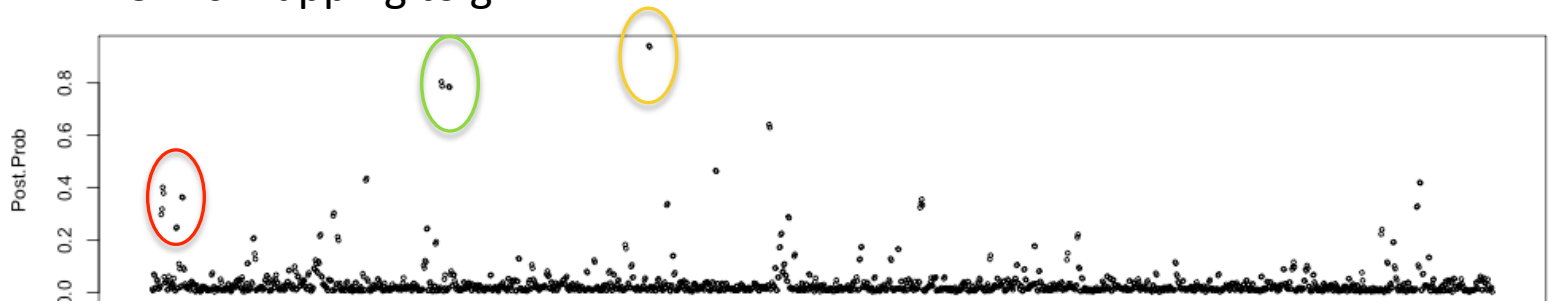
$$\begin{cases} \mathrm{BW} = Xb_1 + \overset{>0}{r_1} g_1 + \overset{>0}{r_4} g_2 + e_1 \\ \mathrm{FI} = Xb_2 + \overset{>0}{r_2} g_1 + \overset{<0}{r_5} g_2 + e_2 \\ \mathrm{FE} = Xb_3 + r_3 g_1 + r_6 g_2 + e_3 \end{cases}$$



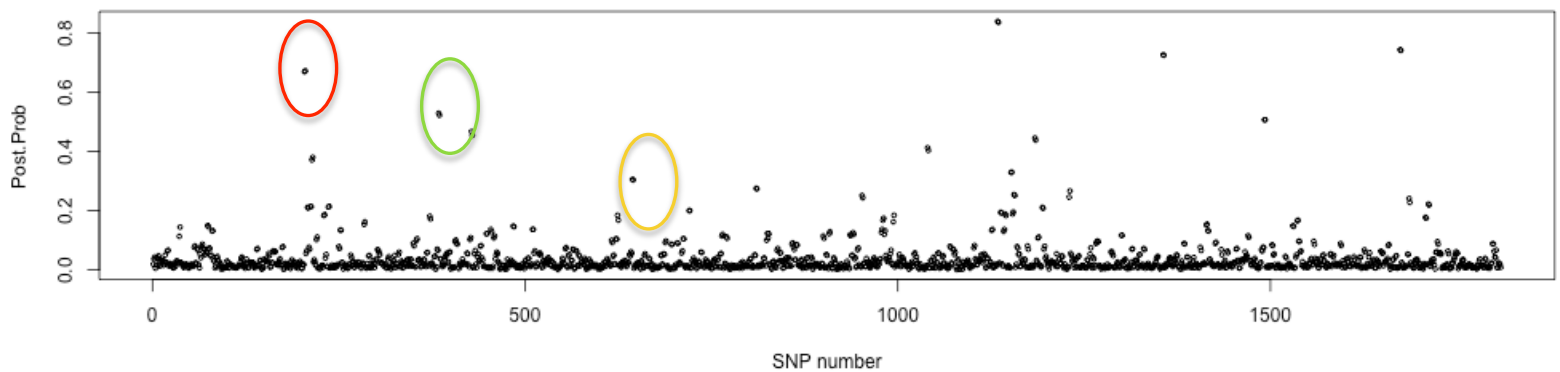
- Map SNPs to g1 (BW+, FI+) and g2 (BW+, FI-) using SNPs and mixture model in g1, g2

Two genomic values explaining BW, FI, FE

SNPs mapping to g1



SNPs mapping to g2



Interpreting SNP effects

$SNP_i \xrightarrow{a_i} g_2 \xrightarrow{0.047} BW$

$g_2 \xrightarrow{-0.037} FE$

Two genomic values explaining BW, FI, FE

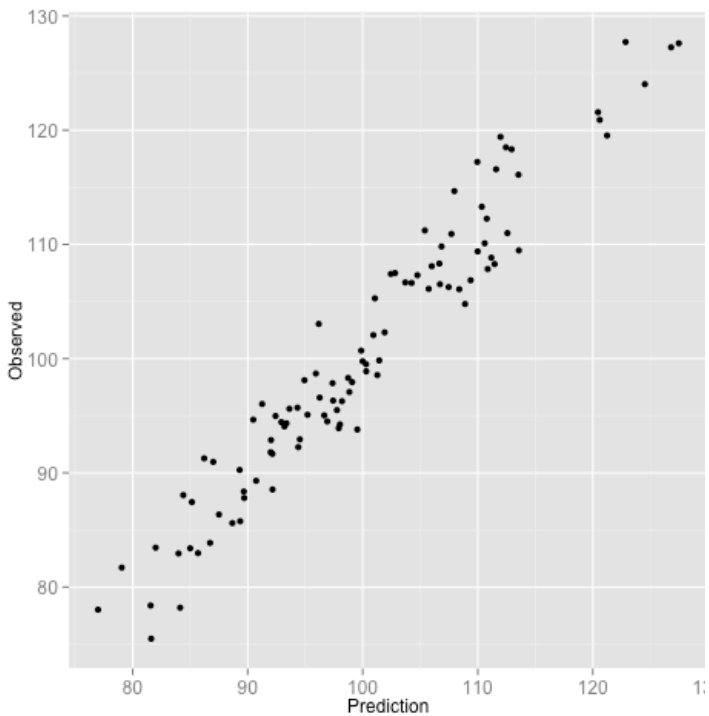
Trait	Correlations g1	g2
BW	0.34	0.34
FI	0.40	0.08
FE	-0.07	0.49

Real (agricultural) applications

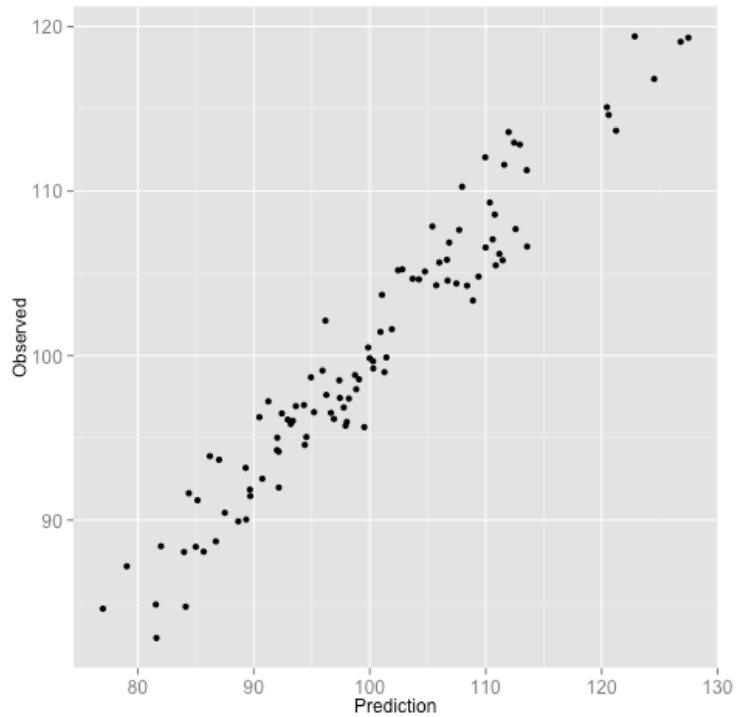
- Pooling data and prediction across breed
  - Consider traits in different breeds as different but correlated traits
  - Disentangle SNPs that contribute to correlation and those that are breed specific
- Multi-environment data
  - Consider traits multi-trait
  - Map common and environment specific SNPs
  - Breeding values for all environments

Learning about biased breeding values

Unbiased predictions



Biased predictions



Can compare between old/new, genotyped/non-genotyped, different accuracies, etc.

Can not compare to anything outside the group

Data design and bias

Fixed

Trait = block + genotype + residual

Biased

Random

Trait = block + genotype + residual

Unbiased

Trait = block + genotype + residual

Biased

Trait = block + genotype + Parent x Parent + res

Unbiased

Biased breeding values

- Getting the scale wrong can be expensive!
- Bias can arise from
  - Confoundment genotypes x environments
  - Pre-selection, preferential treatment
- Can learn about source of bias by looking at correlations of genotypes / blocks / years etc.
  - If you analyse all data in one step

Summary

- Useful cross-fertilization between animals and plants
- Use of sequencing for genotyping
- Refinements in field-data analysis
- Multi-trait models for across breed & GxE

Thank you