Evaluating inbreeding measures using a whole-genome sequenced cattle pedigree

S. Alemu, N.K. Kadri, C. Charlier and T. Druet*

Unit of Animal Genomics, GIGA-R, University of Liège, Belgium

EAAP meeting - Ghent 2019

Introduction

- Inbreeding
  - Negative consequences such as increased occurrence of recessive genetic disorders and inbreeding depression
  - Considered in conservation programs
- Hypotheses for inbreeding depression (ID)
  - Increased homozygosity at rare recessive deleterious alleles
  - Decreased heterozygosity at variants presenting heterozygous advantage (overdominance)

Inbreeding coefficient *F*

- Estimation
  - Pedigree-based or genomic measures
- Numerous genomic estimates:
  - Maximum likelihood, method-of-moments, homozygosity or heterozygosity measures, diagonal elements of the GRM, ROH
- True inbreeding coefficients unknown:
  - Simple comparisons or simulations

Objectives

- Evaluation of inbreeding measures
  - In a livestock population
- Use whole-genome sequence data to mimic scenarios
  - Real genomic structure (LD, allele frequency spectrum, …)
  - Contains causative variants, rare alleles, different functional categories
- Estimation and prediction
- Genome-wide and locus specific

Data

- 266 sequenced Dutch-Holstein cattle
  - Cover > 15x
  - 145 parents
  - 100 offspring with both parents known
  - 13 x 10^6 SNPs
- Estimating of the inbreeding coefficient
  - 37,675 SNPs (50K array)
  - 5,977 SNPs (Low-density – LD array)

Inbreeding coefficients

- Pedigree-based F_{PED}
- Correlations between uniting gametes F_{UG} (GCTA)
- Based on diagonal elements GRM F_{GRM}
- Excess homozygosity F_{HOM} (PLINK)
- Maximum likelihood estimator F_{LIK} (TrioML - Coancestry)
- Homozygous-by-descent (HBD) segments (RZooRoH):
  - Four classes of HBD segments (length 20, 4, 0.8 and 0.2 cM)
  - Proportion of the genome in HBD classes with rate ≤ T F_{HBD-T}
  - Similar to F_{ROH} with 50K

Scores related to inbreeding

- Obtained from the sequence data
- Homozygous mutation load
  - Homozygosity at rare and young alleles
- Homozygosity score at intermediate frequency variants
  - Homozygosity at variants with MAF > 0.15
- Whole genome sequence homozygosity
  - Homozygosity at all variants

Variant classification

- Age of alleles
  - Allele frequency (rare)
  - Derived allele (less SNPs)
  - Not in another breed (not in Belgian Blue): enrichment in young alleles
- Functional annotation
  - All alleles (no assumption)
  - Synonymous versus non-synonymous
  - Excluding intergenic and intronic regions (functional space)

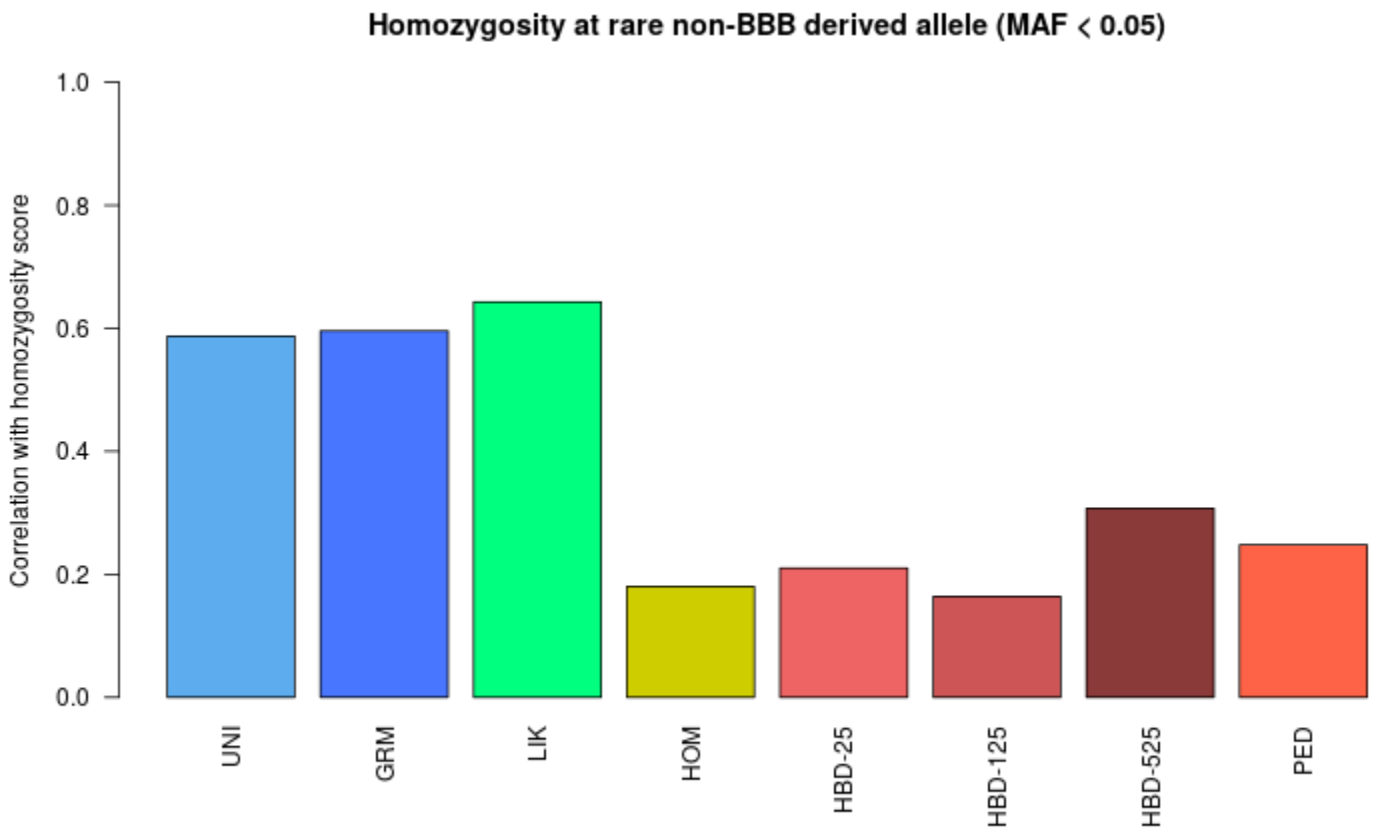
Age of alleles

- Using GEVA (Albers and McVean, 2019):
  - Genealogical Estimation of Variant Age
  - Recombination clock (in generations to TMRCA)

Allele frequency	All derived alleles		Alleles absent from the BBB sample	
	Number of alleles	Average TMRCA	Number of alleles	Average TMRCA
≤ 0.05	20480	641	4633	386
≤ 0.15	23667	1237	2165	575
≤ 0.30	20687	1637	635	773
> 0.30	28111	2606	36	1799

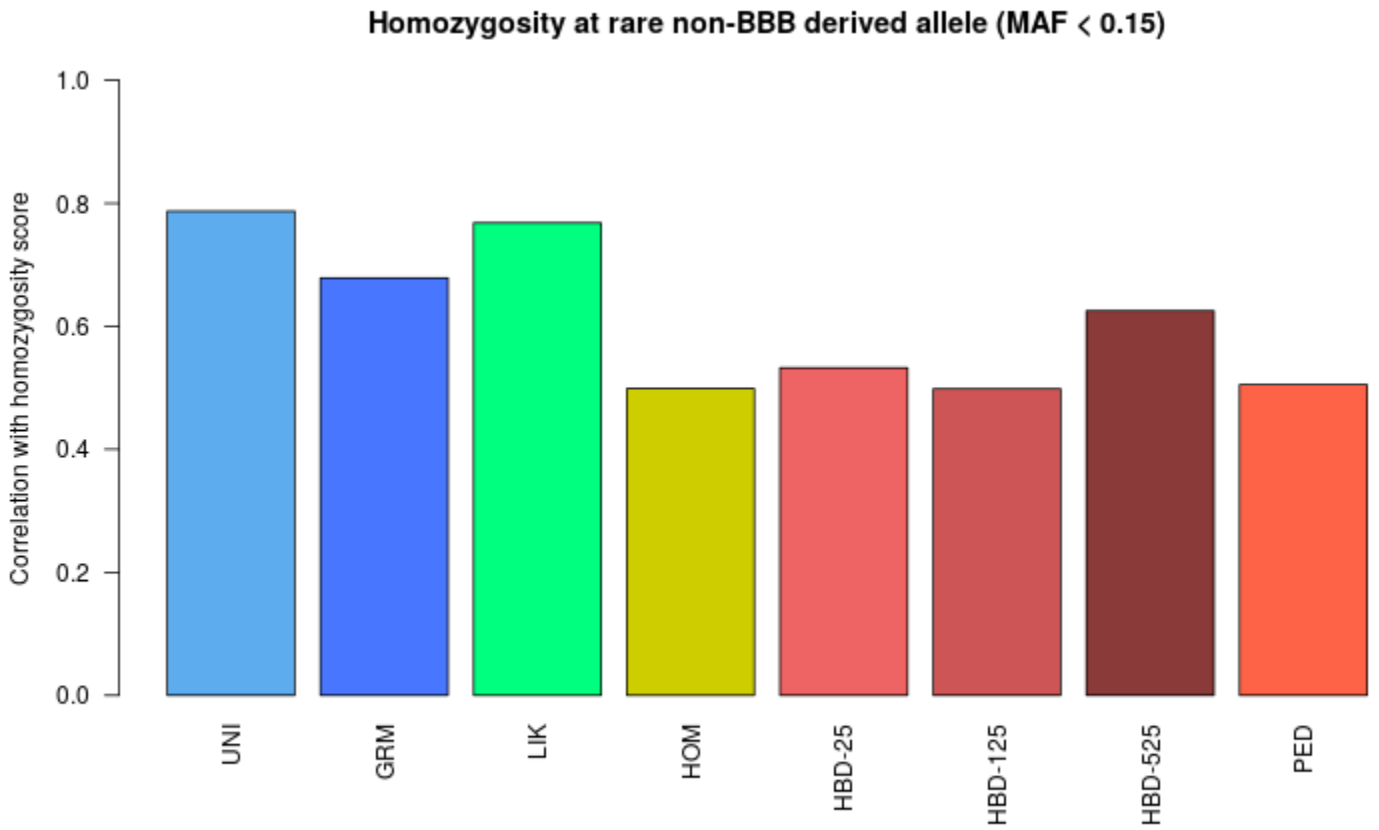
Homozygous mutation load

- Rare recessive deleterious alleles



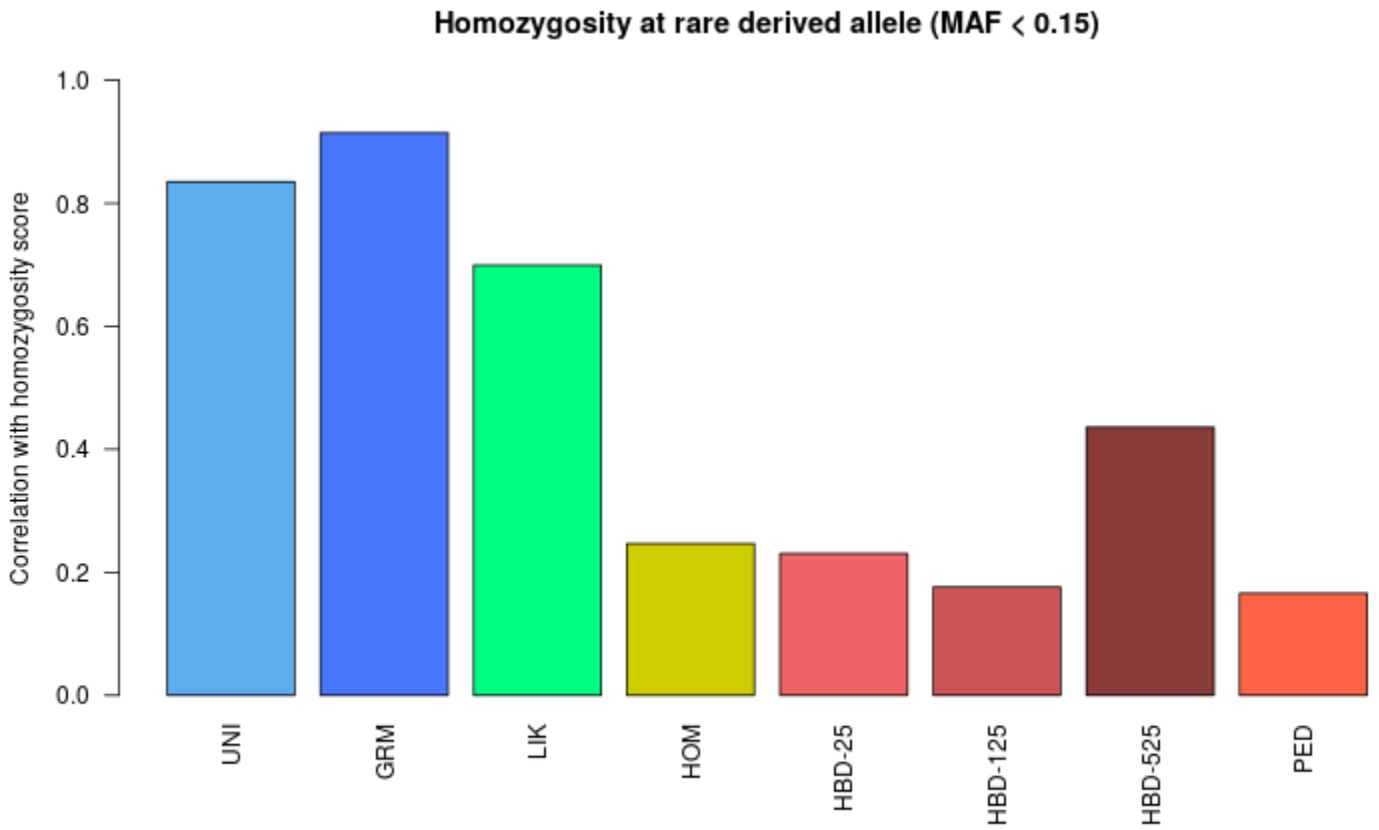
Homozygous mutation load

- Allele frequency < 15%



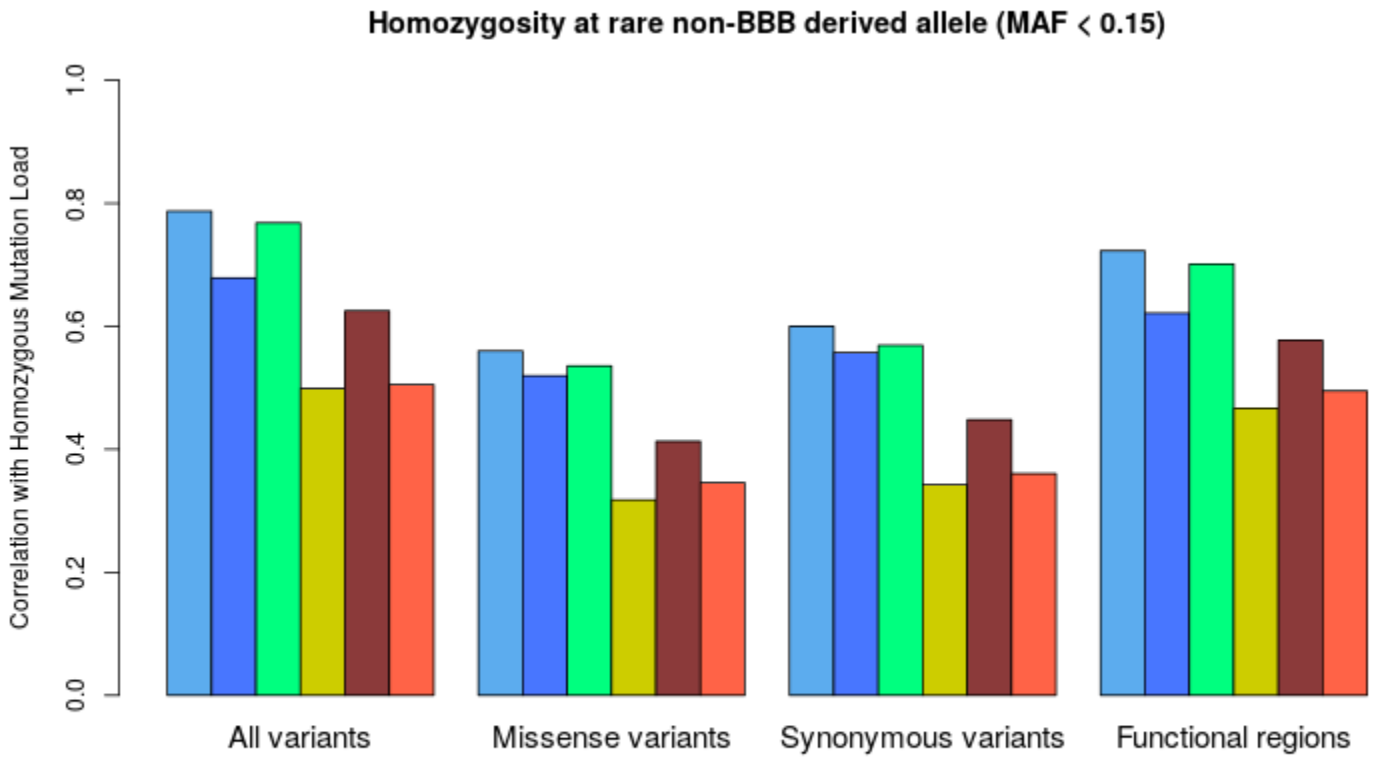
Homozygous mutation load

- Including derived alleles observed in BBB



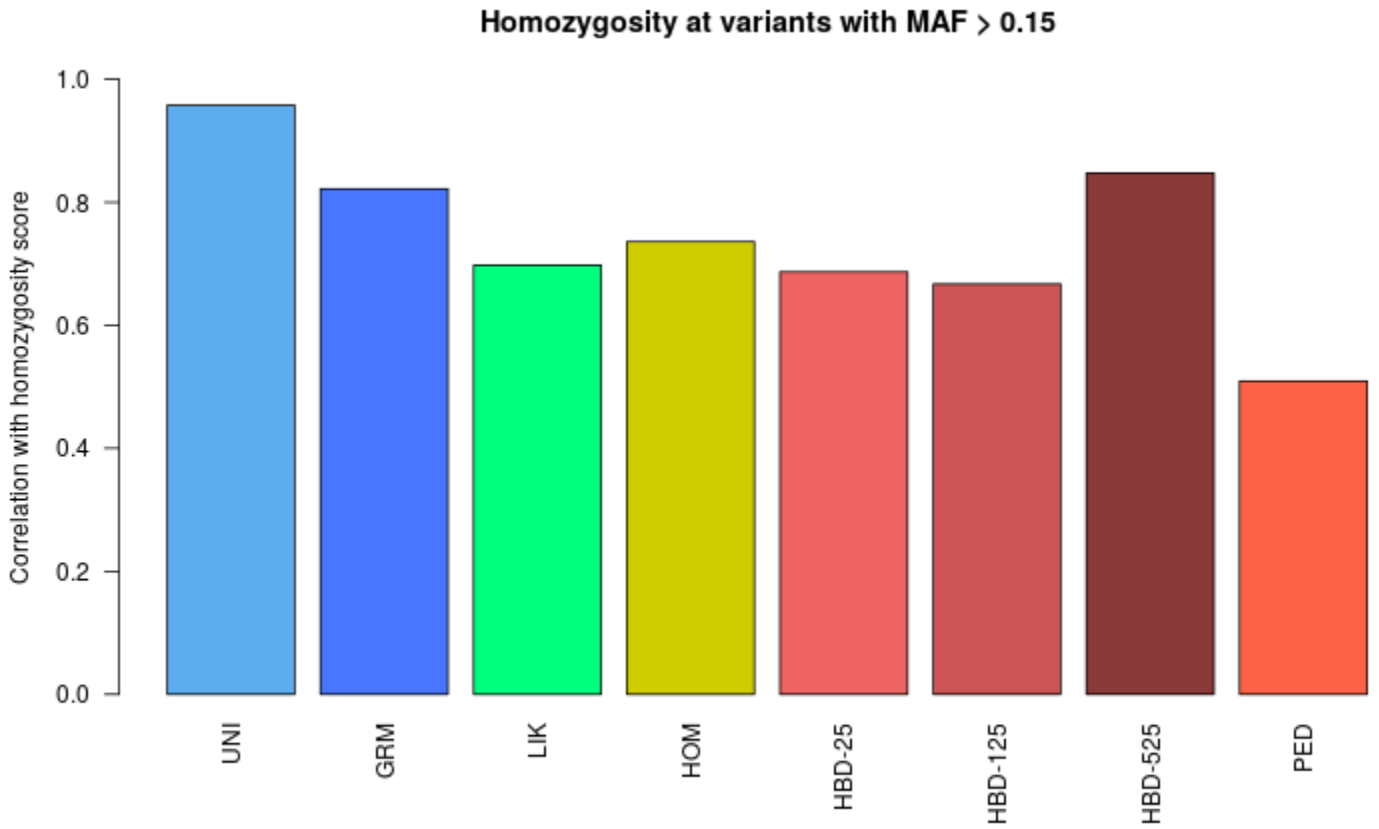
Homozygous mutation load

- For different annotations categories



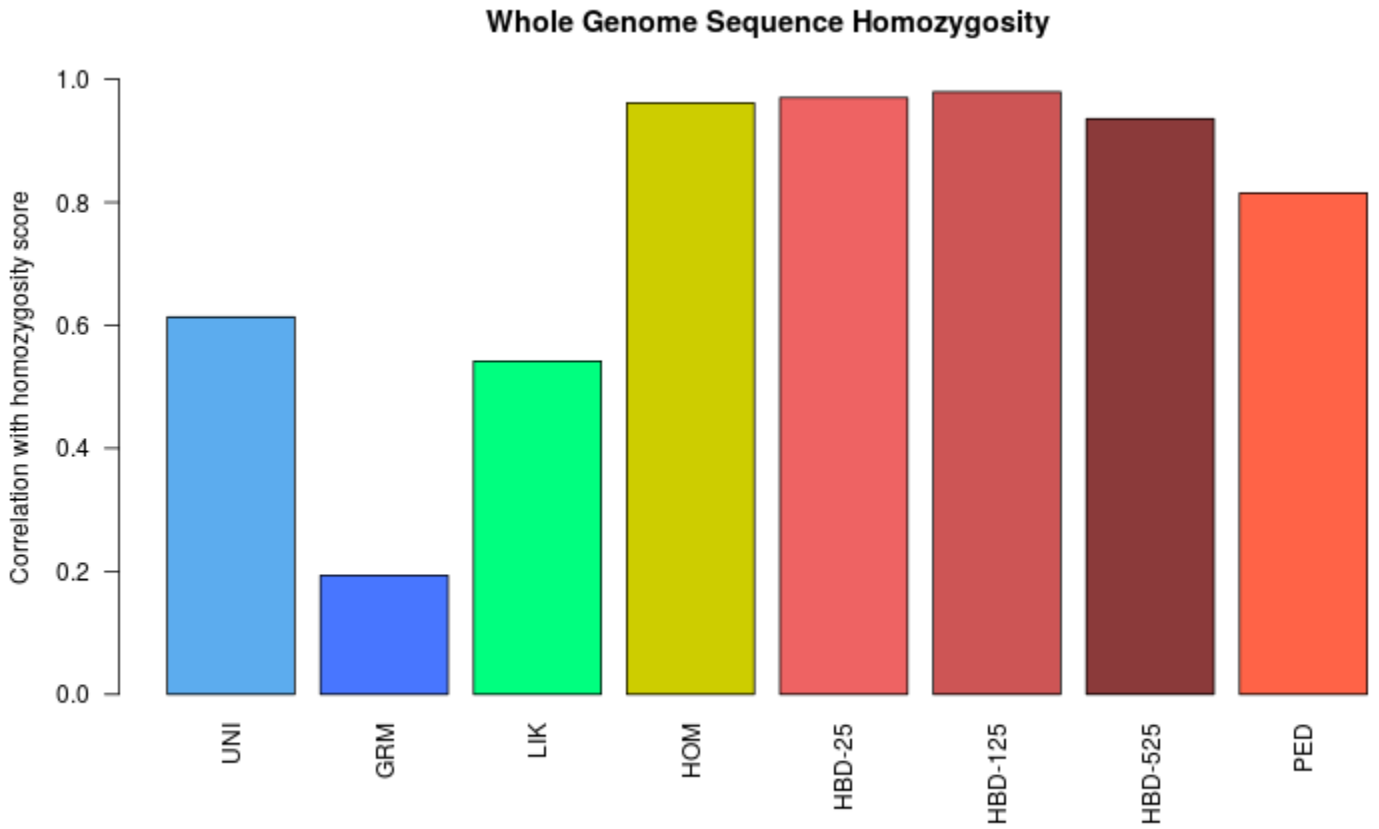
Heterozygous advantage

- Scenario 2: decreased heterozygosity



Whole genome sequence homozygosity

- Homozygosity at all alleles



Using low-density array

- Modest decrease of correlations
- Same ranking with LD array
  - Across scenarios
- Methods still efficient with 6K

Predicting *F*

- Prediction of inbreeding coefficients
  - 100 sequenced trios
  - Use genotypes of parents to predict *F*
  - Compute scores with sequence of progeny
- Same trends across scenarios

Local trends

- Estimate scores in 1 Mb windows
  - Some filtering on number of markers, homozygotes
  - Huge variation
  - Average correlation

	Hom (< 0.15)	Hom (MAF > 0.15)	Hom WGS
F_{PED}	0.02	0.05	0.07
F_{UG}	0.45	0.72	0.67
F_{GRM}	0.43	0.57	0.44
F_{HBD}	0.38	0.73	0.75

Local trends – LD array

- Estimate scores in 1 Mb windows
  - Some filtering on number of markers, homozygotes
  - Huge variation
  - Average correlation

	Hom (< 0.15)	Hom (MAF > 0.15)	Hom WGS
F_{PED}	0.02	0.05	0.07
F_{UG}	0.24	0.48	0.43
F_{GRM}	0.25	0.40	0.32
F_{HBD}	0.35	0.60	0.62

Conclusions

- F_{UG} and F_{GRM} better for inbreeding depression
  - Focus on rare allele (and not young)
- F_{HBD}, F_{HOM} and F_{PED}: proportion genome IBD
  - Improve with young alleles (“breed specific”)
- Lower correlations with NS variants
  - And with homozygosity at rare variants
- For locus-specific estimation and low-density, F_{HBD} performed well

Acknowledgements

Unit of Animal Genomics

Setegn W. Alemu
Naveen Kumar Kadri
Chad Harland
Carole Charlier
Michel Georges

Data set

Erik Mullaart

Funding

fnrs
LA LIBERTÉ DE CHERCHER

erc
European Research Council
Established by the European Commission