Outline of the Armidale Summer Course: Bases and Computing Techniques for Genomic Predictions and GWAS

Instructors: Mehdi Sargolzaei and Daniela Lourenco

Dates: February 20-24, 2023

Brief description: This course encompasses the background needed to perform genomic analyses used in Animal Breeding and Genetics. It involves the generation of genomic data using QMSim (Sargolzaei and Schenkel, 2009), imputation of missing SNP using FImpute (Sargolzaei et al., 2014), and overall genomic analyses using BLUPF90 programs (Misztal et al., 2014).

Day 1: Mehdi

- 1. Introduction to Genomic data
 - a) History of the use of genomic data
 - b) Genomic markers
 - c) Statistics of genomic data
 - d) Genomic files
 - e) Quality control of SNP data
- 2. Simulation of genomic data
 - a) Simulated vs. real data
 - b) QMSim for genomic data simulation
- 3. Introduction to Unix environment and tools
- 4. Exercise: Simulate a genomic dataset with QMSim and use Unix tools for data manipulation and statistics

Day 2: Mehdi

- 1. Introduction to imputation
 - a) Implications of missing genomic data
 - b) Methods for imputation to medium and large density
 - c) State-of-the-art in imputation
- 2. Imputation
 - a) Best practices before and after imputation
 - b) FImpute for imputation
 - c) Accuracy of imputation
- 3. Exercise: Impute missing genotypes with and without pedigree information, and compute imputation accuracy

Day 3: Daniela

- 1. Introduction of BLUPF90 family of programs for the analyses of mixed models including single and multiple traits, maternal effects, and repeatability models
 - a) Renumbering datasets: renumf90
 - b) Estimation of breeding values and VCE: blupf90+
 - c) Estimation of breeding values and VCE Bayesian: gibbsf90+
- 2. Exercise: use of BLUPF90 programs with real data for single and multiple trait models

Day 4: Daniela

- 1. Introduction to Genomic Selection
- 2. Theory of Single-step GBLUP (ssGBLUP)
- 3. Creation and handling of genomic relationship matrices with preGSf90
- 4. Quality control for genomic and pedigree relationships
 - 1. Calling rate
 - 2. Parental exclusions
 - 3. Distributions of diagonals and off-diagonals of G
 - 4. Differences between matched **G** and A_{22}
 - $5. \ Eigenvalues/eigenvectors-population\ stratification$
- 5. Validation techniques for testing genomic models
- 6. Exercises: Application of quality control and use of single-step with BLUPF90 programs in simulated data sets

Day 5:

- 1. Base allele frequencies from Gengler's method
- 2. Estimating SNP effects from GBLUP-based models
- 3. Indirect predictions using SNP effects
- 4. Weighted GBLUP and ssGBLUP
- 5. Genome-wide association studies (GWAS)
- 6. Accounting for unknown relationships in ssGBLUP (UPG and metafounders)
- Exercises: Compute SNP effects from ssGBLUP, run indirect predictions for young animals, and do ssGWAS (variance explained by SNP and p-values) with BLUPF90 programs