« Using whole-genome data in plant breeding: GWAS and Genomic Selection to genetically characterize key traits in large germplasm collections»

Plant breeding has traditionally relied on scoring the phenotype. Dramatic reduction in the price of nucleotide sequencing has led to description of a growing number of whole genome sequence in crop plants and wild species. Combined with affordable genotyping of individuals, often based also on sequencing, this has led to a dramatic increase in marker density, allowing selection at genomic level to be feasible. The course focuses on introduction to the R statistical software and on Genome Wide Association Studies (GWAS) and Genomic Selection (GS) which is becoming possible in an increasing number of plant species, even without a reference genome.

The aim of this course is to provide quantitative and statistical framework to apply for large , potentially unbalanced, dataset using Mixed Linear Models. A brief overview of typical bioinformatics pipelines for the discovery and management of Single Nucleotide Polymorphisms (SNP) will be provided, as well as hands-on experience for the discovery and the analysis of population structure. The course will then provide the participants with in-depth relevant theory of GWAS, with hands-on experience with various case-studies of GWAS. The course will also provide the participants with relevant theory of Genomic Selection using detailed case-studies in plants.

The learning outcomes upon course completion are:

Knowledge

- Understanding of the main statistical concepts of Mixed Linear Model and its use for field trials and quantitative genetics.
- Understanding Association Genetics
- Understanding the principles of Genomic Selection and Prediction
- Knowledge of the assumptions and limitations of different types of GWAS and GS model.

Skills

- Be able to implement a R script for MLM for field trail analysis, including data management and display
- Be able to implement R script for the analysis of population structure in a species using various methods, and displaying data
- Be able to implement R script for GWAS
- Be able to implement R script for GS, using different algorithms, and Cross-Validation procedure.

Pre-course assignments
The attendees will receive some literature and instructions to install software before the start of the Course and should come with their desktop computer. It is asked to prepare a (electronic, not printed) poster (ePoster) on their PhD or research project, that will be presented to the audience in 3 minutes.

Pedagogical Approach

The course is based on short lectures immediately followed by hands-on with case studies, working groups and at plenum. It will include a visit to an experimental station of VIR.

Prerequisite Knowledge

The course is intended for scientists with consolidated knowledge of plant breeding and quantitative genetics, and experience in bio-statistics. Course tutorials will use statistical packages in R. Experience is R is recommended, but is not essential.