

Advanced Course

APPLIED STATISTICAL METHODS IN PLANT GENOMICS Zaragoza (Spain), 18-29 February 2008

PROGRAMME

1. **Opening and introduction (2 hours) (I. Romagosa)**
2. **General statistical theory (estimation and testing) (6 hours) (F. Van Eeuwijk)**
 - 2.1. Parameter estimation
 - 2.1.1. Confidence intervals
 - 2.1.2. Sample size
 - 2.2. Hypothesis testing
 - 2.2.1. Type I and II errors
 - 2.2.2. Power and sample size
 - 2.2.3. Multiple testing
 - 2.3. ANOVA, regression and covariance analysis
3. **Data generation and preparation (8 hours) (M. Vuylsteke)**
 - 3.1. Marker data
 - 3.2. Metabolic data
 - 3.3. Gene expression data
 - 3.3.1. Microarrays
 - 3.3.2. Design of two-colour microarray experiments
 - 3.3.3. Reading microarray data
 - 3.3.4. Exploration of microarray data
 - 3.3.5. Calculation of microarray data
 - 3.3.6. Normalisation of microarray data
 - 3.3.7. Analysis of microarray data
 - 3.3.8. Display of microarray results
 - 3.3.9. Two-channel microarray examples
 - 3.3.10. Affymetrix microarray examples
 - 3.4. Phenotypic data
4. **Mixed models (12 hours) (H. P. Piepho)**
 - 4.1. Variance components models. Estimation and testing procedures for variance components
 - 4.2. Mixed models
 - 4.2.1. Motivation
 - 4.2.2. Fixed and random terms
 - 4.2.3. Estimation of fixed and random effects
 - 4.2.4. Testing of fixed and random terms, model construction
 - 4.3. Examples of mixed models
 - 4.3.1. Models with subsampling
 - 4.3.2. Models for block designs
 - 4.3.3. Split plot models
 - 4.3.4. Repeated measurements models
 - 4.3.5. Spatial models
5. **Genetic map construction from molecular markers (10 hours) (H. Jansen)**
 - 5.1. Recombination and linkage
 - 5.2. Mapping functions
 - 5.3. Map construction algorithms
 - 5.4. Pre- and post map diagnostics
 - 5.5. Map integration

- 6. Dimension reduction, prediction and classification (8 hours) (J. McNicol)**
 - 6.1. Distance and similarity measures
 - 6.2. Dimension reduction techniques (PCA, PCO, MDS)
 - 6.3. Discriminant analysis
 - 6.4. Hierarchical cluster analysis (UPGMA, Neighbour Joining)
 - 6.5. Non-hierarchical cluster analysis (K-means clustering)
 - 6.6. Multivariate regression techniques
 - 6.7. Random forests
 - 6.8. Classification and regression trees
 - 6.9. Support vector machines
 - 6.10. Elastic nets
- 7. Diversity analysis (8 hours) (C. Billot)**
 - 7.1. Distance and similarity measures
 - 7.1.1. Genetic
 - 7.1.2. Non-genetic
 - 7.2. Identifying groups (statistics)
 - 7.2.1. Scaling methods
 - 7.2.2. Clustering
 - 7.3. Population genetic subdivision
 - 7.3.1. Wright's F-statistics
 - 7.3.2. STRUCTURE: model based Bayesian subdivision
 - 7.4. Alternative ways for identifying population genetic structure
- 8. QTL mapping, association mapping, genetical genomics (20 hours) (F. Van Eeuwijk & M. Malosetti)**
 - 8.1. QTL mapping (regression and mixed models)
 - 8.1.1. Marker regression
 - 8.1.2. Simple interval mapping
 - 8.1.3. Composite interval mapping
 - 8.1.4. Multiple QTL models
 - 8.1.5. Multiple trait QTL mapping
 - 8.2. Association mapping
 - 8.2.1. Linkage disequilibrium between marker loci
 - 8.2.2. Linkage disequilibrium decay
 - 8.2.3. Association mapping by mixed models
 - 8.3. Mapping of gene-expression and metabolite data