



HAL
open science

New phenotyping, genotyping and statistical tools for forage breeding

Bernadette Julier, Philippe Barre, Marie Pégard

► **To cite this version:**

Bernadette Julier, Philippe Barre, Marie Pégard. New phenotyping, genotyping and statistical tools for forage breeding. 35. EUCARPIA Fodder Crops and Amenity Grasses section conference, Sep 2023, Brno, Czech Republic. <hal-04172974>

HAL Id: hal-04172974

<https://hal.inrae.fr/hal-04172974v1>

Submitted on 28 Jul 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

New phenotyping, genotyping and statistical tools for forage breeding

Bernadette Julier, Philippe Barre, Marie Pégard

INRAE, URP3F, 866000 Lusignan, France, bernadette.julier@inrae.fr

Various high throughput technologies and the statistical methods needed to use them are now available for forage breeding. From testing in research programmes to implementation in breeding schemes, their implementation is gradually being extended. These technologies, in addition to requiring new equipment, also ask for skills in data management and statistical analyses.

New phenotyping tools

Throughout the last 50 years of forage breeding, new phenotyping tools have been set up to characterise forage yield and various quality components. Resistance to pests and diseases and to abiotic stresses have also been studied as a component of yield and yield stability over time and conditions. The list of traits and protocols, both in field or controlled conditions, continues to grow. In the development of these tools, cost-effectiveness has been and remains a constant concern. One of the most obvious examples of the last decade has seen major developments in image capture and analysis. By using unmanned aerial vehicles or Lidar-type machinery, non-destructive images are recorded. Various cameras (RGB in visible light, multispectral, etc) provide different views of the experiments. Images can be recorded several times, giving access to the dynamics of the process, whether in terms of yield development or plant height, drought onset, etc. Several publications on forage grasses and legumes, including at this conference, have demonstrated the potential of this tool. In short, after image capture, image analysis is carried out to extract plant traits such as height, biomass, drought tolerance, soil cover, disease attack, etc. A good correlation between observed traits and the traits calculated from images is often observed, even if some side effects can alter the correlation. For example, the occurrence of lodging in lucerne field plots, which is a heterogeneous event in the experiment, reduce the accuracy of yield and height prediction.

New genotyping tools

Understanding and elucidating the genetic mechanisms that underlie the traits of interest is essential in plant breeding. In this respect, the last decade has seen the development of high-throughput, cost-effective genotyping tools. For forage species, genotyping by sequencing (GBS) has been widely adopted in research programmes. In these heterozygous and sometimes polyploidy species, GBS markers can be revealed at the individual level, giving access to allele frequency and sometimes to allele dosage, as well as at the population level for which allele frequency is obtained for pooled leaf samples. Restriction enzymes, used to limit the genome complexity, are chosen so the reading depth matches the sequencing effort and missing data is reduced. With GBS, depending on the studies, 10K to 500K were obtained. Whole genome sequencing has been used in a few number of species, generating millions of SNP. Access to the genome sequence of species makes these new tools all the more useful and accurate. Reference genome sequences have been released for most of the major forage species, offering insights on genome structure, in comparison to model species and/or major (grain) species. For some species, several genomes have been released, offering the opportunity of pangenome analyses.

The reference genome sequences are also of major interest for mapping GBS reads and extracting SNP with their position on the sequence. Comparison of experiments is thus feasible.

Thanks to these genotyping tools which produce a dense SNP coverage across the genome, genome-wide association studies (GWAS) are conducted for a number of traits in dedicated populations with

a high diversity. Genome prediction has been tested and promising predictive abilities have been reported. New insights on genetic diversity have been obtained, renewing the knowledge on genetic resources and their potential use in breeding.

Statistical studies

Geneticists make intensive use of statistics. Analysis of increasing amount of data requires computing efforts for phenotyping, genotyping and analysis of genetic determinism of traits.

Interpretation of images are based on analysis of pixel position and/or colour, with the test of various options, so that traits or trait response to a stress are estimated. Field experiments, when conducted on a large number of genotypes as required in genetic studies, often show uncontrolled variation due to soil heterogeneity. In these conditions, the integration of spatial effects that account for uncontrolled variation benefit to the estimation of corrected genotype values.

In genotyping pipelines, read sequences are mapped on reference genome sequences. Series of trimming, application of thresholds and tests are required to ensure the validity of SNP detection and allele frequency.

GWAS and genomic prediction are domains of innovation where new procedures are proposed. Most of them are integrated in dedicated software or packages but the user has to make the appropriate choice of options depending on the dataset and objectives. The risk of detecting false-positive QTL or over-fitted models must be reduced as much as possible.

What next for breeding?

Studies on new genetic tools have shown promising results for forage crops. The transfer from research to breeding is a current challenge. There are three main difficulties: cost-effectiveness, equipment and skills.

Cost-effectiveness is particularly important for genotyping. A genotyping cost close to that of phenotyping cost is needed to implement genotyping in breeding programmes. In addition, for a breeding company, genotyping method must be licence-free, or outsourced to a licenced service provider. GBS is therefore not appropriate. Genotyping methods based on sequencing of captured fragments or amplicons could be considered but the number of SNP are not as high as that of GBS. Whole genome sequencing does not require a licence but remains expensive.

New equipment is also needed: drone or vehicle, camera, computer, memory, computing capacity to process phenotype and genotype data. In addition, the staff must acquire additional skills and change the way they organise their work.

In the coming years, proof of concept will be required to test the effectiveness of breeding schemes that use all these new tools. This activity is included in the BELIS project (Breeding European Legumes for Increased Sustainability) supported by an Innovation-Action programme of Horizon Europe, which will be launched in October 2023 for five years.

Acknowledgements: EUCLEG (Horizon 2020 of European Union n° 727312) and BELIS (Horizon Europe of European Union n°101081878)