ALFALFA BREEDING WITH IMPLEMENTATION OF MOLECULAR TOOLS

Bernadette Julier[1], Marie Pégard[1], Camille Gréard[2], Philippe Barre[1]

ABSTRACT

Genome-wide, high throughput, cost-effective molecular markers have been developed for less than a decade in alfalfa. Their use for breeding this autotetraploid, heterozygous species is not straightforward but highly promising. We propose to browse through their main uses with an emphasis on results from the European project EUCLEG (www.eucleg.eu) and to highlight the interest of using allele frequencies directly estimated on pools of individuals from a population. A first use is the description of genetic resources. With a large set of markers (>100.000), a continuum among the European – American accessions was evidenced but clearly separated from the Chinese accessions. A second domain is the use of markers to identify genes or locus associated to valuable traits and explaining a substantial part of the variation: QTL. Genome wide association studies on a highly diverse panel of alfalfa populations (varieties, landraces, breeding populations) revealed QTL for yield and quality traits such as protein and ADF contents, explaining up to 15% of the variation. Despite the great advantage of dealing with a high level of diversity, a drawback is that the detection of some QTL could be hampered by the genetic structure within the panel. Another possibility to seek QTL is reverse genetic with allele mining in candidate genes. We used this method to find potentially interesting alleles involved in plant growth or digestibility. A third use is genomic selection based on all available markers to predict the phenotype of an individual from its genotype with a calibration set up on a training population. Within a highly diverse panel and a training set of about 270 populations, we obtained predictive ability ranging from 0.50 to 0.66 for yield, protein and ADF contents measured at two locations for two years. These values are relatively high compare to other studies with less diversity and seem promising. The use of genomic selection within breeder plant material has to be demonstrated but the expected genetic gain per year is huge (more than six times compared to phenotypic selection) especially regarding the decrease of selection cycle duration. We conclude that these different uses of molecular markers could renew alfalfa breeding programs for any trait of current and future interest.

Key Words: Association genetics, genetic diversity, genomic selection, *Medicago sativa*, marker

INTRODUCTION

Molecular markers have been proved to be useful tools to speed up genetic progress in many breeding programs as for dairy cows or maize. Their implementation in alfalfa breeding is lagging behind because of still recent development of genome-wide cost-effective genotyping methodologies and because of the genome complexity of this autotetraploid species. The first complete genomic sequence has been released in 2020 only (Chen et al., 2020). We propose to

[1]Bernadette Julier (bernadette.julier@inrae.fr), Marie Pégard (marie.pegard@inrae.fr), Philippe Barre (philippe.barre@inrae.fr), INRAE, P3F, 86600 Lusignan, France,
[2]Camille Gréard (camille.greard@cerience.fr), Cérience, 86600 Saint-Sauvant, France
In: Proceedings, 2022 World Alfalfa Congress, San Diego, CA, November 14–17. UC Cooperative Extension, Plant Sciences Department, University of California, Davis, CA 95616. (See http://alfalfa.ucdavis.edu for this and other alfalfa conference Proceedings.)

go through three main applications of these now available markers. (1) The description of genetic diversity with molecular markers to manage the genetic diversity within the breeding programs and decipher alfalfa expansion history in the world. (2) The use of markers to identify genes or locus that explain a substantial part of the variation (at least 10%) for quantitative traits. This could be obtained by association studies (correlation between genetic and phenotypic data) and/or by allele mining in candidate genes (reverse genetics). (3) The genomic selection based on all available markers to predict the phenotype of an individual from its genotype with a calibration set up on a training population. We emphasize the advantage of studies carried out at the population level that ease the phenotyping steps while allele frequencies are estimated on pooled DNA. Examples are mainly taken from the results of the European project EUCLEG (www.eucleg.eu) to illustrate the findings.

PANELS OF POPULATIONS AS A KEY MATERIAL

As alfalfa varieties, breeding materials, landraces and wild materials are heterogeneous populations, most studies aiming at analyzing genetic diversity are based on populations. Genotyping at the population level, with pools of individuals, have been successfully established in alfalfa with the GBS (Genotyping By Sequencing) methodology (Julier et al., 2018). This protocol has been optimized with a selection of two restriction enzymes that further reduce genome complexity and limit the number of missing data (Julier et al., 2021). Such a genotyping was cost-effective compared to the genotyping of a minimum of 30 individuals required to represent each population.

In addition to the knowledge obtained on genetic diversity, breeders are looking for molecular tools that could increase the genetic gain. Genome-wide association studies (GWAS) and genomic selection (GS) are dedicated methods to identify markers associated to phenotypic variation and to build prediction equations, respectively. The markers are then used to select individuals that either carry the best alleles and / or have the best genetic prediction. Up to now, specific populations are produced, they are mainly progeny of polycross from chosen parents (Annicchiarico et al., 2015; Li et al., 2015). If well chosen, this panel of individuals directly composes the breeding material on which the selection is applied. However, as most breeding traits are quantitative traits and require repeated measurements, offspring of each progeny must be obtained for phenotypic evaluation in dense swards. Instead, the panel for GWAS and GS can be composed of populations, so that seeds are directly available to conduct phenotypic evaluation in dense swards. Such a panel of populations offers the possibility to directly conduct both diversity and genetic analyses. An additional aspect is to consider the possibility to extend the diversity of the population. When choosing a panel of individuals, the extension of diversity requires new crosses that may be complex to connect to the initial progeny. With a panel of populations, the addition of new accessions is straightforward. If the same genotyping method is applied and if the phenotyping experiments are connected with control populations, the datasets may be joined. In addition, when the panel of populations is large enough, subsets of populations on targeted diversity can be sampled. A drawback could be the population substructure hampering the detection of QTL co-segregating with the kinship. In that case, crosses are needed to obtain linkage disequilibrium only based on physical links and not on substructure.

Below, we are reporting experiments conducted with panels of populations that we hypothesized to be efficient for genetic analyses.

A NEW VIEW OF ALFALFA GENETIC DIVERSITY

On a set of 400 cultivated populations, the GBS markers revealed a significant structure with continuous variation among European and American origins (Figure 1). The accessions from China were clearly different from the Western origins and from the two populations with ssp. *falcata* genetic background (Pégard et al., 2021). The breeding material of five European breeders has been plotted on the graph. Depending on the breeders, the range of diversity they provided for this study was either narrow (Figure 1 B) or more diverse (Figure 1 A) but in no case their material was close to North American nor Chinese diversity. Our aim is now to genotype the whole range of *M. sativa* species complex including natural populations in order to draw the history of alfalfa expansion in the world, from ancient until recent times.

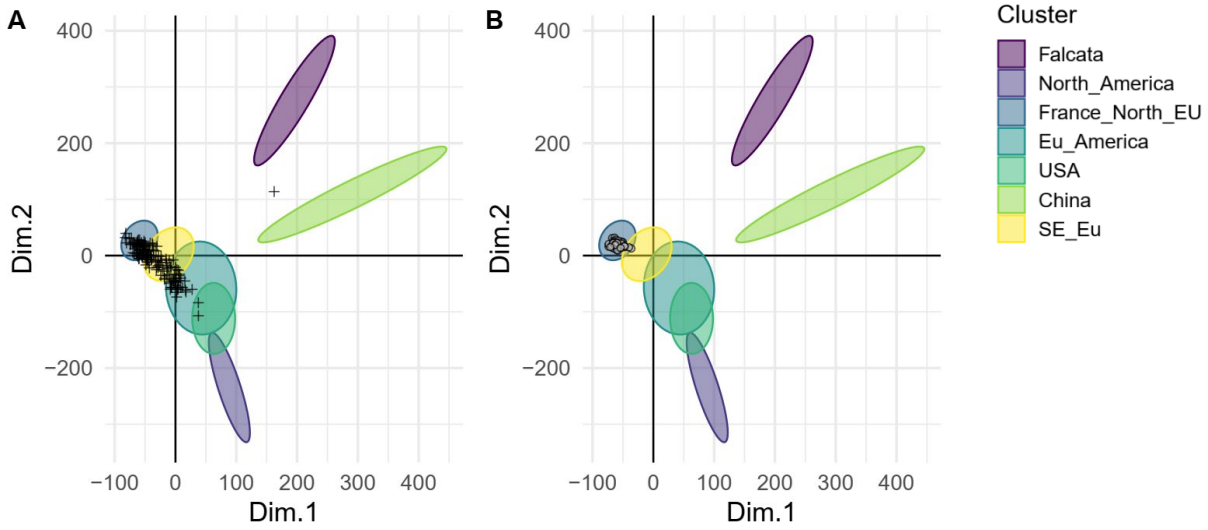

Figure 1. Graphical representation of the first two dimensions of a principal component analysis (PCA) for 400 alfalfa populations genotyped with ~ 100 000 GBS markers. Dim.1 and Dim.2 explained 4.9% and 3.3% of total genetic variation. The ellipses clustered populations after a Discriminant analysis on principal components and the clusters were mainly related to the geographical origins of the populations. Materials of European breeders were plotted on this PCA (black symbols), the cases of two breeders are represented (A) relatively large diversity, (B) narrow diversity of the breeding pools.

MARKERS ASSOCIATED TO TRAIT VARIATION: QTL

The panel of 400 populations was evaluated in field plots in two locations from 2018 to 2021 for forage yield and quality traits. More than 200 000 GBS markers (less than 5% missing data) were used for the analyses. With Genome Wide Association Studies (GWAS), significant QTL were observed for some but not all traits, each explaining up to 15% of the phenotypic variation. This results highlight that a few major loci act at explaining a part of the genetic variation, in addition to many markers with minor effects. The high genome coverage contributed to identify such major loci in this forward genetic strategy.

Another interesting strategy for seeking QTLs is to use previous knowledge in particular from model species in which gene effects on the phenotype have been demonstrated. These genes are perfect candidates for further investigation in other species. A reverse genetic strategy is to mine natural allele diversity for these genes in plant material. Such an approach was conducted in five genes involved in lignin pathway, growth and stress resistance (Gréard et al., 2018). Non-

synonymous variants were detected, more often in wild than in cultivated material. These variants could be used in breeding programs to select promising individuals. The huge advantage of this strategy is that once a variant is identified, it is highly probable that the mutation is the causal one so the linkage between the allele variant and the causal mutation cannot be broken.

GENOMIC SELECTION

The potential of genomic selection can be evaluated from the predictive ability that is the correlation between the true genetic values and the predicted values by using a genomic predictive equation. On our panel composed of 400 populations, with a large training subset of 270 populations, the predictive ability ranged from 0.50 to 0.66 depending on the trait (Table 1). These values of predictive ability were higher than those already published on panels of individuals.

The potential of using predictive equations established on population panels for the selection of individuals was proved to be relevant (Cericola et al., 2018). In these conditions, renewed versions of alfalfa breeding schemes can be conceived, in which the duration of a breeding cycle drops from 10 years in phenotypic selection to 2 years in genomic selection. Considering moderate trait heritability measured in the design and the phenotypic standard deviation, the genetic gain per year could be about 6 times higher in genomic selection than in phenotypic selection (Table 1). This calculation did not take into account the necessity to update the predictive equation which could be done in parallel with a short delay at each cycle.

Table 1. Predictive ability estimated for dry matter yield, ADF and protein contents combined over multiple cuts in Lusignan (France) and Novi Sad (Serbia)

Trait	h^2	Phenotypic standard deviation	Predictive ability	Genetic gain per year with phenotypic selection	Genetic gain per year with genomic selection
Dry matter yield (t/ha)	0.26	1.1	0.58	0.01	0.09
ADF content (%DM)	0.22	39.8	0.50	0.46	2.87
Protein content (%DM)	0.29	40.9	0.66	0.62	3.89

CONCLUSION

Promising results have been obtained with molecular markers in alfalfa. A new overview of genetic diversity offer prospects to better exploit untapped genetic resources in breeding programs in all the regions of the world. Anonymous markers or variants in candidate genes that explain a part of genetic variation can be used to select promising individuals. Genomic prediction, based on a panel of populations, reached high predicting ability. Genetic gain per year is expected to be improved by 6-fold in renewed breeding schemes. This concept still has to be proved, with the measure of achieved genetic grain. Implementation of genomic selection in breeding schemes also requires to optimize its cost efficiency. Two components are important, (1) production of genomic predictive equations and identification of major QTL for all breeding traits, (2) lowering of genotyping costs.

REFERENCES

Annicchiarico P., Nazzicari N., Li X.H., Wei Y.L., Pecetti L., Brummer E.C. (2015) Accuracy of genomic selection for alfalfa biomass yield in different reference populations. BMC Genomics 16. DOI: 10.1186/s12864-015-2212-y.

Cericola F., Lenk I., Fe D., Byrne S., Jensen C.S., Pedersen M.G., Asp T., Jensen J., Janss L. (2018) Optimized Use of Low-Depth Genotyping-by-Sequencing for Genomic Prediction Among Multi-Parental Family Pools and Single Plants in Perennial Ryegrass (Lolium perenne L.). Frontiers in Plant Science 9. DOI: 10.3389/fpls.2018.00369.

Chen H.T., Zeng Y., Yang Y.Z., Huang L.L., Tang B.L., Zhang H., Hao F., Li W., Li Y.H., Liu Y.B., Zhang X.S., Zhang R., Zhang Y.S., Li Y.X., Wang K., He H., Wang Z.K., Fan G.Y., Yang H., Bao A.K., Shang Z.H., Chen J.H., Wang W., Qiu Q. (2020) Allele-aware chromosome-level genome assembly and efficient transgene-free genome editing for the autotetraploid cultivated alfalfa. Nature Communications 11. DOI: 10.1038/s41467-020-16338-x.

Gréard C., Barre P., Flajoulot S., Santoni S., Julier B. (2018) Sequence diversity of five Medicago sativa genes involved in agronomic traits to set up allele mining in breeding. Mol Breed 38. DOI: 10.1007/s11032-018-0898-7.

Julier B., Barre P., Lambroni P., Delaunay S., Thomasset M., Lafaillette F., Gensollen V. (2018) Use of GBS markers to distinguish among lucerne varieties, with comparison to morphological traits. Mol Breed 38:133. DOI: 10.1007/s11032-018-0891-1.

Julier B., Blugeon S., DelaunaynS., Mappa G., Ruttink T., Pégard M., Barre P. (2021) Optimisation of GBS protocols for efficient genotyping of forage species, in: S. Hartmann, et al. (Eds.), 34th Meeting of the Eucarpia Fodder Crops and Amenity Grasses Section, Palacky University Olomouc, online. pp. 71-74.

Li X.H., Wei Y.L., Acharya A., Hansen J.L., Crawford J.L., Viands D.R., Michaud R., Claessens A., Brummer E.C. (2015) Genomic Prediction of Biomass Yield in Two Selection Cycles of a Tetraploid Alfalfa Breeding Population. Plant Genome 8. DOI: 10.3835/plantgenome2014.12.0090.

Pégard M., Barre P., Julier B. (2021) New insights on the genetic structure of lucerne with GBS markers, in: S. Hartmann, et al. (Eds.), 34th Meeting of the Eucarpia Fodder Crops and Amenity Grasses Section, Palacky University Olomouc, online. pp. 17-20.

ACKNOWLEDGEMENTS

This project has received funding from the European Union's Horizon 2020 Program for Research & Innovation under grant agreement n°727312. We thank the seed companies and institutes that provided us with seeds of their accessions (AgriObtentions, Alforex, ART, Barenbrug, Cérience, CGS Sementi, CREA, DLF Seeds, Forage Genetics, Foratec Argentina, IFVCNS, IKBKS, INTA, INRAE-CRG Prairies, Lantmännen, NordGen, RAGT-2n). The staffs of INRAE (URP3F, FERLUS and GeT-PlaGe) and IFVCNS (Djura Karagić, Dragan Milić, Miroslav Zorić) are thanked for their deep investment to collect genotyping and phenotyping data.