HAL open science

Genomic prediction of lucerne forage yield and quality

Marie Pégard, Julien Leuenberger, Bernadette Julier, Philippe Barre

To cite this version:

Marie Pégard, Julien Leuenberger, Bernadette Julier, Philippe Barre. Genomic prediction of lucerne forage yield and quality. Eucarpia - Section Fodder Crops and Amenity Grasses Meeting, Sep 2021, En ligne, Germany. hal-03323297

HAL Id: hal-03323297

https://hal.inrae.fr/hal-03323297v1

Submitted on 20 Aug 2021

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.

Genomic prediction of lucerne forage yield and quality

Marie Pégard[1], Julien Leuenberger[1], Bernadette Julier[1] and Philippe Barre[1]

[1] INRAE P3F, 86600 Lusignan, France

marie.pegard@inrae.fr

Abstract. Genomic prediction has proven its efficiency in numerous animal and plant species. In this study, we used diverse lucerne varieties and populations to test the predicting ability of genomic prediction. Several parameters, such as the number of markers, the population size and the addition of QTL effects, were tested for their effect on the quality of prediction. Based on a large number of SNPs (227 K) obtained by GBS and phenotypes observed in different locations, our results showed a good quality of predicting ability for dry matter yield, ADF (acid detergent fiber) and protein content, especially with a large training population size (around 0.6). The predicting ability is improved by the integration of QTL information directly in the model (above 0.8). A reduction of number of markers (less than 100K) did not alter much the predictive ability. Our results show an accurate prediction of the phenotype of populations *via* genomic prediction models that could speed up the creation of new lucerne varieties.

Keywords: alfalfa, genomic selection, GBS.

1 Introduction

Lucerne (*Medicago sativa* L.) is an important crop worldwide, it plays a major role in farm protein autonomy and also offers multiple ecosystem services. It is a cross-pollinating, auto-tetraploid species, which makes its improvement more complex. Various authors have shown the interest of genomic selection in plant and animal species. Genomic selection can be defined as a prediction of the breeding values based on the genotypic data to select the best individuals for further breeding. Some studies have been published on genomic evaluation of lucerne. They report poor or moderate accuracies (0-0.65) for yield and quality traits, with a number of SNPs from 8K to 44K and a number of individuals between 75 and 274 [1]–[4]. In the present study, we evaluated the ability of genomic prediction models coupled with QTL detected by GWAS (Genome Wide Association Study) to predict the phenotype of diverse genetic resources composed by 400 populations (pools of individuals), genotyped with a large number of markers (227K SNPs).

2 Material and methods

The 400 lucerne accessions comprised cultivars, advanced breeding material and landraces. The accessions were genotyped by genotyping-by-sequencing (GBS) in pools of a hundred individuals and the reads were mapped on the reference genome [5]. After the application of several quality filters, 227K SNP were obtained. The 0.568 % of missing values were imputed by the minor allele frequency. Phenotypic data of the 400 accessions (dry matter yield, acid detergent fiber (ADF) and protein content) were obtained in trials established in 2018 at three locations (Lusignan - FRA, Novi Sad - SRB, , Store Heddinge – DNK) and scored during two years (2019 and 2020). The phenotypes were adjusted for the micro-environmental variation inside each trial, the year and the location effect. Cross-validation was used to assess the prediction quality combined with an independent Test Set representing 20% of the 400 populations (80 populations among the 400). We tested the effect of different population sizes (75%, 50% or 25% of the 320 remaining accessions), and the number of markers on the prediction ability. Several sets of markers corresponding to different marker densities were tested. Based on a window of a given physical size, five sets of markers were designed: 74K, 29K, 12K, 5K and 1.3K SNP. We also tested the influence of integrating QTL information in the model as fixed effects (QGBLUP) compared to the classical GBLUP model. The QTL were detected by Multi-Locus Mixed Model approaches [6]. The predicting ability computed as the Pearson's correlation between the observed phenotype and the predicted value, provided an estimate of the prediction quality.

3 Results

The prediction quality was relatively high: between 0.30 and 0.80 (Fig. 1A). A decrease in the size of the training population had a modest negative impact on the quality of the prediction, whatever the trait.

The GWAS analysis revealed 19 QTL for dry matter yield (explaining between 0.016% and 13.5% of the phenotypic variation), 15 for ADF content (explaining between 0.09 % and 8.6 %) and 15 for protein content (explaining between 0.003% and 15.9 %). The addition of these QTL to the prediction model improved the prediction quality above 0.80 when large training populations were used (Fig. 1A).

The results obtained with the different marker sets showed little impact of the number of markers on the predicting ability (Fig. 1B). The data set with the lowest numbers of markers (1397), however, clearly showed a lower predictive ability than the sets with a higher number of markers. There was a plateau between the set of 5,794 markers (1 marker every 100,000 bases) and the set of 74,175 markers (1 marker every 50 bases). A limited number of markers seems to be sufficient to reach a good quality of prediction in this present dataset and in these conditions (less than 100K). However, the large number of markers was requested to detect the QTL that increased the quality of prediction. We can imagining to use a large number of markers to detect the QTL and then combine them with a small set of markers for the rou-

tine prediction. This would accelerate the improvement of lucerne and facilitate the creation of new cultivars. The next step will be to see if these prediction equations accurately predict the current elite material of breeders.

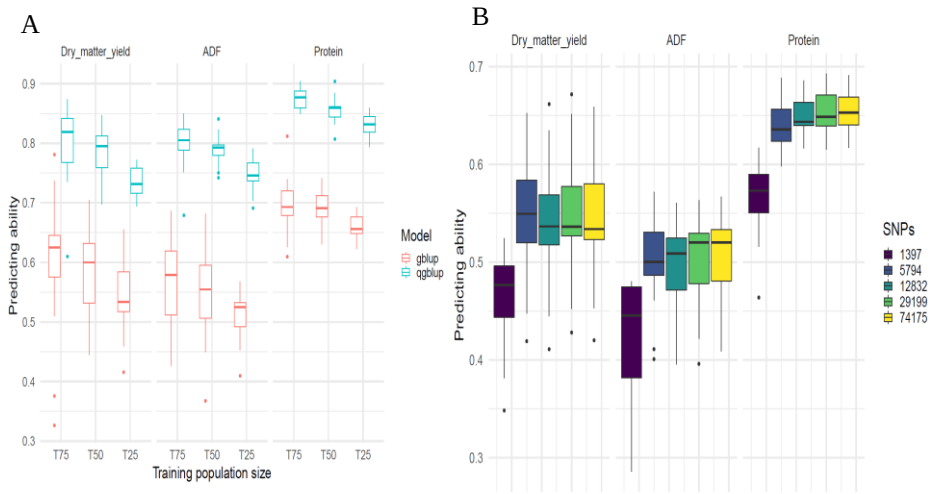

Fig. 1. Predicting of the Test Set (80 accessions) for three traits (dry matter yield, ADF and protein content). **A** : Predicting ability depending on the size of the training population. Two models are presented; the GBLUP in pink and the QGBLUP in blue where QTL are integrated in the model as fixed effects. **B** : Predicting ability of the T50 training population size depending on the number of markers used in the model.

References

[1] P. Annicchiarico, N. Nazzicari, X. Li, Y. Wei, L. Pecetti, and E. C. Brummer, 'Accuracy of genomic selection for alfalfa biomass yield in different reference populations', *BMC Genomics*, vol. 16, no. 1, p. 1020, Dec. 2015, doi: 10/f8ddkp.

[2] E. Biazzi *et al.*, 'Genome-Wide Association Mapping and Genomic Selection for Alfalfa (Medicago sativa) Forage Quality Traits', *PLoS ONE*, vol. 12, no. 1, p. e0169234, Jan. 2017, doi: 10/f9kvzb.

[3] C. Jia *et al.*, 'Genomic Prediction for 25 Agronomic and Quality Traits in Alfalfa (Medicago sativa)', *Front. Plant Sci.*, vol. 9, 2018, doi: 10/gjjz78.

[4] X. Li *et al.*, 'Genomic Prediction of Biomass Yield in Two Selection Cycles of a Tetraploid Alfalfa Breeding Population', *Plant Genome*, vol. 8, no. 2, Jul. 2015, doi: 10/ggjp3c.

[5] H. Chen *et al.*, 'Allele-aware chromosome-level genome assembly and efficient transgene-free genome editing for the autotetraploid cultivated alfalfa', *Nature communications*, vol. 11, no. 1, pp. 1–11, 2020.

[6] V. Segura *et al.*, 'An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations', *Nature genetics*, vol. 44, no. 7, p. 825, 2012, doi: 10/f34dkw.