

Genome-enabled and phenotypic selection of alfalfa for widely-diversified cropping environments

Annicchiarico, P.*, Nazzicari, N., Bouizgaren, A., Hayek, T., Laouar, M., Cornacchione, M., Basigalup, D., Brummer, E.C., Pecetti, L.

Council for Agricultural Research and Economics (CREA), Centre for Animal Production and Aquaculture, Lodi, Italy
* e-mail: paolo.annicchiarico@crea.gov.it

Keywords: adaptation, drought tolerance, genomic selection, genotyping-by-sequencing, salt tolerance

Introduction

The decrease of rainfall amount due to climate change, and the reduction of irrigation water caused by growing water demand for non-agricultural uses, emphasize the importance of breeding novel alfalfa varieties that are more tolerant to drought under rain-fed cropping or modest supplemental irrigation, and more tolerant to salt to exploit low-quality, saline irrigation water. A sharp need for such improved germplasm is emerging in the Mediterranean basin, where alfalfa plays a key role for forage production. Drought-tolerant cultivars are needed also elsewhere, for example in Argentina, where they could allow to expand westward the alfalfa rain-fed cropping. Alfalfa typically features very low rates of genetic yield gain (Annicchiarico et al. 2015a), urging the development of cost-efficient marker-based selection.

The multi-site yield testing of alfalfa varieties and landraces across countries of the western Mediterranean basin revealed outstanding genotype \times environment (GE) interaction of cross-over type associated with three major types of target environments: i) rain-fed or irrigated environments featuring limited spring-summer water available and low salinity, ii) salt-stress environments, and iii) moisture-favorable environments (Annicchiarico et al. 2011). Eco-physiological research identified various mechanisms that contribute to specific-adaptation responses (Annicchiarico et al. 2013). Moderately wide cultivar adaptation may be desirable, given the wide year-to-year climatic variation of sites in this region.

A Mediterranean reference population of alfalfa was developed from elite germplasm within the ERA-Net project REFORMA. A genotype training set was sorted out from it, to verify the ability of genomic selection (Heffner et al. 2009) to predict breeding values for biomass yield across a range of widely-diversified cropping environments. Genotype breeding values were estimated according to responses of their half-sib progenies, as required by the crop outbred system (Annicchiarico et al. 2015a). The genotyping of parent genotypes was carried out by Genotyping-by-Sequencing (Elshire et al. 2011) after optimizing for alfalfa some elements of its protocol (Annicchiarico et al. 2017). Concurrently, the project verified the ability of managed-stress environments of Italy to predict genotype yield responses in distant agricultural environments, following earlier work showing good ability to predict cultivar responses across agricultural environments of Italy (Annicchiarico and Piano 2005). This study aimed to provide an initial assessment of the ability of genomic selection to predict breeding values in each cropping environment and across stress environments, on the basis of preliminary yield data and GBS-generated data.

Materials and Methods

The Mediterranean reference population was developed by repeated intercrossing of genotypes from the drought-tolerant Sardinian landrace Mamuntanas, the salt-tolerant Moroccan landrace Erfoud 1 and the Australian variety SARDI 10 [which is widely adapted across moisture-favorable and drought-prone sites: Annicchiarico et al. 2011], as described in Annicchiarico et al. (2015b). Some 128 half-sib progenies were simultaneously phenotyped in three managed-stress (MS) environments of northern Italy that featured contrasting drought-stress level, two drought-prone rain-fed sites of Algeria (Alger) and western Argentina (Santiago del Estero), one drought-prone Moroccan site managed with limited supplemental irrigation (Marrakech), and a Tunisian site (Médénine) irrigated with moderately saline (9.37 dS/m) water (Figure 1). Dry-matter yield was recorded over five months in the moisture-favorable MS environment, and over a period ranging from nine months (Santiago del Estero) to two years (intense-drought MS) in the other environments. Data from one more year of evaluation will be available in the future for most sites.

Annicchiarico et al. (2015b) described the GBS analysis of the parent genotypes (which used the *ApeKI* restriction enzyme and the KAPA *Taq* polymerase) and the genotype SNP calling, which envisaged two homozygote classes AAAA and aaaa and one heterozygote class pooling AAAa, AAaa and Aaaa variants (given the inability to produce sufficient SNP data to estimate reliably the allele dosage for heterozygous loci). We retained for analyses 9,269 polymorphic SNP markers with less than 20% missing data across genotypes, imputing missing data by the KNNI method (K = 4). Genome-enabled predictions were assessed by two models, i.e., Ridge Regression BLUP and Support Vector Regression with linear kernel, which stood forward in the comparison of models reported in Annicchiarico et al. (2015b).

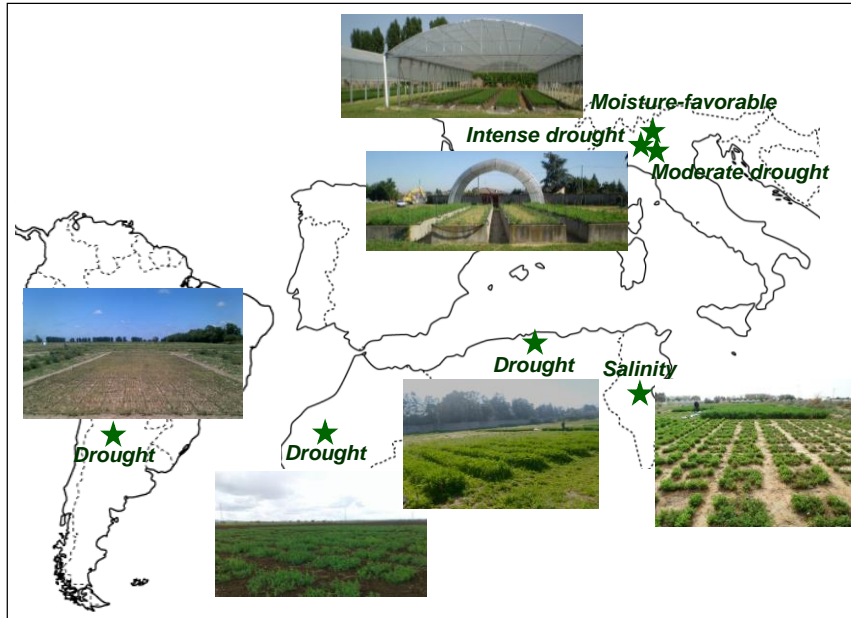


Figure 1. Test environments for biomass yield phenotyping of 128 alfalfa half-sib progenies

Results

Estimates of genetic correlation for yield of half-sib progenies between pairs of environments were mostly low and occasionally negative, indicating the large size of GE interaction even across drought-prone environments. The ability of MS environments of Italy to predict yield responses in distant agricultural environments was generally modest, probably because of differences for other environmental factors (e.g., temperature pattern). High genetic correlation occurred between MS environments with moderate drought and moisture-favorable conditions. The large size of GE interactions was confirmed by cross-over interactions occurring among top-yielding half-sib progenies as revealed by AMMI analysis.

The two genomic selection models exhibited similar predictive ability (data not shown). Environment-specific genomic predictions, which are reported in Table 1 for the best-predicting model, were fairly high for the moisture-favorable MS environment, and poor for the stress-prone sites of Alger, Medénine and Santiago del Estero. These sites will produce at least one more year of data in the future, and the possibility to achieve better genomic predictions using yield data over a longer crop cycle was supported by higher predictive abilities obtained for data of the last harvest in Alger and Medénine (Table 1).

Table 1. Predictive ability of genomic selection for alfalfa breeding values of biomass yield in managed-stress (MS) environments of Italy and agricultural environments (as correlation of modelled and observed data averaged across 100 10-fold stratified cross-validations, for the best-predictive of two models)

Test environment	Total yield	Last harvest yield ^a
MS, intense drought	0.12	0.22
MS, moderate drought	0.18	0.10
MS, moisture-favorable	0.36	.
Alger (Algeria)	0.05	0.12
Marrakech (Morocco)	0.19	0.23
Médenine (Tunisia)	0.05	0.13
Santiago del Estero (Argentina)	0.10	.

^a Excluding environments with relatively short experiment duration

One possible reason for less accurate predictions in stress environments was their trend towards higher experiment error than favorable environments, as shown by error coefficients of variations for intense drought ($CV_e = 27.9\%$), moderate drought ($CV_e = 21.2\%$) and moisture-favorable ($CV_e = 13.9\%$) MS environments. One additional reason might be the fact that the reference population derived from stress-tolerant material and, as such, was characterized by only moderate genetic variation for stress tolerance, as suggested by greater genetic coefficient of variation under moderate drought ($CV_g = 15.7\%$) than intense drought ($CV_g = 10.3\%$). However, the adopted genetic base was that of practical regional interest.

The value of genomic predictions ought to be assessed in relation to opportunities offered by phenotypic selection. A comparison of genomic vs half-sib progeny-based phenotypic selection in terms of predicted genetic gain per year equates to comparing ($i_G r_A / t_G$) vs ($i_P H / t_P$), where r_A and H stand for genome-enabled predictive ability and the square root of narrow-sense heritability, respectively, and i and t are the standardized selection differential and the selection cycle duration in years (including final intermating), respectively, for genomic (G) and a phenotypic (P) selection (Annicchiarico et al. 2017). Since H lies in the range 0.39-0.45 based on narrow-sense heritability values summarized in Annicchiarico (2015), assuming $H = 0.47$, $r_A = 0.12$, $t_G = 1$, $t_P = 4$ and same evaluation costs implies about same efficiency of genomic and phenotypic selection. However, considering also the seemingly 7-8 times lower cost per genotype of genomic selection relative to half-sib progeny-based selection (Annicchiarico et al. 2017) implies, for the hypothetical scenario of 1400 genomically-evaluated and 200 phenotypically evaluated individuals and 20 selected parents ($i_G = 2.54$; $i_P = 1.75$), nearly 50% greater efficiency of genomic over phenotypic selection under the assumption of $r_A = 0.12$. Hence, most r_A values reported in Table 1 should not be seen as unfavorable.

An additional assessment focused on genomic selection ability to predict alfalfa breeding values for total yield averaged across all stress environments (i.e., all those in Table 1 except moisture-favorable MS). Phenotypic data were previously standardized to relative yields within environments. We obtained a prediction accuracy of 0.14, which is not unfavorable when considering that predictions related to breeding values for wide adaptation across quite diverse environments. In this case, half sib progeny-based phenotypic selection would have quite low H value (owing to the outstanding GE interaction), and its cost would be much greater compared with a single-environment phenotypic selection scenario.

Conclusions

Alfalfa breeding for the western Mediterranean basin is hindered by outstanding GE interaction that sets a limit to selection for wide adaptation and can hardly be coped with by selecting in MS environments of distant regions. Our preliminary results indicate that genomic selection tends to be more predictive and convenient for favorable or moderately favorable environments than stress-prone ones. In these latter environments, however, also phenotypic selection may lack accuracy because of high experiment error. Genomic selection is particularly promising for

selecting resilient varieties with relatively wide adaptation. For the same material, genomic selection proved valuable for key forage quality traits (Biazzi et al. 2017).

Acknowledgement

This research was carried out in the framework of the EraNet-ARIMNet project 'Resilient, water- and energy-efficient forage and feed crops for Mediterranean agricultural systems (REFORMA)'.

Bibliography

- Annicchiarico, P. 2015. Alfalfa forage yield and leaf/stem ratio: narrow-sense heritability, genetic correlation, and parent selection procedures. *Euphytica* 205: 409-420.
- Annicchiarico, P. and Piano, E. 2005. Use of artificial environments to reproduce and exploit genotype \times location interaction for lucerne in northern Italy. *Theor. Appl. Genet.* 110: 219-227.
- Annicchiarico, P., Pecetti, L. and Tava, A. 2013. Physiological and morphological traits associated with adaptation of lucerne (*Medicago sativa* L.) to severely drought-stressed and to irrigated environments. *Ann. Appl. Biol.* 162: 27-40.
- Annicchiarico, P., Barrett, B., Brummer, E.C., Julier, B. and Marshall, A.H. 2015a. Achievements and challenges in improving temperate perennial forage legumes. *Crit. Rev. Plant Sci.* 34: 327-380.
- Annicchiarico, P., Nazzicari, N., Li, X., Wei, Y., Pecetti, L. and Brummer, E.C. 2015b. Accuracy of genomic selection for alfalfa biomass yield in different reference populations. *BMC Genomics* 16: 1020.
- Annicchiarico, P., Nazzicari, N., Wei, Y., Pecetti, L. and Brummer, E.C. 2017. Genotyping-by-sequencing and its exploitation for forage and cool-season grain legume breeding. *Frontiers in Plant Science* 8: 679.
- Annicchiarico, P., Pecetti, L., Abdelguerfi, A., Bouizgaren, A., Carroni, A.M., Hayek, T., M'Hammadi Bouzina, M. and Mezni, M. 2011. Adaptation of landrace and variety germplasm and selection strategies for lucerne in the Mediterranean basin. *Field Crops Res.* 120: 283-291.
- Biazzi, E., Nazzicari, N., Pecetti, L., Brummer, E.C., Palmonari, A., Tava, A. and Annicchiarico, P. 2017. Genome-wide association mapping and genomic selection for alfalfa (*Medicago sativa*) forage quality traits. *PLoS ONE* 12: e0169234.
- Elshire, R.G., Glaubits, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S. and Mitchell, S.E. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* 6: e19379.
- Heffner, E.L., Sorrells, M.E., and Jannink, J.L. 2009. Genomic selection for crop improvement. *Crop Sci.* 49: 1-12.