Summary:

Objectives, Description, Main Results & Conclusions

Growing plant-based eating habits boost legume popularity. Legumes offer rich nutrition and health benefits, having a functional food status. However, nutritional compositional data on some promising legume species is limited, hindering their use. Lathyrus sativus (grass pea) is one such understudied species, being an important source of protein and calories in drier areas, with ability to respond to climate challenges (Gonçalves et al., 2022). In this work, untargeted nuclear magnetic resonance (NMR) spectroscopy profiling was used for a broad and unbiased assessment of seed flour metabolite changes of a collection of 200 grass pea accessions, field trialled over three years. NMR assessment of this collection, for which significant genotype-by-sequence information (5,651 SNPs) exist (Sampaio et al., 2021), allowed also to identify the main genomic regions controlling its metabolome variability, through genome wide association study (GWAS). For that, NMR spectra were consolidated into 514 NMR buckets/metabolome features (NMRprocflow, MetaboAnalyst) and their relative concentrations adjusted means calculated using linear mixed models.

From the metabolomic profiles multivariate analysis, a clear year differentiation was observed. 376 NMR features showed genotypic differences, reinforcing their potential for grass pea improvement, and justifying their GWAS. Using a threshold of −log10(p)=4, 435 SNPs associated with 204 NMR features were identified, prioritizing their annotation. Annotation is ongoing by matching spectral information from 1D and 2D spectra to reference compounds in available databases. Amongst identified compounds are amino acids and derivatives, sugars, sugar-alcohols and organic acids.

The combined genetic analysis highlighted multiple loci impacting compounds abundance through a network of interactions, where individual loci may affect more than one compound and vice versa.

The generated knowledge on metabolite variation and respective genetic basis will support precision breeding efforts to answer consumers' nutritional quality concerns, increasing the use of grass pea in diets.

Bibliography

● *Gonçalves et al (2022) http://doi.org/10.3390/agronomy12061324*

● *Sampaio et al (2021) https://doi.org/10.1002/tpg2.20154*

606/215. GENOMIC SELECTION FOR ALFALFA: THE CHALLENGE OF THE AUTOTETRAPLOID GENOME

Authors:

Nelson Nazzicari[1], Luciano Pecetti[2], Nicolo Franguelli[2], Barbara Ferrari[2], Paolo Annicchiarico[2]

Work centre:

(1) Italy, (2) CREA-ZA

Summary:

Objectives

Alfalfa, the most economically important forage legume worldwide, features modest genetic progress due to long selection cycles and high extent of non-additive genetic variance associated with its autotetraploid genome. Genomic selection (GS) has a special interest for this crop, yielding the promise of substantially shorter half-sib progeny-based selection cycles of parent genotypes for synthetic varieties. This prospect implies GS model construction by a training population whose genotyped parents are phenotyped on the ground of their half-sib progenies. However, the autotetraploid genome complicates also GS predictions because of the substantial sequencing effort and associated costs needed for a reliable allele dosage estimation. Moreover, several strategies do exist for SNP marker calling, filtering, and resulting genome representation, with no clear-cut choice for optimal configurations. This study aims to investigates the effects of such choices and to provide a guideline for both the genotyping phase and the bioinformatics analysis linked to GS in alfalfa. We used Genotyping by Sequencing-generated data and focused on traits of different genetic complexity, i.e., dry biomass yield in moisture-favorable (FE) and drought stress (SE) environments, leaf size, and onset of flowering, assessed on 143 half-sib progenies of a genetically-broad European reference population.

Concise description of the work (materials & methods)

We tested three genome representations: 1) proper tetraploid dosage (AAAA, AAAa, AAaa, Aaaa, aaaa); 2) pooled diploid dosage (with the three heterozygote classes Aaaa, AAaa and AAAa pooled in the single Aa class, together with AA and aa for homozygotes); and 3) allele ratios, where each marker is represented as the observed ratio between A alleles and the total of A+a alleles. These genetic configurations highly depend on other filtering parameters such as the allowed pre-imputation rate of missing marker (mpm) and minimum required number of reads per genetic locus. To compare the various

scenarios, we envisaged 10-folds cross-validated predictive ability as the main performance metric. All regressions used ridge regression BLUP as regression model.

Main Results

The number of resulting SNP markers varied from 2387 to 19668 depending on the configuration. Allele ratios maximized the average GS predictive ability (0.206), followed by tetraploid dosage (0.183) and diploid dosage (0.180). The predictive ability of biomass yield decreased markedly in the stress environment (averaging 0.124 in SE vs. 0.349 in FE) along with a reduced broad-sense heritability (0.34 vs. 0.54), in agreement with earlier studies. Predictive abilities averaged 0.335 for leaf size and 0.170 for onset of flowering, the latter possibly hindered by narrow phenotypic variation and genetic control. Filtering on missing rate favoured higher thresholds, while no clear pattern was found for the minimum number of reads. A second result of this study was the release of Legpipe2, a SNP calling pipeline that was used to produce the markers used in the analyses. Legpipe2 is open source, python3-based, modular, easily customizable and supports orthogonal logging and single-file configuration. It is available at https://github.com/ne1s0n/legpipe2.
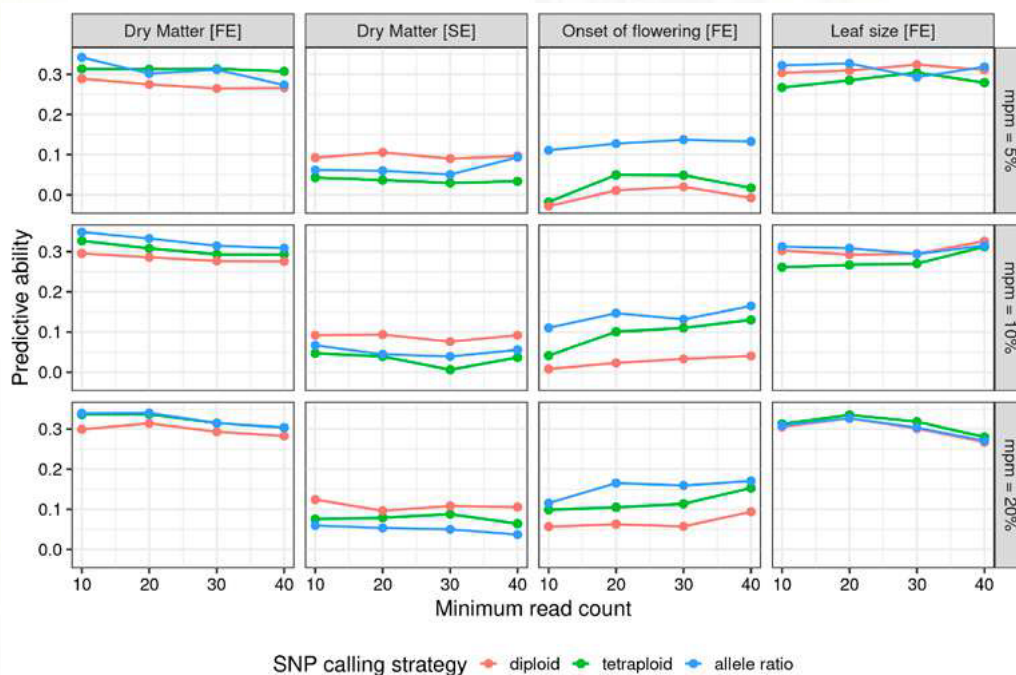

Figure 1. Predictive abilities for three traits in favourable (FE) or drought stress (SE) environment, three levels of filtering on maximum allowed missing rate per marker (mpm), four levels of minimum required number of reads per SNP data point and three SNP representation strategies.

Conclusions

This study shows that SNP ratios is a valid and simple data representation strategy that can improve the GS predictive ability in tetraploid alfalfa by circumventing the difficulties of polyploid dosage calling. It also confirms the greater challenge of predicting biomass yield breeding values in stress environments.

Bibliography

- *Annicchiarico, P., Nazzicari N., Pecetti L.. 2021. Genomic selection for higher yield and quality in alfalfa. In: L.X. Yu, C. Kole (eds), The alfalfa genome. Springer, Cham, Switzerland, p. 191-210.*
- *Nazzicari, N. and Biscarini, F., 2022. Stacked kinship CNN vs. GBLUP for genomic predictions of additive and complex continuous phenotypes. Scientific Reports, 12(1), p.19889.*

606/234. GENETIC DISSECTION OF SEED PROTEIN CONCENTRATION IN PEA USING MULTIPLE DIVERSE MAPPING POPULATIONS

Authors:

Krishna Kishore Gali[1], Ambuj Jha[2], Junsheng Zhou[2], Vijay Reddy Lachagari[3], Bunyamin Tar'an[4], Judith Burstin[5], Grégoire Aubert[5], Dengjin Bing[6], Gene Arganosa[2], Tom Warketin[2]