#047

Evaluation of transcriptome and DNA methylation data for the prediction of hybrid performance in oilseed rape.

Stefan Scholten
Felix Seifert, Susanne Edelmann, Christian Werner, Christian Rockmann, Heike Pospisil, Rod Snowdon, Björn Usadel, Amine Abbadi, Gunhild Leckband

Department of Crop Sciences, Göttingen, Germany

Accurate prediction of hybrid performance on the basis of parental information is crucial for guiding the efficient selection of superior genotypes in oilseed rape (Brassica napus L.) hybrid breeding programs. While predictive ability for genetic data was demonstrated using the Brassica 60 k genotyping array, transcriptome and epigenetic data are expected to complement the characterization of parental genotypes by genomic data and capturing epistatic interactions, thus providing potential to enhance prediction accuracy.

To test the predictive potential of these data types, we generated quantitative messenger RNA (mRNA) and small RNA (sRNA) data by next generation sequencing and quantitative DNA methylation data by reduced representation bisulfite sequencing from young seedlings of the parental lines of two breeding populations. The breeding populations consist of unbalanced crosses of 8 and 4 male-sterile maternal lines and 174 and 200 paternal doubled haploid (DH) lines, resulting in 311 and 437 hybrids, respectively. The hybrids were evaluated for hybrid performance of yield and additional agronomical traits with various genetic architecture in multi-year and multi-environment field trials.

We employed a binomial distribution model to identify genetic, transcriptomic and epigenetic features associated with low and high hybrid performance. Associated features were used to construct distance measures between the inbred lines. These distances and the performance data of the hybrids were then used to estimate regression parameters with a linear regression model. To evaluate and compare the various data types for their predictive ability we employed a Type-1 cross validation scheme with all maternal lines and 25% of the paternal DH lines of the respective population as training set.

Our results demonstrate predictive potential for all three data types. Especially, mRNA and sRNA showed consistently higher prediction accuracies for hybrid performance of yield compared to genetic data while the gain in prediction accuracy varied dependent on the trait and the population under investigation.