A new 7K SNP genotyping array for oat

Andreas Polley[1], Joerg Plieske[1], Eva Grafarend-Belau[1], Outi Manninen[2], Merjia Vetelainen[2], Alf Ceplitis[3], Pernilla Vallenback[3], Muath Alsheikh[4], Heike Gnad[1], and Martin W. Ganal[1]

[1]SGS INSTITUT FRESENIUS GmbH TraitGenetics Section, Am Schwabeplan 1b, 06466 Seeland/Gatersleben, Germany
[2]Boreal Plant Breeding Ltd., Myllytie 10, 31600 Jokioinen, Finland
[3]Lantmännen Lantbruk SE, Udda Lundqvists Väg 11, 26881 Svalöv, Sweden
[4]Graminor Breeding A/S, Hommelstadvegen 60, 2322 Ridabu, Norway

Contact: Martin Ganal, Martin.Ganal@sgs.com

Summary:
A new 7K SNP (Single Nucleotide Polymorphism) oat genotyping array has been developed based on markers from the previously published oat 6K array and markers identified through transcriptome sequencing. This array contains 6,631 unique functional markers that cover the entire oat genome. As far as possible, these markers have been located on the oat reference genome. This 7K array provides a new resource for genetic relationship analysis, genome mapping, association studies, and Genomic Selection through SNP genotyping.

Introduction:
SNP genotyping is now the standard for genetic analyses in crop plants. For hexaploid oat (*Avena sativa*), a first SNP genotyping array was published by Tinker, *et al.* (2014). This 6K array contained 5,743 SNP markers, of which a considerable number did not work or did not produce distinct clusters. This was not unexpected, since the scoring of markers in hexaploid plant species in not easy, especially when more than one genome is detected with an assay (Ganal, *et al.*, 2019). Approximately 3,500 markers were reported to be functional in the investigated germplasm (Tinker, *et al.*, 2014). This low number of markers has made it necessary to increase the number of useful and polymorphic oat SNP markers for routine genotyping in genetic analyses and Genomic Selection within oat breeding.

Experimental approach:
At the time of the development of our 7K array, the oat genome was not fully sequenced. Thus, we generated transcriptome data from seedlings based on a set of twelve oat lines provided by Boreal, Lantmännen, and Graminor. The obtained sequences were assembled for each line and subsequently screened for SNPs using bioinformatic tools (CLC Genomics Workbench and in-house generated software). 10,850 identified candidate SNPs were subsequently screened within a large panel of oat lines (1,584 samples including controls) using a newly designed Illumina SNP array. Furthermore, we added 3,250 functional SNP markers from the published 6K array to this Testing Array on the Illumina Infinium HTS platform. The 14,100 markers on the Testing Array analyzed with this oat panel were individually scored for their polymorphism and, more

importantly, for cluster separation. Once this analysis had been completed, 7,102 markers were selected for our new 7K Illumina Infinium SNP array on the 96-sample XT platform for cost-efficient routine genotyping (Genotyping Array).

Results:
For the 10,850 new candidate SNPs identified from transcriptome sequencing, only 4,004 markers (37%) were found to fulfill our quality criteria regarding cluster separation and minor allele frequency (at least more than 2 lines homozygous with the opposite allele). The main reasons for this low success rate are that many of the candidate markers from the transcriptome sequencing were (1) markers that were only polymorphic between the three genomes (i.e., all samples look heterozygous), (2) did not produce well-separated clusters, probably due to additional SNPs in the assay sequence, (3) had very tight clusters where the separation between the three allelic states was not clear, (4) were represented in only a single line, and (5) were SNPs close to introns that were not accurately identified in the bioinformatic analysis. A dropout rate of around 50% or more was reported for the oat 6K array, as well as for the hexaploid wheat 90K array (Wang, *et al.*, 2014). As expected, from the 3,250 functional oat markers that had been selected from the 6K oat array, the vast majority (3,097 markers, or more than 95%) produced good data. After accounting for loss of markers during array design, 7,102 functional and polymorphic markers were included on our final Genotyping Array.

The new 7K Genotyping Array with the 7,102 selected markers was subsequently tested on a set of oat lines for functionality and polymorphism. Taking into account marker dropouts (that did not produce any signal due to failure in array manufacturing) and markers that did not show three clearly defined clusters, 6,631 markers were classified as functional and useful (>93%). A list of the marker sequences is available from martin.ganal@sgs.com, until the information concerning the array is fully published. Recently, a new 3K oat array was announced by Jason Fiedler at USDA-ARS in Fargo, ND, USA. This 3K array contains approximately 1,000 markers from the 6K array (Tinker, *et al.*, 2014). Thus, there is a significant overlap between the 7K array and the 3K USDA-ARS array.

Based on the cluster distribution results, the vast majority of the markers (approx. 85%) detected more than one genome, and fewer than 1,000 markers were classified as genome-specific. Of the high-quality markers, more than 6,000 markers were polymorphic in the tested oat material from the three European breeding companies. Concerning minor allele frequency (MAF), the markers from the transcriptome sequencing had a somewhat higher minor allele frequency (approx. 38% with MAF>0.1) than the 6K oat array markers (approx. 20% with MAF>0.1), which is probably due to the fact that the 6K oat markers were derived from a different SNP data source (mainly North American germplasm).

Meanwhile, the hexaploid oat genome has been sequenced (https://wheat.pw.usda.gov/jb?data=/ggds/oat-ot3098v2-pepsico; Kamal, *et al.*, 2022). In order to increase their usefulness, we have used the flanking sequences of the 6,631 good SNP markers to locate them on one of the recent versions of the oat genome (PepsiCo version 2). 3,479 markers could be assigned to single genome positions in the oat genome. 2,225 markers

could be assigned to two or three genome positions (mostly homoeologous positions), and the remainder could not be assigned or were assigned to more than three positions. The highest percentage of mapped markers was on the D-genome, followed by the C-genome, with the lowest level of markers being assigned to the A-genome.

Conclusions:
A new oat 7K SNP array has been developed that contains 6,631 functional high-quality markers. Through the combination of markers from the previously published 6K array and novel markers from transcriptome sequencing, this array should be suitable for genotyping a large spectrum of hexaploid oat material both from Europe and North America. The 6,631 markers provide a marker density of more than 1 marker per centiMorgan over the hexaploid oat genome. Since most of the markers will be polymorphic in analyzed oat material, the new 7K oat array will be useful for genetic mapping of mendelian traits and QTLs, genetic relationship analyses, and marker-trait association studies. Finally, the marker density will also be sufficient to perform large scale Genomic Selection in oat breeding. At this time, the usefulness of this array has been validated through the analysis of more than 10,000 oat lines.

References:
Ganal, *et al.* (2019) High-Throughput Genotyping for Cereal Research and Breeding. In: Applications of Genetic and Genomic Research in Cereals (Editors: Thomas Miedaner, Viktor Korzun), eBook ISBN: 9780081022139

Kamal, *et al.* (2022) The mosaic oat genome gives insights into a uniquely healthy cereal crop. Nature, 606: 113-119, doi.org/10.1038/s41586-022-04732-y

Wang, *et al.* (2014) Characterization of polyploid wheat genomic diversity using a high-density 90,000 single nucleotide polymorphism array. Plant Biotechnology Journal, 12: 787-796, doi.org/10.1111/pbi.12183

Tinker, *et al.* (2014) A SNP genotyping array for hexaploid oat. Plant Genome, doi.org/10.3835/plantgenome2014.03.0010