

# Development of optimized haplotype-based genotyping arrays for cost-efficient use in research and plant breeding of wheat

Martin W. Ganal, Jörg Plieske, Jana Lemm, Eva-Maria Graner, Andreas Polley  
 TraitGenetics GmbH, Am Schwabeplan 1b, D-06466 Stadt Seeland OT Gatersleben, Germany  
 Web: [www.traitgenetics.com](http://www.traitgenetics.com) e-mail: [contact@traitgenetics.com](mailto:contact@traitgenetics.com)

## Introduction

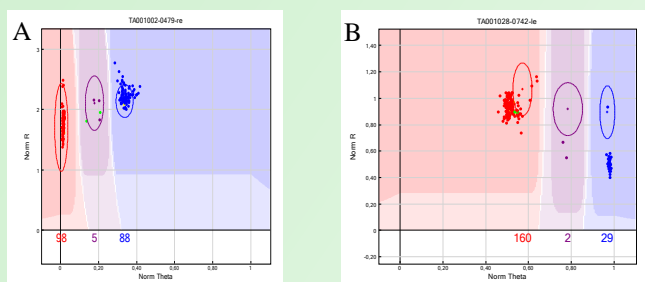
Using available genotyping arrays, large sets of varieties and SNP markers have been analyzed by TraitGenetics for wheat during our internal research activities. Based on these data, we have generated marker databases that include marker quality data, mapping positions, allele frequency data and haplotype information for many thousands of markers in wheat. Such information is crucial for the development of optimized marker sets for plant breeding that can be analyzed in a very cost-effective way. We present here a summary of our current efforts to develop optimized genotyping arrays in wheat.

## SNP Genotyping in Wheat

SNP genotyping in hexaploid bread wheat (*Triticum aestivum*) has made significant progress through the development of high density genotyping arrays such as the 90K Illumina Infinium array (Wang et al., 2014). We have used this array for the internal genotyping of more than 1000 wheat lines and varieties.

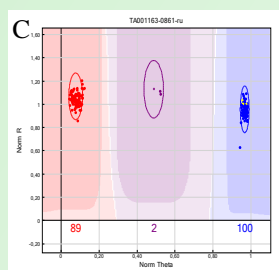
## Genotype calling and marker quality in hexaploid wheat

All markers on the 90K array were assessed for their quality in hexaploid wheat. Specifically, the distribution of the clusters and number of failed samples were determined.



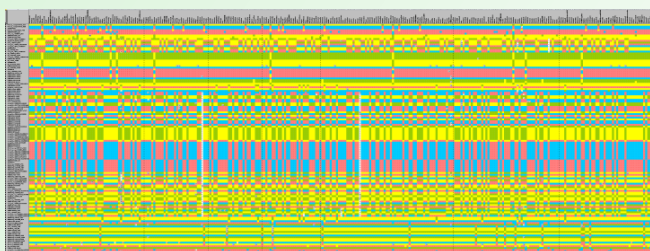
### Examples of cluster patterns in wheat

- A) Clustering of an SNP marker in the allohexaploid bread wheat. Clusters are highly condensed since two monomorphic genomes are in the background (actual detection is AAAAAA, AAAAAAB and AAAABB).
- B) Clustering of an SNP marker when only two genomes are detected. Clusters are less condensed since only one monomorphic genome is in the background (actual detection is BBBB, BBBA and BBAA).
- C) Typical clustering of an SNP marker when only one genome is detected (genome-specific). Detection is AA, AB and BB.



## Allele frequency and haplotype structure

Genotyping data from all internally genotyped wheat lines were collected and assembled in a database to determine the allele frequency of each marker. Together with reliable mapping data for 45832 markers assembled from published data (Wang et al., 2014) and from various mapping populations analyzed at TraitGenetics, the data for all lines and varieties were ordered based on the mapping data to identify the regional haplotype structure in the wheat genome. Many additional markers could be placed into the haplotype blocks, when they showed the same pattern in all investigated lines even without available mapping information.

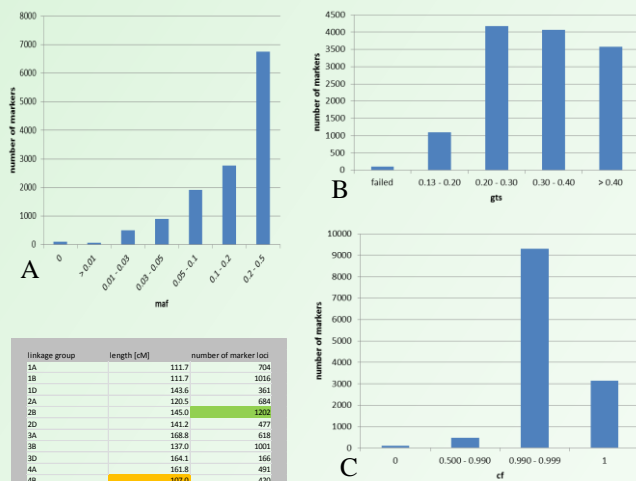


## Analysis of allele frequencies and haplotype structure

Genotype data for a number of wheat varieties for a specific chromosomal region are shown. Markers are ordered according to their map position to reveal the specific haplotype structure. This demonstrates that many marker create redundant data even in diverse wheat lines and varieties.

## An optimized genotyping array for hexaploid wheat

Assessing marker quality data (cluster types and cluster separation) for each marker, allele frequency in wheat material, the number of uncalled samples, and reducing the number of markers per haplotype block to on average 2, we developed an optimized 15K array with 12,905 markers.



linkage group	length [cM]	number of marker loci
1A	111.7	704
1B	111.7	1016
1D	143.6	361
2A	120.5	684
2B	141.0	388
2D	141.2	477
3A	168.8	618
3B	137.0	1001
3D	164.1	366
4A	161.8	491
4B	307.0	420
4D	112.9	86
5A	117.7	770
5B	179.6	1031
5D	167.8	234
6A	122.2	814
6B	110.4	943
6D	122.3	203
7A	167.5	762
7B	142.0	762
7D	196.1	184
<b>Total</b>	<b>2949.9</b>	<b>12905</b>

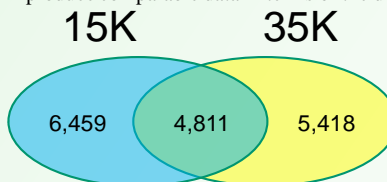
- D A genome: 4843 marker
- B genome: 6375 marker
- D genome: 1687 marker

## Quality data for the optimized 15K wheat array

- A) Allele frequencies (maf) of the employed markers in a wide range of wheat germplasm
- B) Quality data (Gentrin scores) for the markers on the array (>0.2 permits the calling of heterozygous samples in hexaploid wheat)
- C) Proportion of wheat germplasm that can be called with the individual markers (1= all analyzed lines were called)
- D) Number of markers on each chromosome and in each genome of hexaploid wheat

## Current activities

We have also analyzed the same material with the 35K Affymetrix Breeders array. These data show that our 15K array and the 35K array produce comparable data in terms of the detected haplotype blocks.



## Haplotype identification for the 15K and 35K array

11,270 haplotype blocks for our 15K Illumina array  
 10,229 haplotypes blocks for the 35K Affymetrix Breeders array  
 16,688 haplotype blocks in total with both arrays.

At the moment, we are further analyzing all available data in order to design a second version of an optimized array with approximately 20,000 markers that identifies most haplotype blocks in a wide range of wheat material (expected to be available around the middle of 2016).

## Reference

Wang et al. Characterization of polyploid wheat genomic diversity using a high-density 90,000 single nucleotide polymorphism array. *Plant Biotechnol J*. 12:787-796 (2014)

This research was in part supported by the German Ministry of Education and Research (BMBF) and the EU (FP 7). We thank our various collaboration partners in the private and public sector that provided material for the analysis of marker diversity in wheat.

**The optimized 15K wheat array is available for genotyping at TraitGenetics for 29-49 € per sample (depending on sample number) and has already been used for more than 50,000 samples. If you are interested in using this array, please do not hesitate to contact us.**