

Large scale SNP genotyping with optimized molecular marker sets for cost-efficient plant breeding in the Brassica species (*B. napus*, *B. oleracea* and *B. rapa*)

Jörg Plieske, Jana Lemm, Andreas Polley and Martin W. Ganal

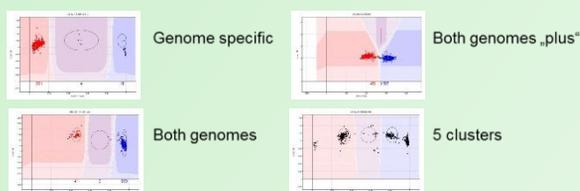
TraitGenetics GmbH, Am Schwabeplan 1b, D-06466 Stadt Seeland OT Gatersleben, Germany
 Web: www.traitgenetics.com e-mail: contact@traitgenetics.com

Introduction

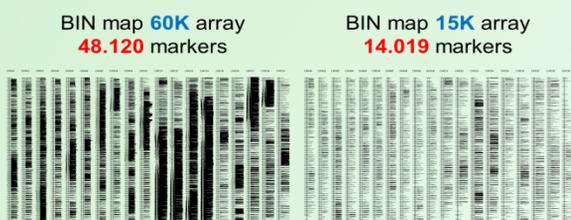
Through the development of the large Brassica 60K Illumina Infinium array, it has now become routine to generate a wealth of genotyping data for individual Brassica lines and varieties. However, with costs in the area of approximately 100 € per sample, this array is quite expensive. Furthermore, many of the genotype data generated in this way constitute of redundant information since many markers are in perfect linkage disequilibrium in breeding material and varieties.

Using the Brassica 60K genotyping array, large sets of varieties have been analyzed at TraitGenetics for *Brassica napus*, *B. oleracea* and *B. rapa* during our own internal research activities. Based on these data, we have generated marker databases that include mapping data, allele frequency data and haplotype information for the markers on the 60K array. 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 our current activities in the Brassica species towards the development of an optimized, high quality genotyping array that can be used in *Brassica napus* and its diploid ancestors at much reduced cost.

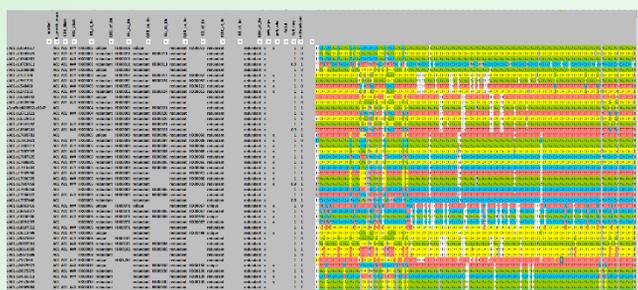
Illumina SNP calling for *Brassica napus*



Genetic BIN map of *B. napus* derived from various populations



Genotype data for a number of winter oilseed rape varieties for a specific chromosomal region. Markers are ordered according to their map position and specific haplotype structure.



Summary

With genotyping data from large sets of lines and varieties we have now selected specific marker sets based on polymorphism information content, haplotype structure, marker quality, genome-specificity in allopolyploid species, genetic map position and other quality features that are relevant for plant breeding. In this way, we can now offer very cost-effective molecular marker analyses based on an optimized array with around 13,000 markers (approx. 29-49 € including array, sample processing and the generation of the genotype table) and also with individual marker sets for various breeding applications including marker-assisted selection, cost-effective genome selection, marker-assisted backcrossing, variety identification and other purposes.

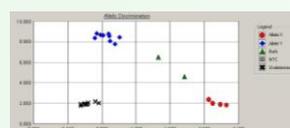
If you are interested in using these arrays or marker sets through our services, please do not hesitate to contact us.

With genotyping data generated from a large set of lines derived from various sources and countries in combination with mapping information from various mapping populations, we have investigated in detail the extent of LD and marker haplotype groups in elite *Brassica napus* material as well as in *B. rapa* and *B. oleracea* to identify haplotype-specific markers of high quality. Taking all this information together, we have subsequently generated an optimized genotyping array with around 13,000 markers for routine use in genetic analyses and breeding including genomic selection of oilseed rape and Brassica vegetables.

Number of markers on the array: **13.714** (100%)
 Number of functional markers: **13.465** (98%)
 Number of polymorphic markers in **984** *B. napus* samples: **12.792** (93%)
 Number of polymorphic markers with MAF >0.03: **11.917** (87%)

This 15K array can now be used at much reduced costs compared to other arrays and without much loss of information compared to larger and more expensive arrays.

In parallel, we have generated an optimized marker collection based on individual SNP markers (KASP) for variety identification, variety purity analysis, marker-assisted backcrossing and other purposes.



(Top) Example for a C genome-specific marker (x are failed *B. rapa* A genome samples)
 (Right) Distribution and quality of the selected KASP markers for the A and C genome.

Chromosome	No. of markers	functional	%	genome-specific	%
A01	35	35	100.0	30	85.7
A02	37	36	97.3	28	77.8
A03	35	33	94.3	24	72.7
A04	35	35	100.0	24	68.6
A05	35	35	100.0	21	60.0
A06	35	33	94.3	28	84.8
A07	35	34	97.1	27	79.4
A08	35	35	100.0	30	85.7
A09	35	34	97.1	25	73.5
A10	35	35	100.0	32	91.4
C01	35	35	100.0	28	80.0
C02	35	32	91.4	25	78.1
C03	35	33	94.3	31	93.9
C04	35	33	94.3	28	84.8
C05	35	35	100.0	34	97.1
C06	35	34	97.1	31	91.2
C07	35	35	100.0	28	80.0
C08	35	34	97.1	29	85.3
C09	35	34	97.1	31	91.2
	660	650	97.7	534	82.2