GENETIC STRUCTURE OF SYNTHETIC *BRASSICA NAPUS* L. POPULATIONS

Iulian GABUR[1], Rod J. SNOWDON[1], Christian OBERMEIER[1]

e-mail: iulian.gabur@agrar.uni-giessen.de

Abstract

The crop species *Brassica napus* L. has significant economic importance around the world. However, the complex evolutionary history and vast geographical distribution of oilseed rape has contributed vastly to genetic population structure investigations. Constant breeding efforts, for use for oilseed rape as food for human consumption, and fodder for livestock, have generated new phenotypic diversity. In this study, we used crosses among very diverse morphotypes as *Brassica oleracea* (turnip rape), *conv. capitata var. medullosa* (Cavalier rouge), *conv. capitata var. sabauda* (Savoy 'Wirsing'), *conv. botrytis var. alboglabra* (broccoli); *Brassica rapa* (turnip), *var. trilocularis* (yellow sarson), *var. chinensis* (bok choy); *Brassica cretica; Brassica montana.* Until now, genetic studies had insufficient genotypes to determine the relationship of oilseed morphotypes and their genetic population structure. We used 18,272 single nucleotide polymorphism markers in a synthetic nested association mapping diverse panel of 200 *B. napus* accessions that included crosses of five very diverse parental lines and a common elite accession. Results on population genetic structure and phylogenetic analyses revealed, as expected, five subpopulations that were largely reflective of phenotypes. The results of this study have provided improved resolution to the genetic and phylogenetic relationships of a synthetic panel within the Brassicas species. Understanding genetic diversity available is key to the future genetic study and constant improvement of this important agronomical crop species.

Key words: oilseed rape, NAM, diversity panel, SNP, phylogenetic analysis

Brassica napus (2n=4x=38, AACC, also known as canola and oilseed rape) has an allotetraploid genome and is the result of a relatively recent spontaneous interspecific hybridisation between two diploid progenitors, *Brassica rapa* (2n=2x=20, AA) and *Brassica oleracea* (2n=2x=18, CC), respectively (U, 1935). Oilseed rape is known to contain the intact chromosome complements of *B. rapa* and *B. oleracea* (Axelsson *et al*, 2000).

Friedt and Snowdon (2010) suggest that the allopolyploidization events occurred as a result of co-cultivation of the two diploid progenitors in close geographical proximity, in the Mediterranean region. Recently, Chalhoub *et al*, (2014) published the oilseed rape reference sequence genome of *B. napus* for the cultivar Darmor-*bzh*. In the same study it was revealed that the oilseed rape genome has undergone 72-fold multiplications and has a high genetic redundancy. By comparison of orthologous gene families with the diploid progenitors, the species *Brassica napus* seems to not be older than 7,500 to 12,500 years, thus, a plant species with a relatively short domestication history.

Oilseed rape is economically important crop and brings high value in cereal crop rotations by providing positive influence on yields of subsequent crops of wheat and barley (Christen and Sieling, 1993). In many regions across Europe, Canada and Australia, oilseed rape has an important role in soil rejuvenation and management strategies of monocotyledonous cereal diseases and pests.

During its cultivation *Brassica napus* evolved in various climate zones and displays a wide genetic variation in regard to vegetative growth and winter hardiness. Geographical differentiation among spring, semi-winter and winter oilseed rape cultivars is under genetic control of vernalisation requirement and flowering time. Nowadays, winter oilseed rape is the most heavily cultivated oilseed crop and a major source of biodiesel in Europe. Spring oilseed rape is present in North America (particularly Canada) and some parts of China. On the market also intermediary types of oilseed rape can be found which are suitable for Asia, the Indian subcontinent and Australia (Bus *et al*, 2011; Snowdon *et al*, 2006). Worldwide, oilseed rape is

[1] Department of Plant Breeding, IFZ Research Centre for Biosystems, Land Use and Nutrition, Justus Liebig University Giessen, Heinrich-Buff-Ring 26-32, 35392 Giessen, Germany

the third most important oilseed crop after soybean and oil palm that have greater oil productions (FAOSTAT data, 2015: http://faostat.fao.org/).

Within the last decades breeding strategies aimed at improving the oil and meal quality of oilseed rape. Moreover, oilseed rape becomes an efficient substitute of fossil energy by promoting clean, eco-friendly and renewable resources. Another important breeding objective was the improvement of seed quality for human consumption. In the last 30 years, the so-called "double-low" (00) seeds with zero erucic acid and low glucosinolate content were developed and commercialized on the world market. However, the 00-seed quality created a very strong selection bottleneck and limited the genetic diversity in modern oilseed cultivars.

Overcoming the genetic diversity erosion in modern *Brassica napus* spring and winter modern cultivars is a continuing challenge for breeders. Development of a nested association mapping (NAM) panel representing the current genetic diversity within winter type *B. napus*, but also introducing new genetic diversity from the progenitors of *B. napus* is a suitable genetic resource that will allow breeders to capture novel loci for the improvement of important agronomical traits in oilseed rape.

MATERIAL AND METHOD

Plant material

A subset of a *Brassica napus* nested association mapping (BnNAM) population was used in this study. The BnNAM population consists of 50 genetically diverse winter oilseed rape accessions (20 exotic *B. napus*, 30 resynthesized *B. napus*) crossed with an elite doubled haploid winter-type oilseed rape cultivar (DH5ON). Each of the 50 subpopulations is composed of ~50 doubled haploid lines per cross (where both parents are natural *B. napus*) or ~50 single backcross recombinant inbred lines (BC1-RILs) for crosses with one resynthesized *B. napus* parent (Snowdon *et al*, 2015). The present study used five BnNAM subpopulations with a total of 200 BC1-RILs derived from synthetic *B. napus* founders carrying a very diverse genetic background (*table 1*).

DNA Isolation

The 200 accessions were grown in a greenhouse at Justus Liebig University, Department of Plant Breeding, Giessen, Germany. Fresh leaf tissue was harvested from the youngest true leaf (leaf 3–4) 4 weeks after planting. Leaf tissue was frozen using liquid nitrogen and samples were stored at −80°C until DNA was isolated. Total genomic DNA isolation was conducted using the DNeasy 96 plant kit (Qiagen, USA) with minor modifications of the protocol (unpublished data). DNA concentration was quantified using Qubit fluorometer and NanoDrop Fluorospectrometer (Thermo Fisher Scientific, Waltham, MA).

SNP genotyping and quality control

The entire BnNAM panel was genotyped with the 60K Illumina Infinium Brassica single nucleotide polymorphism (SNP) array containing 52,158 SNP probes. Using the Darmor-*bzh* reference v4.1 (Chalhoub et al., 2014), 28,073 SNP marker were anchored to the reference using BLASTN as described by Qian *et al* (2014).

Population structure and phylogenetic analysis

Population structure analyses were conducted using the R package "Selection tools" (http://population-genetics.uni-giessen.de/~software/). Population stratification was determined by calculating the kinship matrix and the first five principal components. Only markers with MAF < 0.05 were included in the analysis. To examine the phylogenetic relationship between the oilseed rape accessions an unweighted pair group matrix algorithm (UPGMA) was used. This analysis considers all samples independently and groups them in k-means clusters based on the heterogeneity among subpopulations. Initially, the genetic distance among genotypes was calculated using the Roger's Distance algorithm and all samples were transposed into a matrix. Therefore, we used all 200 genotypes independent of their breeding history. Results were displayed using the R package "ape".

Table 1

Parents, genetic origin and composition of synthetic *Brassica napus* subpopulations used for population structure analysis

Parental lines	Type	Accesion name	Number of RILs	Mother	Variety	Father	Variety
PBY033	Synthetic	H149	48	*B. oleracea* conv. *capitata* var. *medullosa*	Cavalier rouge'	*B. rapa chinensis*	Pak Choi
PBY034	Synthetic	H165	28	*B. oleracea* conv. *capitata* var. *sabauda*	Savoy 'Wirsing'	*B. rapa chinensis*	Pak Choi
PBY040	Synthetic	RS13/6	53	*B. rapa chinensis*	Chinese cabbage	*B. oleracea* var. *alboglabra*	Broccoli
PBY050	Synthetic	CRY1	41	*B. rapa* spp. *troilocularis*	Yellow Sarson	*B. cretica*	
PBY052	Synthetic	MOY4	31	*B. rapa* spp. *Troilocularis*	Yellow Sarson	*B. montana*	
PBY061	**Elite**	**DH5ON**	**-**	***B. napus ssp. napus***	**Oase**	***B. napus ssp. napus***	**Nugget**

RESULTS AND DISCUSSIONS

The high-density Brassica 60k SNP Illumina consortium array (Illumina, San Diego, CA, USA) containing in total 52,157 SNP probes was used for genotyping of the diversity set. After filtering for markers showing polymorphism within the diversity panel and filtering for markers which could be unambiguously anchored to the *B. napus* Darmor-*bzh* reference genome using very stringent BLAST alignment parameters (zero mismatches of the SNP flanking sequences). We identified 28,073 single-locus probe SNPs. Therefore, these markers mapped at a single physical genome position were used in downstream analyses.

For population structure analysis, 10,005 SNPs with MAF < 5% and missing call > 10% were also eliminated, leaving only high-quality, polymorphic, single-locus SNPs. After filters were applied to SNPs and samples, the final *B. napus* data set included a total of 18,068 high-confidence SNP markers scored on 200 NAM accessions. These SNP markers were evenly distributed along the 19 *B. napus* chromosomes, with an average of one SNP every 27,702 bp on the A sub-genome, and 47,344 bp on the C sub-genome, respectively.

The A sub-genome contains 48.5 % SNPs, and the C sub-genome 51.5 %, respectively. Resulted values are in accordance with the expected distribution based on size in Mb. In *B. napus* the A chromosomes are relatively smaller (in total ~ 238 Mb) than the C chromosomes (in total ~ 405 Mb) according to the Darmor-*bzh* reference genome (Chalhoub *et al*, 2014).

The 18,068 SNPs were used to evaluate the genome-wide patterns of population structure in this globally representative and diverse nested association mapping *B. napus* panel. K-means clustering approaches determined that five subpopulations (K = 5) was the optimal clustering for the 200 accessions based on the 18,068 high quality SNP markers. Phylogenetic analysis based on Rogers genetic distance (Rogers, 1972) supported a major influence of the founder parents on the population structuration (*figure 1*).

Thus, for further analyses of population stratification a principal component analysis (PCA) was performed. To quantify the genetic differentiation between the five subpopulations detected, we estimated the identity-by-state values using 18,068 high density quality-filtered polymorphic SNP markers available for the NAM panel, similar to Qian (2014). Data for 4 synthetic lines was excluded from population structure analysis as it contained more than 10% missing SNP calls per genotype. The first three principal components (PC) explained 9.3 % (PC1), 7.7% (PC2) and 6.9% of the variation in the genotypic data, respectively. Based on the founder parent, the accessions were separated by PCA into 5 subfamilies (H149, H165, RS13/6, CRY1, MOY4) (*figure 2*).

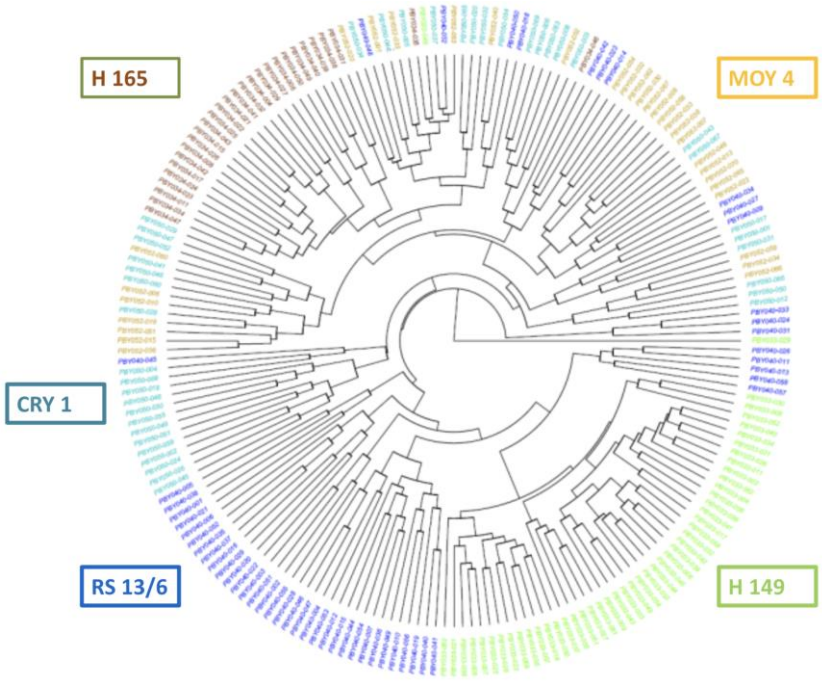

Figure 1 **Unweighted pair group matrix algorithm (UPGMA) tree describing genetic relationships among 200 winter type synthetic *Brassica napus* accessions**

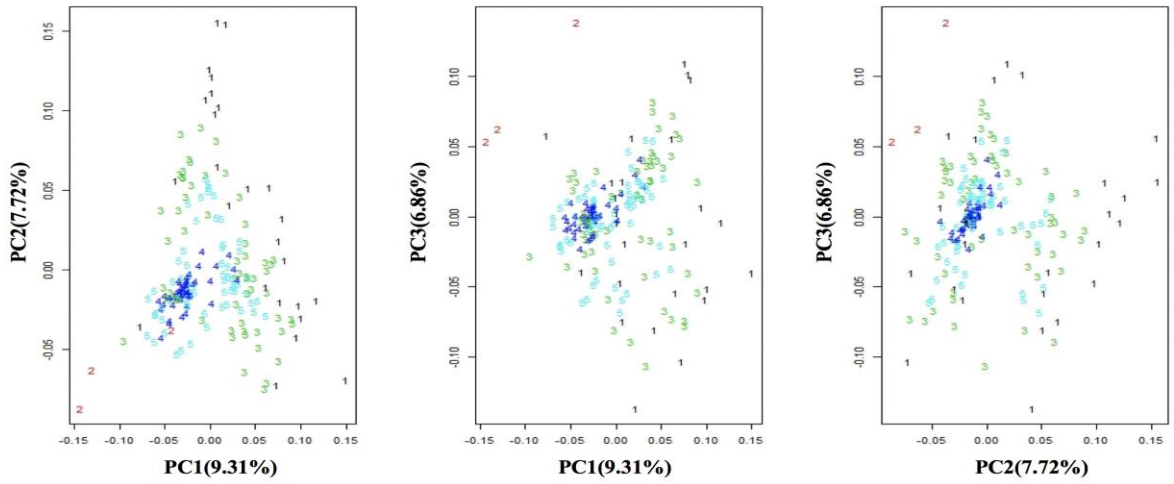

Figure 2 **Comparative principal coordinate analysis (PCA) of subgenomic genetic diversity, measured across 18,068 SNP markers per 196 lines from synthetic *Brassica napus* populations using PC1, PC2 and PC3.**

PCA and UPGMA tree analysis, using only 200 nested association mapping *B. napus* winter genotypes to analyse population structure, resulted in subdivision into the same subpopulations by PCA analysis, whereby the principal component accounting for genetic diversity was smaller than with inclusion of the outliers in the full set of 196 lines (*figure 1 and 2*). The results of the UPGMA tree analysis corresponded with a high similarity to the PCA. This result is in agreement with the phylogenetic analysis performed for the synthetic nested association mapping panel, where the variance explained by PC1 was higher. Despite the differentiation between subpopulations, the overall shape of the PCA plot for each two principal components was similar. As expected, the subfamilies derived from crosses with exotic synthetic lines, as *B. cretica* and *B. montana*, exhibit a higher genetic diversity.

CONCLUSIONS

Using polymorphic, single-copy, genome-wide SNP markers to analyse genetic diversity in winter *B. napus* is a useful tool for breeding. We found evidence of potential new genetic resources coming from exotic material, as *B. cretica* and *B. montana*. Results indicate that new genomic regions could be associated with important agronomical traits on the A- and C-subgenome. Access to new breeding pools is particularly important for efforts to increase seed quality traits and disease resistance in modern oilseed rape cultivars.

ACKNOWLEGMENTS

Funding was received within the European PLANT-KBBE-IV research funding programme for the French-German consortium GeWiDis (Exploiting Genome Wide Diversity for disease resistance improvement in oilseed rape) in Germany by the federal Ministry of Education and Research (BMBF, grant 031A325) and in France by the L'Agence nationale de la recherche (ANR, grant ANR-13-KBBE-0004-01), respectively.

Iulian Gabur acknowledges funding through a scholarship from the German Academic Exchange Service (DAAD). SNP genotyping data were provided by the PRE-BREED-YIELD consortium (BMBF grant no. 0315964).

REFERENCES

Axelsson T., Bowman C.M., Sharpe A.G., Lydiate D.J., Lagercrantz U., 2000 - *Amphidiploid Brassica juncea contains conserved progenitor genomes*. Genome 43: 679-688

Bus A., Körber N., Snowdon R.J., Stich B., 2011 - *Patterns of molecular variation in a species-wide germplasm set of Brassica napus*. Theoretical and Applied Genetics 123: 1413–1423.

Chalhoub B., Denoeud F., Liu S., Parkin I.A., Tang H., Wang X., et al, 2014 - *Early allopolyploid evolution in the post-Neolithic Brassica napus oilseed genome*. Science 345: 950-953.

Christen O., Sieling K., 1993 - *The effect of different preceding crops on the development, growth and yield of winter barley*. Journal of Agronomy and Crop Science 171: 114–123.

Friedt W., Snowdon R., 2010 - *Oilseed Rape*. In: Oil Crops Vollmann J and Rajcan I, (eds), pp. 91–126. Springer New York, New York, NY.

Qian L., Qian W., Snowdon R.J., 2014 - *Sub-genomic variation patterns as a signature of breeding in rapeseed*. BMC Genomics 15: 1170

Rogers J.S., 1972 - *Measures of Genetic Similarity and Genetic Distance*. In: Studies in Genetics VII, University of Texas Publication 7213, Austin, 145-153

Snowdon R.J., Lühs W., Friedt W., 2006 - *Oilseed rape*. In: Kole C. (ed) Genome Mapping and Molecular Breeding, Vol. 2: Oilseeds. Springer Verlag, Heidelberg, NewYork/Berlin/Toyko, pp.55-114.

U N., 1935 - *Genome Analysis in Brassica with Special Reference to the Experimental Formation of B. Napus and Peculiar Mode of Fertilization*. Japanese Journal of Botany, 7:389-452