LINKAGE DISEQUILIBRIUM IN THE GENOME OF SYNTHETIC *BRASSICA NAPUS* POPULATIONS

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

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

Abstract

New technologies as high-density single-nucleotide polymorphism (SNP) genotyping arrays are a powerful tool that can give valuable insight into patterns of linkage disequilibrium (LD) in the recent domesticated *Brassica napus* genome. This study used the Brassica 60k SNP Illumina consortium genotyping array to assess the distribution of LD and haplotype structure in a diverse panel of 200 synthetic lines of winter oilseed rape (*Brassica napus*). Pairwise LD analysis was conducted within the A- and C-subgenomes of oilseed rape. Results revealed that LD decayed, on average, more rapidly in the A-subgenome (0.15 Mb) than in the C-subgenome (2.00 Mb). Our findings suggest the presence of a strong selection of large genomic regions associated with important agronomical traits, especially on the C-subgenome. These results imply that during oilseed rape artificial and natural selection, the C-subgenome was of particular interest for breeders. Increasing the genetic diversity and recombination rate on the whole genome level is of crucial importance for future breeding efforts

Key words: oilseed rape, SNP, linkage disequilibrium analysis, LD

Most important agronomic traits in crops are known to be controlled by complex quantitative trait loci (QTL). In oilseed rape (or canola: *Brassica napus*), the world's second most important oilseed crop, numerous studies have reported QTL for various agronomic, developmental, seed quality and disease resistance traits since the first genetic mapping of QTL in this crop (Uzunova *et al*, 1995). Typically, bi-parental genetic mapping populations are used to detect genomic regions involved in important agronomical traits. However, mapping resolution can be limited in conventional QTL mapping approaches by population size, low number of recombination events and polymorphism. Alternatively, genome wide association studies (GWAS) use genetically unrelated collections or populations of breeding lines and are very useful for QTL localization (Flint-Garcia *et al*, 2003). In contrast to bi-parental QTL mapping, GWAS is based on linkage disequilibrium (LD) and utilises a higher number of recombination, therefore improving the mapping resolution. A major prerequisite for analysis of linkage disequilibrium patterns is the availability of numerous molecular markers spanning every chromosome.

Nowadays, in many important agronomical crops, genome-wide marker screening is done using single-nucleotide polymorphism (SNP) markers. High-density SNP arrays like the Brassica 60 k SNP Illumina consortium array (Illumina, San Diego, CA, USA) allows high-resolution linkage disequilibrium (LD) analysis in various population sizes.

LD analyses are an important tool that provide insight into many evolutionary events across the crop breeding history and can directly guide breeders in their search for genetic diversity in various gene pools. Moreover, it can reflect the history of mutations and gene conversion events in a genome.

LD, at a conceptual level, describes the non-random association of alleles at two or more loci caused by genetic linkage. Among the factors that influence the LD decay estimates in crop genomes are, in particular, species, gene pool and population size. For example, in *Arabidopsis* LD decay has been estimated, on average, from 50 kb (Nordborg *et al*, 2005) to over 250 kb (Hagenblad *et al*, 2002), whereas in maize is between 0.5-7.0 kb (Remington *et al*, 2001; Ching *et al*, 2002; Palaisa *et al*, 2003) and 1–10 kb (Yan *et al*, 2009), while in rice 75–150 Kb (Mather *et al*, 2007). In different *B. napus* populations the average LD decay estimates were varying greatly, from 1–2 cM (Ecke *et al*, 2010) to 20 cM (Zou *et al*, 2010) and 0.25-0.30 Mb (A-subgenome), and 2.00-2.50 Mb (C-subgenome) (Qian *et al*, 2014). Previous

[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

studies have suggested that SNPs in strong LD are typically organized into haplotype blocks. In a crop genome, the genetic variation is defined by abundance and conservation of haplotype blocks. This information is important for accurate understanding of associations between molecular markers, genes and agronomical relevant phenotypic traits.

Brassica napus is a relatively recent allo-tetrapolyploid (AACC, 2n = 38), derived from interspecific hybridisation events between *B. rapa* (AA, 2n = 20) *and B. oleracea* (CC, 2n = 18) (Allender *et al*, 2010). The recent alloploidisation event makes *B. napus* an interesting model for investigating genome-wide and subgenome specific linkage disequilibrium patterns.

The objectives of this study are to evaluate the extent of LD decay and variation in the distribution of haplotype block size within the A- and C-subgenomes of synthetic *Brassica napus* NAM populations.

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 very diverse genetic background.

DNA Isolation

The 200 accessions were grown in a greenhouse at the 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), we anchored 28,073 SNP marker using BLASTN as described by Qian et al. (2014). Initially, monomorphic SNP markers that exhibited a minor allele frequency (MAF) <0.05 and a failed call frequency >90% were removed from the SNP data set. Subsequently, only polymorphic SNP markers that were previously anchored to the Darmor-*bzh* reference v4.1 were included in further analysis.

Analysis of linkage disequilibrium

To investigate chromosome-wide and genome-specific patterns of linkage disequilibrium, the squared allele-frequency correlations (r2) between pairs of SNPs were calculated. Only markers with max.10% missing data and MAF ≥ 0.05 were included in the analysis. Haplotype blocks were defined with the confidence interval method described by Gabriel et al. (2002) in Haploview version 4.2 (Barrett et al. 2005) and with the R package 'GenABEL' (Aulchenko et al. 2007). Additionally, haplotype blocks were defined for each significant associated marker when markers flanked after anchoring to the *Brassica napus* Darmor-*bzh* reference were in strong LD (r2 > 0.4), as described by Hatzig et al. (2015).

RESULTS AND DISCUSSIONS

Linkage disequilibrium (LD) decay rate in the *Brassica napus* nested association mapping (NAM) subpanel revealed regions with highly conserved LD blocks. The mean physical distance at which the pairwise genotypic marker-marker associations (r2) decayed below a threshold of 0.2 was 0.15 Mb for the whole genome and the mean LD decay within the NAM population was 2 Mb, similar to the LD decay described within a Chinese semi-winter oilseed rape panel by Qian (2014). Generally, on the C chromosome a higher LD decay was observed (*figure 1*). As the genome donors of 3 of the 5 parents of the used NAM panel (H149, H165, RS13/6) have chinese semi-winter oilseed rape ancestors, the observed differences in LD decay between different chromosomes was expected. However, this result indicates that a higher mapping resolution could be obtained in the NAM panel using genome-wide association studies for important agronomical traits. Within the selected NAM population was a higher LD decay was observed in the A-subgenome, compared to the C-subgenome. This should be taken into consideration in the further

detailed analyses of identified genome regions involved in marker-trait associations.

The 18,068 polymorphic SNP markers used in our study were sufficient to perform a preliminary genome-wide analysis of haplotype block structure in the *B. napus* nested association mapping panel. In particular, we showed that some oilseed rape chromosomes carry extremely large segments of highly conserved LD in synthetics-derived lines, and that this phenomenon is a mostly observed on the C- subgenome chromosomes. This may indicate that interspecific hybridization with *B. rapa* has increased the recombination rates of A-subgenome chromosomes, causing more rapid LD decay and shortened LD blocks in the A-subgenome.

Another contributing factor to LD decay pattern in *Brassica napus* is likely to be artificial and natural selection for domestication and adaptation traits. During geographical adaptation and human selection many traits as flowering time or seed quality were crucial for domestication, although strong selection at a locus may reduce diversity and increase LD and haplotype block size in the genomic region (Rafalski *et al*, 2004).

Stronger LD and longer-range haplotype blocks were detected on chromosomes C01, C02, C04 and C09 suggest selection the corresponding genomic regions.

Regions of interest for major agronomical traits could be delimited by the use of local LD patterns that include significant SNP marker-trait associations. This approach provides a valuable method to identify and localize potential factors influencing crop traits, as potential candidate genes. As an example, very strong LD blocks identified on the C-subgenome correspond to previous published studies findings in various *B. napus* populations, as seed quality traits (Qian *et al*, 2014), seed germination and seedling development (Hatzig *et al*, 2015) and flowering time genes (Schiessl *et al*, 2017). This reflects extreme selection during breeding for specific agronomical traits. In recent years considerable progress has been made in increasing genetic diversity on the C-subgenome, by introducing new donors, in order to improve disease resistance. Typically, strong LD blocks could be used when defining QTL confidence intervals from potential strong candidate genes.

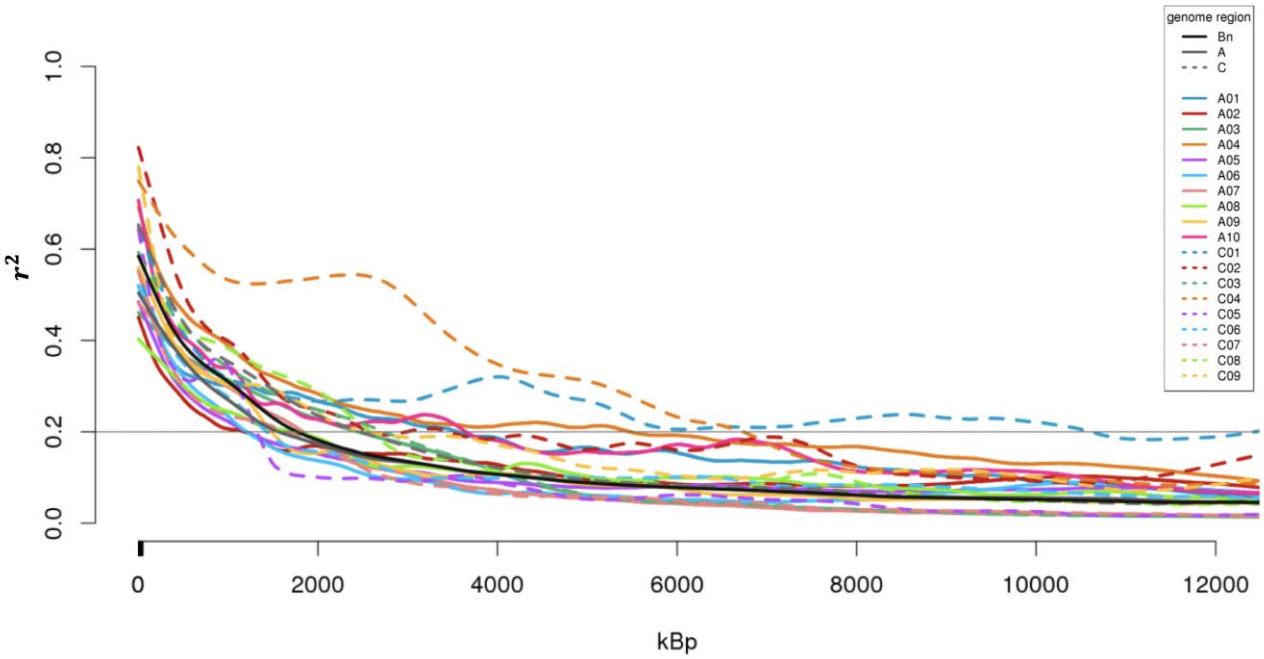

Figure 1 **Patterns of linkage disequilibrium (LD, r2 = 0.2) across the 19 *Brassica napus* chromosomes, measured with 18,068 SNP markers. The solid lines represent LD decay on the A-subgenome chromosomes, while the dashed lines represent LD decay on the C-subgenome chromosomes.**

CONCLUSIONS

Analysis of genetic diversity and knowledge about recombination rates are of crucial importance for future breeding programs. New large-scale automated phenotyping platforms have the potential to reveal the broad variability across *Brassica napus* populations. Phenotypic variability is influenced by both genetic and environmental factors. Complementing automated phenotyping platforms with polymorphic, single-copy, genome-wide SNP markers will enable

scientists and breeders to better analyse and understand genetic diversity in winter *B. napus*. Our results found evidence of selections hotspots, especially on the C-subgenome. Future breeding improvements imply the identification of positive alleles for poorly selected traits and introgression in new, elite cultivars. Additionally, access to new genetic diversity is particularly important for breeding of important traits as seed quality and disease resistance in modern oilseed rape cultivars. This study provides a basis for the establishment of genomic selection tools for improved oilseed rape.

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

Allender C.J., King G.J., 2010 - *Origins of the amphiploid species Brassica napus L. investigated by chloroplast and nuclear molecular markers*. BMC Plant Biol 10:54.

Aulchenko Y.S., Koning D.J., Haley C., 2007 - *Genome wide rapid association using mixed model and regression: a fast and simple method for genome wide pedigree-based quantitative trait loci association analysis*. Genetics 177: 577–585.

Barrett J.C., Fry B., Maller J., Daly M.J., 2005 - *Haploview: analysis and visualization of LD and haplotype maps*. Bioinformatics 21: 263–265

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.

Ching A., Caldwell K.S., Jung M., Dolan M., Smith O.S., Tingey S., Morgante M., Rafalski A.J., 2002 - *SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines*. BMC Genet 3:19.

Ecke W., Clemens R., Honsdorf N., Becker H.C., 2010 - *Extent and structure of linkage disequilibrium in canola quality winter rapeseed (Brassica napus L.)*. Theor Appl Genet 120:921–931.

Flint-Garcia S.A., Thornsberry J.M., Buckler E.S., 2003 - *Structure of linkage disequilibrium in plants*. Annu Rev Plant Biol 54:357–374.

Gabriel S.B., Schaffner S.F., Nguyen H., Moore J.M., Roy J., et al., 2002 - *The structure of haplotype blocks in the human genome*. Science 296: 2225–2229

Hagenblad J., Nordborg M., 2002 - *Sequence variation and haplotype structure surrounding the flowering time locus FRI in Arabidopsis thaliana*. Genetics 2002, 161:289–298.

Hatzig S.V., Frisch M., Breuer F., Nesi N., Ducournau S.., Wagner M.H., et al, 2015 - *Genome-wide association mapping unravels the genetic control of seed germination and vigor in Brassica napus*. Front. Plant Sci. 6:221

Mather K.A., Caicedo A.L., Polato N.R., Olsen K.M., McCouch S., Purugganan M.D., 2007 - *The extent of linkage disequilibrium in rice (Oryza sativa L.)*. Genetics 177:2223–2232.

Nordborg M., Hu T.T., Ishino Y., Jhaveri J., Toomajian C., Zheng H., Bakker E., Calabrese P., Gladstone J., Goyal R., Jakobsson M., Kim S., Morozov Y., Padhukasahasram B., Plagnol V., Rosenberg N.A., Shah C., Wall J.D., Wang J., Zhao K., Kalbfleisch T., Schulz V., Kreitman M., Bergelson J., 2005 - *The pattern of polymorphism in Arabidopsis thaliana*. PLoS Biol 3:e196.

Palaisa K.A., Morgante M., Williams M., Rafalski A., 2003 - *Contrasting effects of selection on sequence diversity and linkage disequilibrium at two phytoene synthase loci*. Plant Cell 15:1795–1806.

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

Rafalski A., Morgante M., 2004 - *Corn and humans: recombination and linkage disequilibrium in two genomes of similar size*. Trends Genet 20:103–111.

Remington D.L., Thornsberry J.M., Matsuoka Y., Wilson L.M., Whitt S.R., Doebley J., Kresovich S., Goodman M.M., Buckler E.S., 2001 - *Structure of linkage disequilibrium and phenotypic associations in the maize genome*. Proc Natl Acad Sci U S A 98:11479–11484.

Schiessl S., Huettel B., Kuehn D., Reinhardt R., Snowdon R.J., 2017 - *Targeted deep sequencing of flowering regulators in Brassica napus reveals extensive copy number variation*. Sci Data. 4: 170013.

Uzunova M., Ecke W., Weißleder K., Röbbelen G., 1995 - *Mapping the genome of rapeseed (Brassica napus L.). I. Construction of an RFLP linkage map and localization of QTLs for seed glucosinolate content*. Theor Appl Genet 90:194–204.

Yan J.B., Shah T., Warburton M.L., Buckler E.S., McMullen M.D., Crouch J., 2009 - *Genetic characterization and linkage disequilibrium estimation of a global maize collection using SNP markers*. PLoS One 4:e8451.

Zou J., Jiang C., Cao Z., Li R., Long Y., Chen S., Meng J., 2010 - *Association mapping of seed oil content in Brassica napus and comparison with quantitative trait loci identified from linkage mapping*. Genome 53:908–916.