Analysis of linkage disequilibrium in canola-quality winter rapeseed (*Brassica napus* L.)

Wolfgang Ecke, Rosemarie Clemens, Heiko C. Becker

Department of Crop Sciences, Georg-August-University Göttingen, Von-Siebold-Str. 8, 37075 Göttingen, Germany
E-mail: wecke@gwdg.de

Abstract

The extend and structure of linkage disequilibrium (LD) was analysed in a set of 90 canola quality winter rapeseed lines derived from six breeding companies using 471 AFLP markers. To determine recombination frequencies the AFLP markers were also scored in 94 lines of a segregating doubled haploid population from a cross between the winter rapeseed variety 'Express' and a resynthesized rapeseed line, 'R53′. Only 11.36% of the 110,685 pair wise combinations of the 471 AFLP markers were in significant LD with a mean r^2 of 0.13, showing a low overall level of linkage disequilibrium in canola quality winter rapeseed. Mean r^2 values of linked and unlinked marker pairs in significant LD of 0.32 and 0.08, respectively, indicated that the major cause for linkage disequilibrium was linkage while linkage disequilibrium due to population structure, genetic drift or selection was negligible. Among linked markers linkage disequilibrium was high for closely linked markers but decayed rapidly within a few cM. The results are discussed with respect to the breeding history of rapeseed and the feasibility of a global approach to association analysis.

Key words: linkage disequilibrium, AFLP markers, rapeseed, Brassica napus, association mapping

Introduction

QTL mapping in rapeseed has usually been carried out in segregating populations derived from biparental crosses using interval mapping. This approach allows the identification of loci contributing to a quantitative trait and the estimation of their effects but is somewhat limited in its resolution due to the limited number of recombination events available in a segregating population, resulting in confidence intervals for the QTL positions in the range of several cM up to several tens of cM (van Ooijen 1992; Darvasi et al. 1993). Furthermore, only QTL can be detected that are polymorphic between the two parents.

An alternative approach could be association analyses using natural populations or, in the case of crop plants, collections of varieties or breeding lines. Due to the higher number of recombination events in such a material a higher resolution could be achieved than in QTL mapping (Ewens and Spielman 2001; Jannink et al. 2001). In addition, the genetic diversity would not be limited to the diversity occurring between two parental lines. On the other hand, this approach is strongly dependent on the extent of linkage disequilibrium (LD) in the plant population used. A global approach to association analysis, in which the whole genome is scanned for QTL using anonymous markers, is only feasible if significant linkage disequilibrium extends over at least a few cM.

Linkage disequilibrium, that is the non-random association of alleles at different loci (Flint-Garcia et al. 2003), can be caused by linkage, population stratification, genetic drift or selection. Its structure and extend varies strongly between plant species and even between populations of the same species (reviewed in Flint-Garcia et al. 2003), depending on the population history. As such, it can only be determined empirically. To determine the applicability of a global approach to association mapping in current canola-quality winter rapeseed, linkage disequilibrium was analysed in a population of 90 winter rapeseed varieties.

Materials and Methods

Plant materials: Linkage disequilibrium was analysed in a set of 90 canola quality winter rapeseed varieties and breeding lines – further referred to as LD population – from six different breeding companies, e.g. DSV, KWS, Limagrain-Nickerson, NPZ, SW Seed, and Syngenta. Recombination frequencies between markers were determined in 94 lines of a segregating doubled haploid population (F_1DH) from a cross between an inbred line of the winter rapeseed variety 'Express' and a resynthesized rapeseed genotype, 'R53′.

Marker analysis: DNA was prepared from 0.1 g young leave material from greenhouse grown plants using Nucleon PhytoPure Genomics DNA Extraction Kits (GE Healthcare, Germany) according to the manufacturers instructions. AFLP analyses were carried out following the protocol of Vos et al. (1995), modified according to Kebede and Kopisch-Obuch (unpublished). After the final selective amplification, carried out multiplexed with one unlabeled MseI primer and four different EcoRI primers labelled with four different fluorescent dyes, the amplifications products were separated on an ABI PRISM 3100 Genetic analyser (Applied Biosystems, Germany). The markers were scored using GeneMapper Software Version 3.7 (Applied Biosystems, Germany).

Data analysis: Linkage disequilibrium between marker pairs, expressed as r^2 (Flint-Garcia et al. 2003), and its significance level was determined with the program TASSEL (Zhang et al. 2006) based on the marker data from the LD

population. Recombination frequencies (rf) were determined by counting the number of recombinations between marker pairs in the segregating doubled haploid population and dividing the count by the number of informative genotypes. The frequencies were rounded to full percentages.

Results

Using 60 primer combinations a total of 1129 AFLP markers could be scored in the segregating doubled haploid population. From these markers, 775 proved to be polymorphic in the LD population, with allele frequencies ranging from 0.01 to 0.5. For the analysis of linkage disequilibrium in rapeseed, 471 markers with allele frequencies higher than 0.1 were selected. R^2 values were calculated for all possible 110,685 pair wise combinations of the 471 markers and the level of significance was determined for each combination. With a mean r^2 of 0.025 the overall linkage disequilibrium in the LD population was small. Furthermore, only 12,574 or 11.36% of the marker pairs were in significant ($p \leq 0.05$) LD. With a mean r^2 of 0.13 linkage disequilibrium was higher among the significant marker pairs, but still quite low in absolute terms.

In a genome like the rapeseed genome with 19 chromosome pairs, most marker pairs will not be physically linked but involve markers residing on different chromosomes. Fig. 1 shows the number of marker pairs at different levels of recombination. Between recombination frequencies of 4 to 34% the number of marker pairs per class varies between about 100 and 200, but at frequencies beyond 34% the numbers sharply increase, indicating the inclusion of an increasing fraction of unlinked markers. From the total of 12,574 marker pairs in significant LD, 2232 had recombination frequencies equal or less than 34%. These marker pairs, that can be considered as linked, had an average r^2 of 0.32 while, at 0.08, the mean r^2 of the remaining 10,342 significant marker pairs, that mainly include unlinked marker pairs, was much smaller.

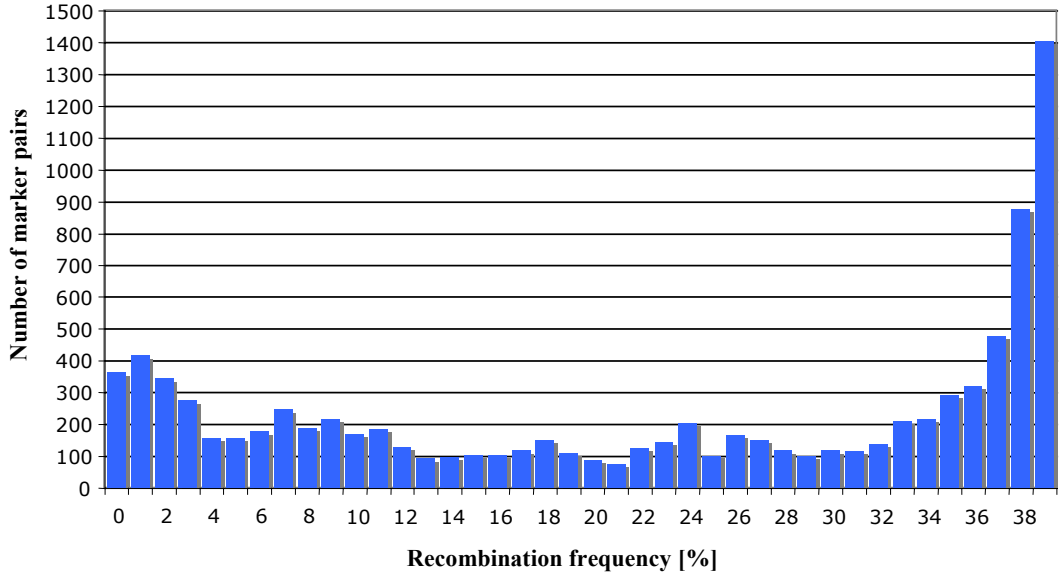

Fig. 1: Number of marker pairs at different recombination frequencies.

Fig. 2 shows a more detailed analysis of the relationship between recombination frequency and linkage disequilibrium in the LD population. For closely linked markers (rf= 0%) linkage disequilibrium is high with a mean r^2 of 0.51 and 81% of the marker pairs in significant LD. From there linkage disequilibrium decays rapidly with increasing recombination frequencies. At a recombination frequency of 7% the mean r^2 is already below 0.1 and the fraction of significant marker pairs has dropped to 42%.

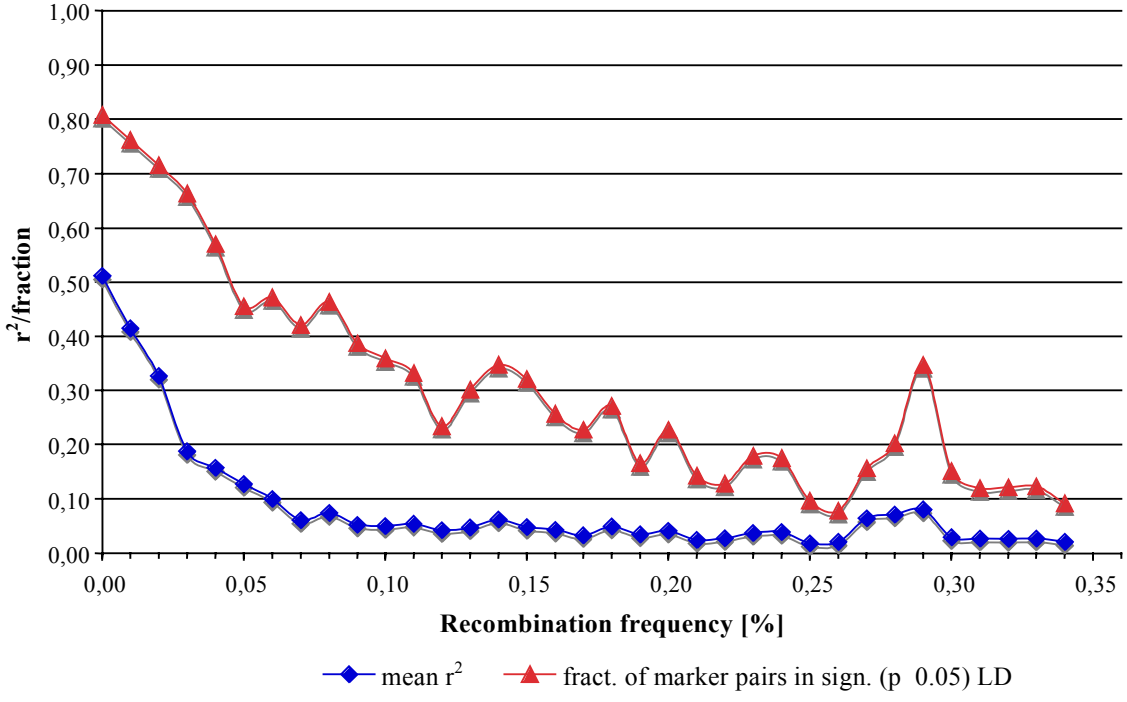

Fig. 2: Relationship between linkage disequilibrium and recombination frequency in winter rapeseed

Discussion

The 90 winter rapeseed lines that made up the LD population used to analyse linkage disequilibrium in rapeseed were selected from the materials of six major breeding companies. They include current varieties and breeding lines as well as older varieties, all of canola quality. As such, the LD population should be representative of the gene pool currently used in the breeding of canola quality winter rapeseed.

Our results show that the overall level of linkage disequilibrium in this gene pool is low, as indicated by the small percentage of marker pairs in significant LD and the low mean r^2 of these pairs. The comparison between the mean r^2 values

of linked and unlinked markers clearly shows that the major cause for linkage disequilibrium in the rapeseed genome is linkage. Furthermore, the very low mean r^2 of 0.08 of the unlinked markers, which is close to 0, indicates that linkage disequilibrium due to other factors that can cause LD, e. g. population stratification, selection or drift, is negligible in winter rapeseed.

This is surprising, considering the recent breeding history of rapeseed. During the last 40 years rapeseed was subject to two rounds of intense selection for two new quality traits: low erucic acid and low glucosinolate content, which were initially discovered in only a few genotypes in the 1960-ties and 70-ties. Current canola quality rapeseed is supposed to be derived from a limited number of crosses between these genotypes and elite breeding lines of that time (Becker et al. 1999). This genetic bottleneck and the following selection cycles were expected to have caused a high level of linkage disequilibrium in canola quality winter rapeseed.

There may be two reasons this level was not discovered. First, the rapeseed genome is distributed across 19 chromosome pairs. Most markers and genes are therefore physically unlinked and subject to independent assortment during gamete formation. This may have allowed any linkage disequilibrium due to genetic drift or selection to decay even within the limited number of breeding cycles that have passed since the introduction of the quality traits. Second, genetic analyses had shown early on that the low erucic acid phenotype is caused by only two genes (Harvey and Downey 1964; Kondra and Stefansson 1965). Later, QTL mapping identified just three major QTL to be responsible for low glucosinolate content (Uzunova et al. 1995). This means that the selection for the two quality traits affected only five regions of the rapeseed genome. Actually, the possibility of a higher level of linkage disequilibrium within and between these regions cannot be excluded by the present results, since this would not have shown up in the summary analysis carried out so far. A more detailed analysis of the regional distribution of linkage disequilibrium in the rapeseed genome will be conducted as soon as the AFLP markers have been localized on the genetic map of rapeseed.

The low level of linkage disequilibrium between unlinked markers is favourable for association analyses in rapeseed since it will preclude the occurrence of a high incidence of false positives, that is, associations not due to linkage. Furthermore, the rapid decay of LD within a few cM will give association analyses a much higher resolution than QTL mapping in segregating populations. On the other hand, useful levels of linkage disequilibrium seems to extend over at least one to two cM, indicating that global association analyses should be possible, although large numbers of markers will be required. Assuming a total length of the rapeseed genome of 2000 cM, a global analysis would need about 1000 to 2000 evenly spaced marker. Since markers on genetic maps are usually not evenly spaced and even at a recombination frequency of 0% not all the marker pairs are in significant LD, the actual number of markers needed will be higher. The analysis of the 60 AFLP primer combinations used in this study took about 5 month, but the major time consuming step was the manual control of all marker scores since the automatic scoring by the GeneMapper software alone would have introduced an unacceptable level of errors. If this obstacle could be overcome or if other marker types like SNPs, which give simpler banding patterns making them more amenable to automatic scoring, and the corresponding high throughput platforms are used, the analysis of several thousand markers in a reasonable time could be feasible.

Acknowledgement

We thank the breeding companies DSV, KWS, Limagrain-Nickerson, NPZ, SW Seed, and Syngenta for providing seeds of their varieties and breeding lines. This work was supported by the Deutsche Forschungsgemeinschaft (DFG), grant no. BE 1854/12-2.

References

Becker HC, Löptien H, Röbbelen G (1999) Breeding: An overview. In: C. Gómez-Campo (ed.). Biology of *Brassica* Coenospecies. Elsevier, Amsterdam, pp. 413-460

Darvasi A, Weinreb A, Minke V, Weller JI, Soller M (1993) Detecting Marker-QTL linkage and estimating QTL gene effect and map location using a saturated genetic map. Genetics 134:943-951

Ewens WJ, Spielman RS (2001) Locating genes by linkage and association. Theor Popul Biol 60:135–39

Flint-Garcia SA, Thornsberry JM, Buckler IV ES (2003) Structure of linkage disequilibrium in plants. Ann Rev Plant Biol 54:357-374

Harvey BL, Downey RK (1964) The inheritance or Erucic Acid content in Rapeseed (*Brassica napus*). Can J Plant Sci 44:104-111

Jannink J-L, Bink MC, Jansen RC (2001) Using complex plant pedigrees to map valuable genes. Trends Plant Sci 6:337–42

Kondra ZP, Stefansson BR (1965) Inheritance of Erucic and Eicosenoic acid content of Rapeseed oil *(Brassica napus*). Can J Genet Cytol 7:500-510

Uzunova M, Ecke W, Weißleder K, and 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

van Ooijen, JW (1992) Accuracy of mapping quantitative trait loci in autogamous species. Theor Appl Genet 84:803-811

Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Friters A, Pot J, Paleman J, Kuiper M and Zabeau M (1995). AFLP: a new technique for DNA fingerprinting. NAR 23: 4407-4414

Zhang Z., Bradbury PJ, Kroon DE, Casstevens TM, Buckler ES. A software package for association and diversity analysis in plants and animals. Poster, Plant and Animal Genome Conference, San Diego, Jan. 14 - 18, 2006. http://www.maizegenetics.net/index.php?page=bioinformatics/tassel/index.html