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

Wolfgang Ecke, Rosemarie Clemens and Heiko C. Becker

Department of Crop Sciences, Georg-August-University Göttingen, Von-Siebold-Str. 8, 37075 Göttingen, Germany

Objectives:

To analyse the extent and structure of linkage disequilibrium in canola qualtity winter rapeseed

because

- the prospects of QTL mapping by global association analysis are strongly dependent on the linkage disequilibrium in the plant material analysed
- the breeding history of rapeseed with two rounds of intensive selection for two quality traits, erucic acid and glucosinolate content, may have left an imprint on linkage disequilibrium in canola quality winter rapeseed

Approach:

- analyse > 2000 AFLP marker in a population of 87 canola quality winter rapeseed varieties and breeding lines (LD population) mainly derived from six major European breeding companies and in a segregating DH population (mapping population) from a cross between the winter rapeseed variety 'Express' and a resynthesized rapeseed, 'R53'
- map all markers that (1) are polymorphic in both populations and (2) which rare alleles are present in at least 10 genotypes of the LD population (allele frequency > 0.11)
- estimate linkage disequilibrium (r^2) between all possible marker pairs in the LD population using TASSEL (Zhang et al. 2006)

Results: marker analysis

132	AFLP primer combinations analysed in both populations
2161	markers segregating in the mapping population identified
1470	of these were polymorphic in the LD population
820	polymorphic markers mapped in the DH population
787	markers showed allele frequencies > 0.11 in the LD population and were used for the analysis of linkage disequilibrium

Results: genetic map

Link. gr.	No. of markers	No. of loci	Length [cM]	Link. gr.	No. of markers	No. of loci	Length [cM]
N1	26	20	105.9	N12	37	23	85.3
N2	19	16	151.7	N13	64	47	238.8
N3	39	25	136.3	N14	47	22	144.7
N4	26	18	62.4	N15	37	23	150.4
N5	42	29	191.0	N16	34	22	100.7
N6	44	20	127.6	N17	43	27	114.2
N7	36	16	78.0	N18	25	15	99.0
N9	55	25	149.2	N19	55	40	136.0
N10	26	13	76.7	LG5	24	7	20.5
N11	118	50	150.6	LG10	23	15	77.2
				Sum	820	473	2396.2

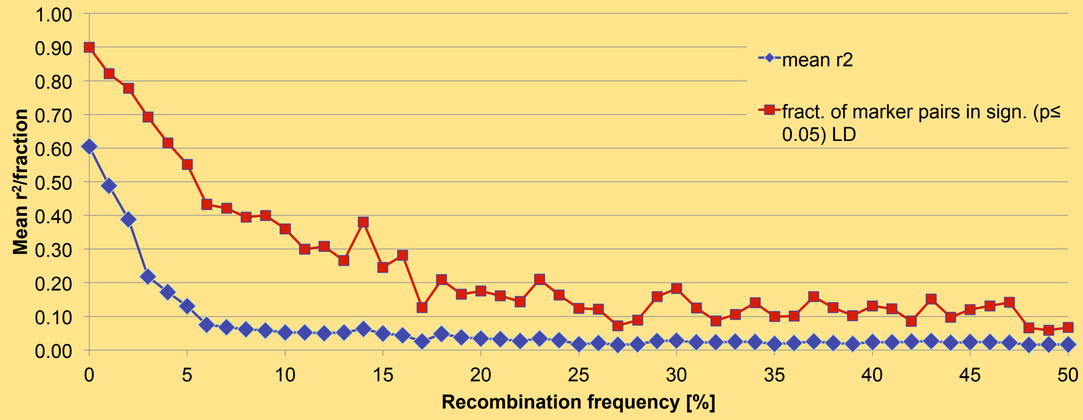
A total of 820 markers were mapped at 473 loci on 20 linkage groups, giving a genetic map covering 2396 cM of the rapeseed genome. For 18 linkage groups the corresponding linkage groups in reference maps could be identified by map alignments. These linkage groups were labelled according to the N nomenclature of linkage groups in rapeseed (Parkin et al. 1995). The remaining two linkage groups were labelled LG5 and LG10

Results: estimation of linkage disequilibrium

	N	mean r^2
All marker pairs	309,291	0.027
Marker pairs in significant (p≤ 0.05) LD	37,937 (12.3%)	0.130
Linked marker pairs	19,651	0.127
Linked marker pairs in significant (p≤ 0.05) LD	6,883 (35.0%)	0.335
Unlinked marker pairs	289,640	0.020
Unlinked marker pairs in significant (p≤ 0.05) LD	31,054 (10.7%)	0.085

With a mean r^2 of 0.027 the overall linkage disequilibrium in the LD population was small. Furthermore, only 12.3% of the marker pairs were in significant LD. Although, at a mean r^2 of 0.13, linkage disequilibrium was higher among these marker pairs, it was still quite low in absolute terms. Showing a mean r^2 of 0.127 and 35% of the marker pairs in significant LD with a mean r^2 of 0.335 levels of linkage disequilibrium were much higher among linked markers than between unlinked markers where the average linkage disequilibrium was at an r^2 of 0.02 and even between the 10.7% of marker pairs in significant LD did not exceed 0.085.

Results: relationship between linkage disequilibrium and recombination frequency in winter rapeseed



A close relationship was observed in the LD population between the average linkage disequilibrium and the recombination frequency of the marker pairs. For closely linked markers (rf= 0%) linkage disequilibrium is high with a mean r^2 of 0.61 and 90% of the marker pairs in significant LD. From there linkage disequilibrium decays rapidly with increasing recombination frequencies. At a recombination frequency of 6% the mean r^2 is already below 0.1 and the fraction of significant marker pairs has dropped to 43%.

Discussion: levels and structure of linkage disequilibrium

The results show that the overall level of linkage disequilibrium in canola quality winter rapeseed 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 0.085 of the unlinked marker pairs in significant LD indicates that linkage disequilibrium due to other factors that can cause LD, e. g. population structure, selection or genetic drift, is negligible in winter rapeseed.

This is surprising, considering the recent breeding history of rapeseed. During the last 40 years rapeseed was subject of 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 1960s and 70s. 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.

On the other hand, the rapeseed genome is distributed across 19 chromosome pairs. Most loci 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. In addition, 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.

Discussion: prospect for association analyses in canola quality rapeseed

The low level of linkage disequilibrium between unlinked markers is favorable 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, where confidence intervals for QTL positions are usually in the range of several cMs up to tens of cMs. On the other hand, useful levels of linkage disequilibrium seem to extend over at least one to two cM, indicating that global association analyses should be possible, although rather large numbers of markers will be required.

Literature

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

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

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

Parkin IAP, Sharpe AG, Keith DJ, Lydiate DJ (1995) Identification of the A and C genomes of amphidiploid Brassica napus (oilseed rape). Genome 38:1122-1131

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

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.