

Extent and structure of linkage disequilibrium in canola quality winter rapeseed (*Brassica napus* L.)

Wolfgang Ecke · Rosemarie Clemens ·
Nora Honsdorf · Heiko C. Becker

Received: 27 May 2009 / Accepted: 12 November 2009 / Published online: 2 December 2009
© The Author(s) 2009. This article is published with open access at Springerlink.com

Abstract Linkage disequilibrium was investigated in canola quality winter rapeseed to analyze (1) the prospects for whole-genome association analyses and (2) the impact of the recent breeding history of rapeseed on linkage disequilibrium. A total of 845 mapped AFLP markers with allele frequencies ≥ 0.1 were used for the analysis of linkage disequilibrium in a population of 85 canola quality winter rapeseed genotypes. A low overall level of linkage disequilibrium was found with a mean r^2 of only 0.027 over all 356,590 possible marker pairs. At a significance threshold of $P = 2.8 \times 10^{-7}$, which was derived by a Bonferroni correction from a global α -level of 0.1, only 0.78% of the marker pairs were in significant linkage disequilibrium. Among physically linked marker pairs, the level of linkage disequilibrium was about five times higher with more than 10% of marker pairs in significant linkage disequilibrium. Linkage disequilibrium decayed rapidly with distance between linked markers with high levels of linkage disequilibrium extending only for about 2 cM. Owing to the rapid decay of linkage disequilibrium with distance association analyses in canola quality rapeseed will have a significantly higher resolution than QTL analyses in segregating populations by interval mapping, but much larger number of markers will be necessary to cover

the whole genome. A major impact of the recent breeding history of rapeseed on linkage disequilibrium could not be observed.

Introduction

QTL mapping in segregating populations derived from biparental crosses has become a common tool in the analysis of quantitative traits in plants. Although this approach allows the identification of loci contributing to a quantitative trait and the estimation of their effects, it has some inherent disadvantages. First, mapping populations have to be developed specifically for the QTL mapping. In addition, only the allelic diversity sampled in the two parents of the cross can be analyzed, that is, QTL where both parents have the same allele cannot be detected. Furthermore, due to the limited number of recombination events available in a segregating population, the resolution is somewhat limited, 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).

An alternative approach could be association analysis or linkage disequilibrium (LD) mapping using natural populations or, in the case of crop plants, collections of varieties and breeding lines. Owing to the higher number of recombination events in such a material, a higher resolution could be achieved than in segregating populations (Ewens and Spielman 2001; Jannink et al. 2001). In addition, the allelic diversity would not be limited to the diversity occurring between two parental lines. Furthermore, this approach could be easily integrated into the breeding process for new crop varieties. For example, Kraakman et al. (2004) used data from official Danish variety trials and mapped AFLP markers to localize QTL

Communicated by M. Kearsey.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-009-1221-0) contains supplementary material, which is available to authorized users.

W. Ecke (✉) · R. Clemens · N. Honsdorf · H. C. Becker
Department of Crop Sciences, Georg-August-University
Göttingen, Von-Siebold-Str. 8, 37075 Göttingen, Germany
e-mail: wecke@gwdg.de

for yield and yield stability in modern two-row spring barley cultivars.

Association analysis is based on the linkage disequilibrium between linked loci and is strongly dependent on the extent and structure of the linkage disequilibrium in the population analyzed. Linkage disequilibrium, the non-random association of alleles at different loci, is created by mutation, admixture between genetically distinct populations, selection and genetic drift, and decays by genetic recombination (Flint-Garcia et al. 2003). Accordingly, the linkage disequilibrium in a population is dependent on the population history and the mating system of the species. Linkage disequilibrium has been analyzed in a number of plant species, either globally, using molecular markers, or locally by sequencing specific genomic segments of up to several hundred kilobytes. In maize, an allogamous species, Tenaillon et al. (2001) observed a rapid decay of linkage disequilibrium within 100–200 bp in a genetically broad material of inbred lines and exotic landraces. On the other hand, in *Arabidopsis thaliana*, an autogamous species, linkage disequilibrium extended over about 1 cM or 250 kb in a global population of 20 accessions from different parts of the world and the decay of linkage disequilibrium with distance was even lower in local populations (Nordborg et al. 2002). In sugar cane, where elite varieties are propagated clonally, Jannoo et al. (1999) observed linkage disequilibrium between RFLP markers to extend over up to 10 cM. Nevertheless, analyzing only inbred lines of maize Remington et al. (2001) found linkage disequilibrium extending over 1.5 kb and Rafalski (2002) reported that linkage disequilibrium in elite germplasm of maize extends over more than 100 kb. Analyzing linkage disequilibrium with SSR markers in the flint and dent germplasm groups used in maize hybrid breeding in Europe Stich et al. (2005) observed very high levels of linkage disequilibrium with 55 and 48% of the linked and unlinked marker pairs, respectively, in significant LD in the flint germplasm group and an average length of LD blocks of 26 cM. The levels of linkage disequilibrium were even higher in the dent germplasm group. Conversely, Kim et al. (2007) observed a rapid decay of linkage disequilibrium within 10 kb in a sample of 19 *A. thaliana* accession. The results from the different populations indicate that population history may be more important than the mating system in determining the level of linkage disequilibrium in a population. A strong influence of population history was also observed in barley and soybean (Caldwell et al. 2006; Hyten et al. 2007). In both crop plants, linkage disequilibrium extends farthest in elite breeding materials while decaying the most rapidly in wild relatives with landraces taking an intermediate position. When comparing SSR and AFLP markers, Stich et al. (2006) also found a strong influence of the marker type on the detection of

linkage disequilibrium. The level of linkage disequilibrium detected in European maize inbred lines was much higher with SSR markers than with AFLP markers, presumably because the former distinguish between more alleles than the latter.

Rapeseed is a partially allogamous species that is bred like an autogamous species with controlled crosses followed by several generations of selfing to develop new varieties. It gained its current importance as a major oil crop in temperate regions only after two rounds of intense selection for two new quality traits: zero erucic acid and low glucosinolate content, which were initially introduced into the breeding material from one donor genotype each in the 1960s and 1970s, respectively. Current elite breeding materials produce seed oil free from erucic acid and a meal low in glucosinolates—a quality termed ‘canola’—and are supposed to be derived from a limited number of crosses between the original genotypes with these quality traits and breeding lines of that time (Becker et al. 1999). Accordingly, the introduction of the two traits may have constituted a genetic bottleneck in the breeding history of rapeseed that, together with the following intense selection for the new traits, could have had a major impact on the level and structure of linkage disequilibrium in current canola quality rapeseed materials.

In rapeseed, QTL mapping in segregating populations is well established and has been used in a number of studies to analyze quality traits such as oil content (Ecke et al. 1995; Zhao et al. 2005; Delourme et al. 2006; Qiu et al. 2006; Zhao et al. 2006), glucosinolate content (Toroser et al. 1995; Uzunova et al. 1995), tocopherol content (Marwede et al. 2005), phytosterol and sinapate ester content (Amar et al. 2008), and the fatty acid composition of the seed oil (Thormann et al. 1996; Zhao et al. 2008) as well as disease resistances such as blackleg (Pilet et al. 1998) or heterosis (Radoev et al. 2008). So far, no study has been published on the application of association analysis in rapeseed or about linkage disequilibrium in rapeseed populations. The objective of this study was to determine the extent and structure of linkage disequilibrium in canola quality winter rapeseed to (1) analyze the prospects for association analysis in current elite breeding materials of this crop plant and (2) to elucidate the impact the introduction of the ‘canola’ quality has had on the linkage disequilibrium in this material.

Materials and methods

Plant materials

Linkage disequilibrium was analyzed in a set of 85 Northern European canola quality winter rapeseed varieties

and breeding lines (Table 1), further called LD population. For the analysis, one individual plant per variety was used. For genetic mapping, a mapping population of 94 doubled haploid lines derived from one F_1 plant of a cross between the winter rapeseed variety ‘Express’ and a resynthesized

rapeseed, ‘R53’, was used. This population had already been used to develop a genetic map in rapeseed comprised mainly of SSR markers (Radoev et al. 2008).

DNA preparation and AFLP analysis

DNA was prepared from 0.1 g of leaf material of 3 weeks old greenhouse grown plants using Nucleon PhytoPure extraction kits (RPN8510, GE Healthcare Bio-Sciences AB, Uppsala, Sweden) following the manufacturer’s instructions.

The *EcoRI* primers used in AFLP analysis were labeled with one of the following four fluorescent dyes: (6, 5) FAM, NED, VIC, or PET (Applied Biosystems, Darmstadt, Germany). AFLP analyses were carried out following the protocol of Vos et al. (1995) modified for multiplexing in the PCR according to F. Kopisch-Obuch (personal communication): 250 ng DNA were digested in 30 μ l RL buffer (10 mM Tris–acetate, 10 mM Mg–acetate, 50 mM K–acetate, 5 mM DTT, pH 7.5) with 4 U *EcoRI* (Fermentas, St. Leon-Rot, Germany) and 4 U *MseI* (New England Biolabs, Frankfurt, Germany) for 1.5 h at 37°. After adding 10 μ l of a mix containing 5 pmol *EcoRI* adapter, 50 pmol *MseI* adapter, 1 mM ATP and 1 U T4-DNA ligase (Promega, Mannheim, Germany) in RL buffer, DNA and adapters were ligated in a time series of different temperatures (3 h 10 min 37°, 3 min 33.5°, 3 min 30°, 4 min 26° and finally 15 min 22°). The final restriction–ligation product (RL) was diluted 1:5 with HPLC grade water. For pre-amplification, 8 μ l of the diluted RL was added to 12 μ l of a reaction mixture, giving final concentrations of 1 \times *Taq* buffer (Solis Biodyne, Tartu, Estonia, Reaction buffer B), 3.125 mM MgCl₂, 0.45 mM dNTPs, 10 pmol *EcoRI*+1 primer, 9 pmol *MseI*+1 primer and 2.5 U *Taq* DNA polymerase (FIREPol, Solis Biodyne). The pre-amplification was carried out in a Biometra T1 Thermocycler (Biometra GmbH, Göttingen, Germany) with the following program: 94° for 30 s, 20 cycles of 94° for 30 s, 56° for 30 s and 72° for 2 min, and a final 5 min at 72°. The pre-amplification product was diluted 1:10 with HPLC grade water. The final AFLP amplification used 6 μ l of the diluted pre-amplification product in a total reaction volume of 20 μ l containing 1 \times *Taq* buffer, 0.36 mM dNTPs, 3.125 mM MgCl₂, 1 U *Taq* polymerase, 7 pmol *MseI*+3 primer, 2 pmol of (6, 5)FAM labeled *EcoRI*+3 primer, 2 pmol of VIC labeled *EcoRI*+3 primer, 4 pmol of NED labeled *EcoRI*+3 primer, and 6 pmol of PET labeled *EcoRI*+3 primer. The protocol for the Thermocycler was as follows: 1 cycle of 94° for 1 min, 65° for 30 s, and 72° for 2 min, 12 cycles of 94° for 30 s, 64.2° for 30 s and 72° for 2 min, 25 cycles of 94° for 30 s, 56° for 30 s and 72° for 2 min, and finally 72° for 5 min.

Table 1 Origin of the 85 canola quality varieties and breeding lines used in the analysis of linkage disequilibrium in rapeseed

Variety	Breeder	Variety	Breeder
Alesi	KWS	Magnum	Syngenta
Remy	KWS	Madrigal	Syngenta
Robust	KWS	Laser	Syngenta
Alaska	KWS	Fortis	Syngenta
Pirola	KWS	Smart	Syngenta
Adder	KWS	Roxet	Syngenta
Milena	KWS	NK Bravour	Syngenta
Allure	KWS	NK Fair	Syngenta
Agalon	KWS	Aviso	SW Seed
K615	KWS	Sansibar	SW Seed
KW1519	KWS	SWGospel	SW Seed
Picasso	KWS	Verona	SW Seed
Lord	KWS	Tenor	SW Seed
KW3077	KWS	Expert	SW Seed
Rodeo	KWS	Musette	SW Seed
Rapid	Limagrain-Nickerson	Kvintett	SW Seed
Boston	Limagrain-Nickerson	Falstaff	SW Seed
Escort	Limagrain-Nickerson	SW Sinatra	SW Seed
Montego	Limagrain-Nickerson	Viking	NPZ
Ontario	Limagrain-Nickerson	Aragon	NPZ
Pacific	Limagrain-Nickerson	Aurum	NPZ
Savannah	Limagrain-Nickerson	Lorenz	NPZ
Missouri	Limagrain-Nickerson	Baros	NPZ
Manitoba	Limagrain-Nickerson	Rasmus	NPZ
Ladoga	Limagrain-Nickerson	Gefion	NPZ
Atlantic	Limagrain-Nickerson	Nugget	NPZ
Cooper	Limagrain-Nickerson	Zephir	NPZ
Licapo	DSV	SLM 0413	NPZ
Capitol	DSV	SLM 0512	NPZ
Idol	DSV	LSF 0519	NPZ
Vivol	DSV	HSL 1032	NPZ
Bristol	DSV	Campari	NPZ
Lirajet	DSV	Caramba	NPZ
Lisabeth	DSV	Express 617	NPZ
Lipid	DSV	Prince	NPZ
Lipton	DSV	Wotan	NPZ
Lisek	DSV	Amor	Petersen/Raps GbR
Contact	DSV	Orlando	Saaten Union
Columbus	DSV	Pollen	Adrien Momont
Lion	DSV	Ascona	SW Seed
Oase	DSV	Duell	Raps GbR
Apex	Syngenta	Jessica	–
Recital	Syngenta		

KWS KWS SAAT AG; DSV Deutsche Saatveredelung AG; NPZ Norddeutsche Pflanzenzucht Hans-Georg Lembke KG

The AFLP products were separated on an ABI PRISM 3100 Genetic Analyser (Applied Biosystems) using 50-cm capillary arrays and GeneScan-500 LIZ size standard (Applied Biosystems). GeneMapper v3.7 software (Applied Biosystems) was used for a semi-automatic marker scoring. Since in GeneMapper v3.7's output, AFLP primer combinations are written as markers and the actual AFLP markers as alleles of these markers a Perl script, 'Extract_marker', was developed to transform GeneMapper's output into a marker matrix. The primer combinations used, the labels of the *EcoRI* primers and the numbers of markers identified with the different primer combinations are listed in Table S1.

Genetic mapping

Genetic mapping of the AFLP markers was based on a framework map previously established in the mapping population by Radoev et al. (2008). Using the program MAPMAKER/EXP V.3.0b, the new markers were assigned to linkage groups by the 'near' command with an LOD threshold of 4.0 and a maximum recombination frequency of 0.4. Linkage groups were then reanalyzed using the 'order' command. Finally, markers that could not be placed by the 'order' command were manually placed using the 'try' command. Double crossovers were identified using MAPMAKER's 'genotype' command and were rechecked in the trace files and, if necessary, corrected, followed by a remapping of the affected markers. Markers with high numbers of double crossovers and markers with strongly disturbed segregations where one class was represented by fewer than 25 genotypes were excluded from the mapping. Linkage groups were named according to the N-nomenclature proposed by Parkin et al. (1995). Recently, a new nomenclature was proposed by the Steering Committee of the Multinational Brassica Genome Project (<http://www.brassica.info>). In this nomenclature, A1–A10 correspond to N1–N10 and C1–C10 to N11–N19.

Analysis of linkage disequilibrium

For the analysis of linkage disequilibrium, only markers with allele frequencies in the LD population of 0.1 or larger were used. This discrimination against rare alleles is justifiable because the information from them is neither useful in the analysis of linkage disequilibrium nor in association analysis. R^2 values of linkage disequilibrium for all pairwise marker combinations and the corresponding significance levels (P values) were calculated using the program TASSEL V.2.0.1 (Zhang et al. 2006). Recombination frequencies between marker pairs were calculated by a Perl script and added to the corresponding rows of the LD table generated by TASSEL. All further statistical and graphical

analyses were carried out in Microsoft Office Excel 2007. The threshold for declaring linkage disequilibrium between two markers significant was derived by a Bonferroni correction from a global α -level of 0.1, resulting in a per test threshold of $P = 2.8 \times 10^{-7}$.

Results

Marker analysis and map construction

By using 132 primer combinations 2161 AFLP markers could be scored in the mapping population. In the LD population, 1,463 of these markers were also polymorphic and 898 showed allele frequencies equal to or larger than 0.1. Of the markers with allele frequencies ≥ 0.1 in the LD population 845 could be mapped in the mapping population. The AFLP markers were mapped within a framework of 167 markers from the earlier map that had been established in the mapping population by Radoev et al. (2008). After the initial map construction, the markers were distributed across 21 linkage groups. By mapping some additional markers from the full set of 2,161 markers, the map could be consolidated in 19 linkage groups, a number corresponding to the 19 chromosomes of the haploid rapeseed genome. Based on map alignments using the SSR markers from the earlier map, 18 of the linkage groups could be named according to the N-nomenclature of rapeseed linkage groups. The last linkage group was designated as N8 by exclusion. The final map (Table 2) has a length of 2,473 cM and comprises 1,032 markers distributed across 551 map positions. Included are 865 new AFLP markers that cover 2,345 cM (95%) of the total map. Individual linkage groups range in length from 77 to 242 cM, holding between 27 and 132 markers. The full map is listed in Table S2.

Levels of linkage disequilibrium

Linkage disequilibrium in canola quality winter rapeseed was analyzed using pairwise combinations of the 845 AFLP markers with allele frequencies ≥ 0.1 in the LD population. With a mean r^2 value of only 0.027 over all 356,590 possible pairwise combinations, the overall level of linkage disequilibrium in the rapeseed genome is very low (Table 3). This conclusion is reinforced by the observation that only 0.78% of marker pairs are in significant LD. With a mean r^2 of 0.122 linkage disequilibrium among physically linked marker pairs, that is pairs where both markers are on the same linkage group, is nearly five times higher than the overall mean. Furthermore, 11.58% of these marker pairs are in significant LD and with a count of 2,658 represent the vast majority of marker pairs in

Table 2 Summary of the genetic map with the AFLP markers used in the analysis of LD in rapeseed

Linkage group	No. of markers	No. of map pos.	Length (cM) ^a	New AFLP markers	Framework markers ^b
N1	33	25	118	26	7
N2	28	21	141	20	8
N3	46	26	122	39	7
N4	32	19	78	27	5
N5	50	30	174	43	7
N6	54	26	135	45	9
N7	41	17	77	37	4
N8	27	16	77	24	3
N9	65	26	133	55	10
N10	37	21	83	27	10
N11	132	49	132	122	10
N12	52	30	161	45	7
N13	82	52	242	67	15
N14	63	28	147	53	10
N15	51	30	142	39	12
N16	44	28	130	35	9
N17	54	34	124	43	11
N18	74	32	140	62	12
N19	67	41	117	56	11
Sum	1032	551	2473	865	167

^a Recombination frequencies were transformed to centimorgan according to the Kosambi mapping function

^b Markers from the map of Radoev et al. (2008)

Table 3 Number of marker pairs and average level of LD (mean r^2) in different classes of marker pairs

Class ^a	All pairs in the class		Pairs in significant LD at $P = 2.8 \times 10^{-7}$	
	n (%) ^b	Mean r^2	n (%) ^c	Mean r^2
All	356,590 (100.00)	0.027	2,775 (0.78)	0.722
Linked	22,951 (6.44)	0.122	2,658 (11.58)	0.729
Unlinked	333,639 (93.56)	0.020	117 (0.04)	0.544

^a All: all marker pairs, linked: pairs of markers from the same linkage group, unlinked: pairs where the markers are on different linkage groups

^b Percentage from all (356,590) marker pairs

^c Percentage from all marker pairs in the class

significant LD indicating that the major determinant of linkage disequilibrium in the rapeseed genome is genetic linkage. Accordingly, only 117 of the unlinked marker pairs are in significant LD and at 0.544, the mean r^2 of these marker pairs is still lower than the mean r^2 of 0.729 of the linked marker pairs in significant LD.

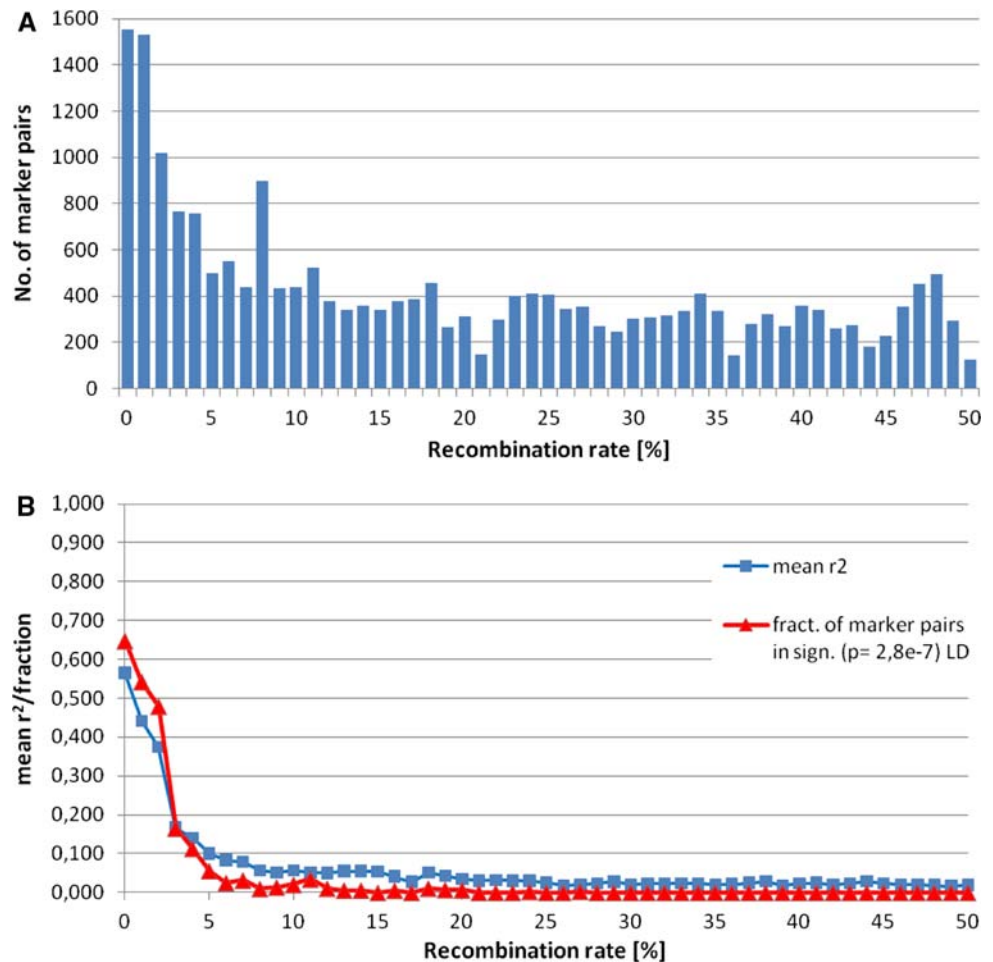
Structure of linkage disequilibrium

To investigate the structure of linkage disequilibrium in the rapeseed genome, the dependency of linkage disequilibrium on distance was analyzed among the physically linked marker pairs. The number of marker pairs at recombination rates from 0 to 50% ranges from 126 to 1,554 (Fig. 1a) providing a solid base for this analysis. Among the linked marker pairs, linkage disequilibrium decays rapidly with

distance (Fig. 1b). Closely linked marker pairs at recombination frequencies of 0–2% show high levels of linkage disequilibrium with mean r^2 values ranging from 0.566 to 0.374, but at a recombination rate of 5%, the mean r^2 is already down to 0.1 and at high distances, it is not significantly different from the overall mean of 0.027. Likewise, the fraction of marker pairs in significant LD decays from 65 to 48% for closely linked marker pairs to 6% at a recombination rate of 5% and 1–3% at intermediate recombination rates (6–20%). With the exception of two marker pairs at 24 and 27%, no marker pairs in significant LD are found at higher recombination rates.

The rapid decay of linkage disequilibrium is also apparent when looking at the distribution of linkage disequilibrium across individual linkage groups (Fig. 2a). Colors indicative of high LD are close to the diagonal

Fig. 1 Relationship between marker distance and linkage disequilibrium. **a** Number of marker pairs at different distances. **b** Average linkage disequilibrium and fraction of marker pairs in significant LD at different distances. Distances between markers of physically linked pairs are given as recombination rates determined in the mapping population and rounded to full percentages. Average linkage disequilibrium is presented as the mean r^2 of all linked marker pairs at a given recombination rate



representing closely linked marker pairs. Nevertheless, there are some regions in the rapeseed genome where linkage disequilibrium extends over somewhat larger distances. For example, on linkage group N5 between markers E45M57-152E and E42M56-162R, there are marker pairs in significant LD as far apart as 16 cM (Fig. 2b), indicating that the decay of linkage disequilibrium with distance may vary between different regions of the rapeseed genome. The LD maps of the full set of 19 linkage groups are shown in Table S3.

The 117 unlinked marker pairs in significant LD are distributed across 20 pairs of genomic segments located on 11 linkage groups (Table 4). The number of involved markers per segment ranges from 1 to 49 and the length of the segments from 0 (one marker or a group of co-segregating markers) to 10.8 cM, in total covering only 34.5 cM of the rapeseed genome. The 20 segment pairs can be subdivided into three classes according to the pattern of associated markers. In the most frequent class, only one marker on one linkage group is in significant LD with several markers on the second linkage group. In the second class only one marker pair in significant LD

is present on the two linkage groups. In the third class, that was only observed once, several markers on one linkage group are in significant LD with several markers on the second linkage group. The full LD maps of the linkage group pairs listed in Table 4 are shown in Table S4.

Discussion

In canola quality winter rapeseed a very low level of linkage disequilibrium was found with a mean r^2 of only 0.027 and a fraction of marker pairs in significant LD ($P = 2.8 \times 10^{-7}$) of only 0.78%. High levels of linkage disequilibrium were limited to closely linked markers at distances of 0–2 cM. Distances between markers have been determined in a doubled haploid mapping population of 94 genotypes. In rapeseed, populations used for mapping have a size ranging from 50 (Parkin et al. 1995) to 445 (Delourme et al. 2006) genotypes. The mapping population used here is near the lower bound of this range. As a consequence, a large fraction of co-segregating marker

Fig. 2 LD maps of individual linkage groups of the genetic map. **a** LD map of linkage group N1. **b** Segment of the LD map of linkage group N5. Below the diagonal the level of linkage disequilibrium between individual marker pairs is indicated, above the diagonal the significance level of the linkage disequilibrium ($P = 2.8 \times 10^{-7}$, $P = 1.4 \times 10^{-7}$, not significant). Groups of co-segregating markers are framed in green

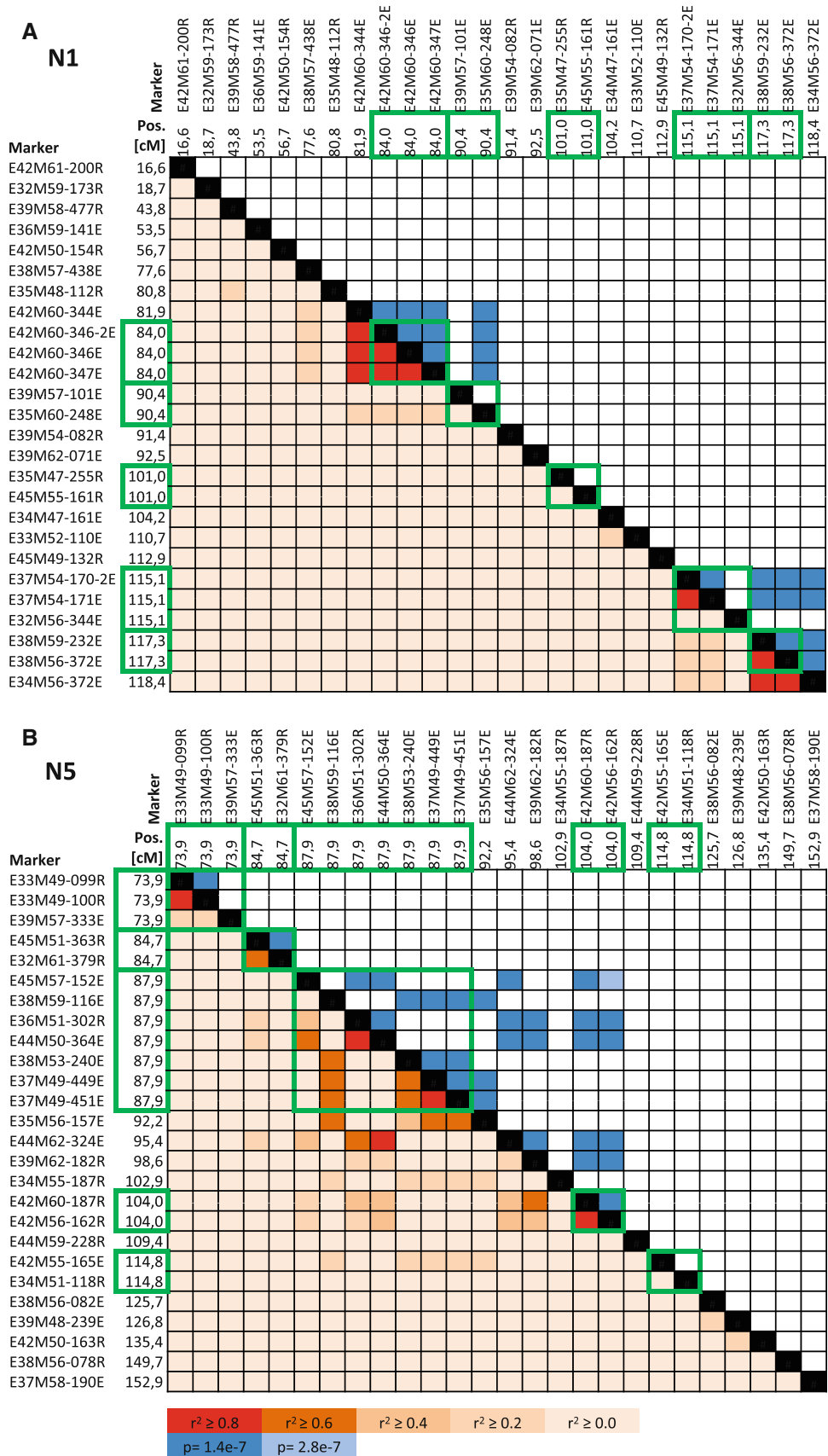


Table 4 Distribution of unlinked marker pairs in significant LD across the linkage groups of the genetic map

No	Class ^a	Mean r^2	LG 1	#Markers	Position/length (cM) ^b			LG 2	#Markers	Position/length (cM)		
					Start	End	Length			Start	End	Length
1	1	0.930	N09	10	83.0	87.3	4.3	N15	1	80.8	80.8	0.0
2	1	0.552	N10	1	25.8	25.8	0.0	N11	49	18.2	26.9	8.7
3	2	0.656	N10	1	25.8	25.8	0.0	N15	1	80.8	80.8	0.0
4	1	0.600	N10	1	25.8	25.8	0.0	N15	2	62.5	62.5	0.0
5	1	0.620	N09	10	83.0	87.3	4.3	N10	1	25.8	25.8	0.0
6	2	0.636	N06	1	76.6	76.6	0.0	N19	1	45.3	45.3	0.0
7	1	0.315	N05	1	109.4	109.4	0.0	N11	6	44.0	54.8	10.8
8	2	0.461	N12	1	60.3	60.3	0.0	N19	1	71.9	71.9	0.0
9	3	0.349	N03	5	0.0	2.2	2.2	N06	4	69.1	71.2	2.1
10	1	0.359	N11	1	19.3	19.3	0.0	N19	4	32.4	35.6	3.2
11	1	0.376	N06	1	17.8	17.8	0.0	N17	7	16.4	17.5	1.1
12	2	0.406	N03	1	76.4	76.4	0.0	N13	1	83.6	83.6	0.0
13	2	0.402	N13	1	25.7	25.7	0.0	N15	1	91.6	91.6	0.0
14	2	0.388	N01	1	91.4	91.4	0.0	N09	1	88.4	88.4	0.0
15	1	0.377	N01	4	81.9	84.0	2.1	N05	1	174.1	174.1	0.0
16	1	0.327	N08	4	51.5	51.5	0.0	N19	1	27.0	27.0	0.0
17	1	0.324	N07	2	70.1	70.1	0.0	N13	1	106.1	106.1	0.0
18	2	0.302	N01	1	53.5	53.5	0.0	N18	1	74.7	74.7	0.0
19	2	0.298	N01	1	118.4	118.4	0.0	N10	1	22.6	22.6	0.0
20	2	0.376	N07	1	76.7	76.7	0.0	N16	1	0.0	0.0	0.0

^a 1: one marker on one linkage group is in sign. LD with two or more markers on the second linkage group, 2: one marker on one linkage group is in sign. LD with one marker on the second linkage group, 3: several markers on one linkage group are in sign. LD with several markers on the second linkage group

^b Position and length of the genomic segments containing the markers in sign. LD on the two linkage groups

pairs were observed (first column in Fig. 1a). Because the markers in these pairs are not truly at the same position in the genome, using a larger mapping population would have resulted in the detection of recombinations between some of these markers, predominantly between markers that are somewhat farther apart than others, placing these marker pairs at distances larger than 0 cM. Because there is clearly a relationship between distance and average LD, the remaining co-segregating marker pairs would have a higher mean LD, indicating that the true decay of LD between 0 and 1 cM should be steeper than that shown in Fig. 1b. Also, using a mapping population of only 94 genotypes has incurred a larger statistical error on the estimates of recombination frequencies between individual marker pairs. On the other hand, the decay of LD with distance was analyzed by averaging LD over many marker pairs at each distance. Especially in the critical range between 1 and 5 cM, the number of marker pairs in each distance class is rather large, ranging from 1,531 to 501. The mean LD values calculated over so many marker pairs should represent good estimates of the average LD at different distances in the rapeseed genome.

The observation of high levels of significant LD over distances of 0–2 cM is, at the level of map distances, similar to the situation reported by Nordborg et al. (2002) in *Arabidopsis* where linkage disequilibrium extended over about 1 cM. In *Arabidopsis* 1 cM corresponds to 250 kb (Nordborg et al. 2002). In rapeseed with its larger genome of 1.2×10^9 bp (Arumuganathan and Earle 1991) and a total map length of about 2,400 cM one cM should, on average, correspond to about 500 kb. This means that at the level of physical distance linkage disequilibrium extends over larger distances in rapeseed than in *Arabidopsis*. This is also true when compared with maize where linkage disequilibrium decays over distances of 100 bp–100 kb, depending on the population analyzed (Remington et al. 2001; Tenaillon et al. 2001; Rafalski 2002). Similar levels of linkage disequilibrium as in rapeseed have been found in elite cultivars of soybean where high LD extended over at least 500 kb, but the decay of linkage disequilibrium was faster in landraces and in the wild relative *Glycine soja* LD decayed within 36–77 kb (Hyten et al. 2007). Likewise, in Asian rice, *Oryza sativa*, linkage disequilibrium extending over more than 500 kb was found in the *temperate*

japonica variety group, but in the *tropical japonica* and the *indica* variety group LD decayed within 150 and 75 kb, respectively (Mather et al. 2007).

The differences in the extent of linkage disequilibrium between the different soybean and rice groups have been attributed to increased self-fertilization in elite breeding materials and differences in outcrossing and recombination rates, respectively. The comparatively far reach of significant linkage disequilibrium between linked markers in canola quality winter rapeseed, extending over 500–1,000 kb, may be due to the genetic bottleneck that current elite breeding materials passed through during the introduction of the canola quality. On the other hand, linkage disequilibrium created by genetic drift and a strong selection should not be limited to linked markers, but only 20 pairs of small genomic segments with 117 unlinked marker pairs in significant LD were found. Half of the segment pairs followed a peculiar pattern (class 1, Table 4) where one marker on one linkage group was in significant LD with several markers on the second linkage group which, in many cases, showed significant LD among themselves. Nine of the remaining ten pairs of segments were of class 2 where a single marker on one linkage group was in significant LD with a single marker on a second group. The class 2 pattern can be considered as a special case of the class 1 pattern. The prevalence of the class 1 pattern among unlinked marker pairs in significant LD may indicate that many of these associations are artifacts. With AFLP markers the single characterizing property of a marker, apart from the primer combination, is the size of the amplified DNA fragment. In a few cases, the fragment scored in the LD population and the identically sized fragment mapped in the mapping population may have been derived from different loci on different linkage groups. If the fragment scored in the LD population was in significant LD with linked loci the wrong mapping position assigned to this fragment due to the mapping of an unrelated fragment in the mapping population would have given rise to a classes 1 or 2 pair of genomic segments with seemingly unlinked marker pairs in significant LD. Taking this into consideration, the true level of linkage disequilibrium between unlinked marker pairs in the rapeseed genome may be even lower than that reported in this study.

Under the assumption of a genetic bottleneck and a strong selection for two different quality traits in the recent breeding history of canola quality winter rapeseed, the near absence of significant LD between unlinked marker pairs was unexpected. There may be two reasons for this result. 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 zero erucic acid phenotype is caused by only two genes (Harvey and Downey 1964; Kondra and Stefansson 1965) that were later mapped by Ecke et al. (1995). Furthermore, 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.

To determine an appropriate threshold for declaring the linkage disequilibrium between two markers significant, a Bonferroni correction was applied. This was necessary because of the large number of marker pairs tested. Without a correction taking into account the multiple testing, a large number of false positives would have occurred. On the other hand, applying a Bonferroni correction leads to a very stringent significance threshold that can considerably increase the number of false negatives, that is cases in which significant LD is not recognized. To alleviate this problem, a global α -level of 0.1 was chosen instead of the more customary 0.05, resulting in the significance threshold of $P = 2.8 \times 10^{-7}$ used in this study.

With respect to an analysis of the prospects for association analysis in rapeseed, there was an additional consideration in choosing a stringent significance threshold. In association analysis, the power to detect a QTL in linkage disequilibrium with a marker is mainly determined by four factors: (1) the population size, (2) the total variance of the trait, (3) the variance caused by the QTL effect, and (4) the fraction of this variance still apparent at the marker locus. The last factor is dependent on the linkage disequilibrium between the marker and the QTL and it can be shown that in a digenic situation, when QTL and marker both have two alleles, the measure r^2 for linkage disequilibrium is also a measure for the fraction of the variance due to a QTL effect still apparent at a marker locus in linkage disequilibrium with the QTL (W. Ecke, data not shown). With the significance threshold applied here, the lowest r^2 values still considered significant were about 0.3 and the mean r^2 of all significant marker pairs was 0.722, levels of linkage disequilibrium that may allow the detection of linked QTL in populations of a size usually used in rapeseed for QTL mapping that have ranged from 105 (Toroser et al. 1995; Thormann et al. 1996) to about 440 (Delourme et al. 2006). In choosing a stringent significance threshold derived from a Bonferroni correction, a subset of marker pairs were selected with a level of linkage disequilibrium that makes it meaningful for a discussion about the prospects of whole-genome association analysis in rapeseed.

The rapid decay of linkage disequilibrium and the near absence of linkage disequilibrium between unlinked markers in canola quality winter rapeseed are, on the one

hand, very favorable for association analysis. It will give association analysis an unprecedented resolution as compared to interval mapping in segregating populations. If a marker is found to be associated with a trait then there will be a high probability that it is closely linked with a QTL for this trait. In addition, with the low level of linkage disequilibrium between unlinked markers the occurrence of false positives, that is markers in association with the trait but not genetically linked to a QTL for that trait, will be rare. This also means that in association analysis in canola quality winter rapeseed, it will not be necessary to take population structure into consideration. Actually, using an evenly distributed subset of 89 of the mapped markers in a principle coordinate analysis, no population structure was detected in the LD population (data not shown). On the other hand, the rapid decay of linkage disequilibrium will necessitate the use of a large number of markers in a whole-genome association analysis. Useful levels of linkage disequilibrium seem to extend over 1–2 cM in the rapeseed genome. Given a size of the genome of approximately 2,400 cM between 1,200 and 2,400 evenly spaced markers would be required to cover the genome at that spacing. Taking into account that even among co-segregating markers not all marker pairs are in significant LD and that markers on a genetic map usually are not evenly spaced, the actual number of markers required for a comprehensive whole-genome association analysis in rapeseed is probably several times larger, bringing it close to the lower bound of the range of 9,600–75,600 SNP markers estimated by Hyten et al. (2007) to be necessary for a whole-genome association analysis in soybean.

Using a large number of markers in an association analysis in canola quality winter rapeseed should allow the localization of most genes and QTL within 1–2 cM. For marker-assisted selection, this will be sufficient but to identify and clone the functional gene a higher resolution would be required. In maize, rice, soybean and barley, there was a correlation between the genetic diversity of the materials analyzed and the level and extent of linkage disequilibrium, with elite breeding materials showing the highest extent and the lowest levels being observed in wild relatives or genetically very broad materials (Remington et al. 2001; Tenaillon et al. 2001; Rafalski 2002; Caldwell et al. 2006; Hyten et al. 2007; Mather et al. 2007). If this pattern should also be true for rapeseed, a two tiered approach where QTL are first mapped in a material with high linkage disequilibrium and then are fine mapped in a material with a rapid decay of linkage disequilibrium as has been proposed for *Arabidopsis* (Nordborg et al. 2002) may be feasible. The QTL could first be mapped in canola quality rapeseed. For fine mapping populations of older, non-canola quality varieties or landraces or even resynthesized rapeseed genotypes could be used. To

determine the resolution that can be achieved by such an approach, linkage disequilibrium would have to be analyzed in the older materials; as yet no such study has been conducted.

Acknowledgments This study was funded by the Deutsche Forschungsgemeinschaft (DFG), Grant nos. BE1854/12-2 and BE1854/12-3. We are grateful to breeding companies Deutsche Saatveredelung AG, KWS SAAT AG, Limagrain-Nickerson GmbH, Norddeutsche Pflanzenzucht Hans-Georg Lembke KG, SW Seed GmbH, and Syngenta Agro GmbH for providing seeds of their varieties and breeding lines for the LD population.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Amar S, Ecke W, Becker HC, Möllers C (2008) QTL for phytosterol and sinapate ester content in *Brassica napus* L. collocate with the two erucic acid gene. *Theor Appl Genet* 116:1051–1061
- Arumuganathan K, Earle ED (1991) Nuclear DNA content of some important plant species. *Plant Mol Biol Rep* 9:208–218
- Becker HC, Löptien H, Röbbelen G (1999) Breeding: an overview. In: Gómez-Campo C (ed) *Biology of Brassica coenospecies*. Elsevier, Amsterdam, pp 413–460
- Caldwell KS, Russell J, Langridge P, Powell W (2006) Extreme population-dependent linkage disequilibrium detected in an inbreeding plant species, *Hordeum vulgare*. *Genetics* 172:557–567
- 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
- Delourme R, Falentin C, Huteau V, Clouet V, Horvais R, Gandon B, Specel S, Hanneon L, Dheu JE, Deschamps M, Margale E, Vincourt P, Renard M (2006) Genetic control of oil content in oilseed rape (*Brassica napus* L.). *Theor Appl Genet* 113:1331–1345
- Ecke W, Uzunova M, Weißleder K (1995) Mapping the genome of rapeseed (*Brassica napus* L.). II. Localization of genes controlling erucic acid synthesis and seed oil content. *Theor Appl Genet* 91:972–977
- Ewens WJ, Spielman RS (2001) Locating genes by linkage and association. *Theor Popul Biol* 60:135–139
- Flint-Garcia SA, Thornsberry JM, Buckler ES IV (2003) Structure of linkage disequilibrium in plants. *Ann Rev Plant Biol* 54:357–374
- Harvey BL, Downey RK (1964) The inheritance of Erucic Acid content in Rapeseed (*Brassica napus*). *Can J Plant Sci* 44:104–111
- Hyten DL, Choi IY, Song Q, Shoemaker RC, Nelson RL, Costa JM, Specht JE, Cregan PB (2007) Highly variable patterns of linkage disequilibrium in multiple soybean populations. *Genetics* 175:1937–1944
- Jannink J-L, Bink MC, Jansen RC (2001) Using complex plant pedigrees to map valuable genes. *Trends Plant Sci* 6:337–342
- Jannoo N, Grivet L, Dookun A, D'Hont A, Glaszmann JC (1999) Linkage disequilibrium among modern sugarcane cultivars. *Theor Appl Genet* 99:1053–1060
- Kim S, Plagnol V, Hu TT, Toomajian C, Clark RM, Ossowski S, Ecker JR, Weigel D, Nordborg M (2007) Recombination and

- linkage disequilibrium in *Arabidopsis thaliana*. Nat Genet 39:1151–1155
- Kondra ZP, Stefansson BR (1965) Inheritance of Erucic and Eicosenoic acid content of Rapeseed oil (*Brassica napus*). Can J Genet Cytol 7:500–510
- Kraakman ATW, Niks RE, Van den Berg PMMM, Stam P, Van Eeuwijk FE (2004) Linkage disequilibrium mapping of yield and yield stability in modern spring barley cultivars. Genetics 168:435–446
- Marwede V, Gül MK, Becker HC, Ecke W (2005) Mapping of QTL controlling tocopherol contents in winter oilseed rape (*Brassica napus* L.). Plant Breed 124:20–26
- Mather KA, Caicedo AL, Polato NR, Olsen KM, McCouch S, Purugganan MD (2007) The extent of linkage disequilibrium in rice (*Oryza sativa* L.). Genetics 177:223–2232
- Nordborg M, Borevitz JO, Bergelson J, Berry CC, Chory J, Hagenblad J, Kreitman M, Maloof JN, Noyes T, Oefner PJ, Stahl EA, Weigel D (2002) The extent of linkage disequilibrium in *Arabidopsis thaliana*. Nat Genet 30:190–193
- 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
- Pilet ML, Delourme R, Foisset N, Renard M (1998) Identification of loci contributing to quantitative field resistance to blackleg disease, causal agent *Leptosphaeria maculans* (Desm.) Ces. et de Not., in winter rapeseed (*Brassica napus* L.). Theor Appl Genet 96:23–30
- Qiu D, Morgan C, Shi J, Long Y, Liu J, Li R, Zhuang X, Wang Y, Tan X, Dietrich E, Weihmann T, Everett C, Vanstraelen S, Beckett P, Fraser F, Trick M, Barnes S, Wilmer J, Schmidt R, Li J, Li D, Meng J, Bancroft I (2006) A comparative linkage map of oilseed rape and its use for QTL analysis of seed oil and erucic acid content. Theor Appl Genet 114:67–80
- Radoev M, Becker HC, Ecke W (2008) Genetic analysis of heterosis for yield and yield components in rapeseed (*Brassica napus* L.) by QTL mapping. Genetics 179:1547–1558
- Rafalski A (2002) Applications of single nucleotide polymorphisms in crop genetics. Curr Opin Plant Biol 5:94–100
- Remington DL, Thornsberry JM, Matsuoka Y, Wilson LM, Whitt SR, Doebley J, Kresovich S, Goodman MM, Buckler ES IV (2001) Structure of linkage disequilibrium and phenotypic association in the maize genome. Proc Natl Acad Sci USA 98:11479–11484
- Stich B, Melchinger AE, Frisch M, Maurer HP, Heckenberger M, Reif JC (2005) Linkage disequilibrium in European elite maize germplasm investigated with SSRs. Theor Appl Genet 111:723–730
- Stich B, Maurer HP, Melchinger AE, Frisch M, Heckenberger M, van der Voort JR, Peleman J, Sorensen AP, Reif JC (2006) Comparison of linkage disequilibrium in elite European maize inbred lines using AFLP and SSR markers. Mol Breed 17:217–226
- Tenaillon MI, Sawkins MC, Long AD, Gaut RL, Doebley JF, Gaut BS (2001) Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). Proc Natl Acad Sci USA 98:9161–9166
- Thormann CE, Romero J, Mantet J, Osborn TC (1996) Mapping loci controlling the concentrations of erucic and linolenic acids in seed oil of *Brassica napus* L. Theor Appl Genet 93:282–286
- Toroser D, Thormann CE, Osborn TC, Mithen R (1995) RFLP mapping of quantitative trait loci controlling seed aliphatic-glucosinolate content in oilseed rape (*Brassica napus* L.). Theor Appl Genet 91:802–808
- 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
- van Ooijen JW (1992) Accuracy of mapping quantitative trait loci in autogamous species. Theor Appl Genet 84:803–811
- Vos P, Hogers R, Sleeker M, Reijmans M, Lee T et al (1995) AFLP: a new concept for DNA fingerprinting. Nucl Acids Res 23:4407–4414
- Zhang Z, Bradbury PJ, Kroon DE, Casstevens TM, Buckler ES (2006) A software package for association and diversity analysis in plants and animals. Poster, plant and animal genome conference, San Diego, January 14–18, 2006. <http://www.maizegenetics.net/index.php?page=bioinformatics/tassel/index.html>
- Zhao J, Becker HC, Zhang D, Zhang X, Ecke W (2005) Oil content in a European × Chinese rapeseed population: QTL with additive and epistatic effects and their genotype–environment interactions. Crop Sci 45:51–59
- Zhao J, Becker HC, Zhang D, Zhang Y, Ecke W (2006) Conditional QTL mapping of oil content in rapeseed with respect to protein content and traits related to plant development and grain yield. Theor Appl Genet 113:33–38
- Zhao J, Dimov Z, Becker HC, Ecke W, Möllers C (2008) Mapping QTL controlling fatty acid composition in a doubled haploid rapeseed population segregating for oil content. Mol Breed 21:115–125