Association mapping of leaf traits, flowering time, and phytate content in *Brassica rapa*

Jianjun Zhao, Maria-João Paulo, Diaan Jamar, Ping Lou, Fred van Eeuwijk, Guusje Bonnema, Dick Vreugdenhil, and Maarten Koornneef

Abstract: Association mapping was used to investigate the genetic basis of variation within *Brassica rapa*, which is an important vegetable and oil crop. We analyzed the variation of phytate and phosphate levels in seeds and leaves and additional developmental and morphological traits in a set of diverse *B. rapa* accessions and tested association of these traits with AFLP markers. The analysis of population structure revealed four subgroups in the population. Trait values differed between these subgroups, thus defining associations between population structure and trait values, even for traits such as phytate and phosphate levels. Marker–trait associations were investigated both with and without taking population structure into account. One hundred and seventy markers were found to be associated with the observed traits without correction for population structure. Association analysis with correction for population structure led to the identification of 27 markers, 6 of which had known map positions; 3 of these were confirmed in additional QTL mapping studies.

Key words: Brassica rapa, genetic variation, phytate, association mapping.

Résumé : La cartographie par association a été employée pour examiner les bases génétiques de la variation chez le *Brassica rapa*, une importante espèce potagère et oléagineuse. Les auteurs ont analysé la variation pour le contenu en acide phytique et en phosphore dans les graines et les feuilles de même que d'autres caractères morphologiques ou liés au développement chez une collection d'accessions variées du *B. rapa*. L'association entre ces caractères et des marqueurs AFLP a ensuite été testée. L'analyse de la structure de la population a révélé l'existence de quatre sous-groupes au sein de celle-ci. Les valeurs pour les caractères à l'étude différaient d'un groupe à l'autre indiquant la présence d'association entre la structure de la population et les caractères, même pour des caractères tels que le contenu en acide phtyique et en phosphore. Les associations entre caractères et marqueurs ont été examinées en faisant abstraction ou non de la structure de la population. Cent soixante-dix marqueurs ont présenté une association avec les caractères observés sans correction pour la structure de la population. Une analyse tenant compte de la structure de la population a permis d'identifier 27 marqueurs, 6 desquels avaient été placés sur une carte génétique et dont 3 ont été confirmés grâce à d'autres études de cartographie QTL.

Mots-clés : Brassica rapa, variation génétique, acide phytique, cartographie par association.

[Traduit par la Rédaction]

Introduction

The traditional method to investigate the genetic basis of variation within the germplasm of a plant species is mapping of quantitative trait loci (QTLs) in specially designed segregating populations. Such populations are derived from crosses of 2 parental lines commonly differing for the trait(s) of interest. Major limitations of this procedure (Flint-Garcia et al. 2003; Gupta et al. 2005) are as follows: (*i*) only those QTLs for which parents differ will be de-

Received 27 February 2007. Accepted 18 July 2007. Published on the NRC Research Press Web site at genome.nrc.ca on 17 October 2007.

Corresponding Editor: H. deJong.

J. Zhao. Laboratory of Plant Breeding, Wageningen University, P.O. Box 386, 6700 AJ Wageningen, the Netherlands; Horticultural College, Hebei Agricultural University, Hebei Province, Baoding 071001, China; Laboratory of Plant Physiology and Genetics, Wageningen University, Arboretumlaan 4, 6703 BD Wageningen, the Netherlands.
M.-J. Paulo. Laboratory of Plant Breeding, Wageningen University, P.O. Box 386, 6700 AJ Wageningen, the Netherlands; Max Planck Institute for Plant Breeding Research, Carl von Linné Weg 10, 50829 Cologne, Germany.
D. Jamar and D. Vreugdenhil. Laboratory of Plant Physiology and Genetics, Wageningen University, Arboretumlaan 4, 6703 BD Wageningen, the Netherlands.
P. Lou and G. Bonnema. Laboratory of Plant Breeding, Wageningen University, P.O. Box 386, 6700 AJ Wageningen, the Netherlands.
F. van Eeuwijk. Laboratory of Plant Breeding, Wageningen University, P.O. Box 386, 6700 AJ Wageningen, the Netherlands; Biometris, Wageningen University, P.O. Box 100, 6700 AC Wageningen, the Netherlands.
M. Koornneef.[1] Laboratory of Genetics, Wageningen University, Arboretumlaan 4, 6703 BD Wageningen, the Netherlands; Max Planck Institute for Plant Breeding Research, Carl von Linné Weg 10, 50829 Cologne, Germany.

[1]Corresponding author (e-mail: Maarten.Koornneef@wur.nl).

 doi:10.1139/G07-078

tected; (*ii*) the precision of QTL location is limited because of the relatively low amount of recombination that can occur in most, often relatively small, biparental offspring populations; and (*iii*) the development of populations takes a considerable amount of time. An alternative strategy is association mapping (Flint-Garcia et al. 2003), which uses the linkage disequilibrium (LD) occurring in natural or breeding populations. Typically more variation can be observed in natural and breeding populations than in segregating populations. Marker–trait associations can be detected in collections of unrelated genotypes when linkage disequilibrium stemming from the linkage between a marker and a causal gene underlying the trait has not yet been completely broken by recombination events. The disadvantage of such collections or populations is that linkage disequilibrium also occurs for reasons other than genetic linkage, such as drift, selection, and admixture of populations.

Association mapping, or LD analysis, has been used widely in humans (Kruglyak 1999) and animals (Farnir et al. 2000; Nsengimana et al. 2004), and many markers have been identified that appear to be closely linked to genes affecting complex diseases (Lander and Schork 1994; Jorde 2000). Association mapping has been applied to plants as well, at the level of both individual genes and the whole genome (Gupta et al. 2005; Mackay and Powell 2007). Understanding the degree of LD across the genome in sampled populations will facilitate the choice of appropriate methods and germplasm collections for genetic association mapping (Varshney et al. 2005). Studies in *Arabidopsis* (Nordborg et al. 2002, 2005), maize (Remington et al. 2001; Tenaillon et al. 2001; Palaisa et al. 2003), rice (Garris et al. 2003), and barley (Kraakman et al. 2004) have shown the impact of biological and historical factors on the extent of LD, explaining the variable degrees of LD. The development of high-throughput genotyping techniques and new statistical methodologies has allowed more efficient use of this genetic approach, resulting in a growing number of publications describing research on marker–trait associations in germplasm or cultivar collections (Aranzana et al. 2005; Breseghello and Sorrells 2006; Kraakman et al. 2006; Malosetti et al. 2007).

The genus *Brassica* has a long history of worldwide cultivation and comprises a large and diverse group of important vegetable, oil, fodder, and condiment crops. The organs consumed as food are different in various *Brassica rapa* morphotypes such as leafy vegetables, turnips, and oil types. Therefore, morphological characteristics such as rosette morphology, leaf shape and structure, enlarged taproot, branching habit, and size of the seedpods differ probably because of directed selection for specific variants. Flowering time and leaf number also vary greatly within *B. rapa*, which is possibly important for the selection of plants to meet growth environments and consumer needs. However, other traits have not or have much less rigorously been subjected to human selection and might show variation independent of the use of the crop. An example of this might be nutrient composition, which is important for future plant breeding programs, provided sufficient variation is present. *Brassica* species and varieties provide a useful source of protein, vitamin C, secondary metabolites such as glucosinolates, and phosphate and other minerals for humans and animals. However, anti-nutritional substances, such as phytic acid in *B. napus* seed meal used as feed for animals (Peng et al. 2001), are an example of nutrient compounds for which reduction is desired. Phytic acid is considered to be an anti-nutritional substance because the highly negatively charged phosphates in this acid form a complex (phytate) with cations (potassium, magnesium, iron, and zinc) that are therefore not bioavailable, resulting in micronutrient (iron and zinc) deficiencies in animals and humans. In fact, there is considerable intraspecific variation in phytate concentration among edible plant parts (White and Broadley 2005). A 3-fold difference in phytate levels between cultivars was observed in *B. napus* (Mollers et al. 1999). The screening of a number of *Arabidopsis thaliana* accessions revealed a wide range of variation in phytate levels, from 7.0 mg to 23.1 mg of phytate per gram of dry seed (Bentsink et al. 2003).

In this study we analysed the variation of phytate and phosphate in a diverse set of *B. rapa* accessions and tested whether association mapping could be used to identify genomic regions controlling these traits. Additionally, we compared the outcome of association mapping for phytate and phosphate content with that for the traits flowering time, leaf edge shape, presence of leaf hairs, and leaf number, for which it is assumed that selection depending on the use of the crop has taken place.

Materials and methods

Plant materials

A collection of 160 *Brassica rapa* accessions encompassing a wide range of morphological types and geographical origins was used in this study. The accessions were obtained from the Dutch Crop Genetic Resources Center (CGN) in Wageningen and the Chinese Academy of Agricultural Sciences (Institute for Vegetable and Flowers (CAAS-IVF) and Oil Crop Research Institute (CAAS-OCRI)) and from T.C. Osborn (University of Wisconsin, Madison, USA), who provided 3 parental lines of mapping populations. The collection includes traditional cultivars, breeding material, and modern cultivars originating from different geographical locations. All of the accessions used in the study and their origins are listed in Table S1[2] and described in Zhao et al. (2005). Most of the accessions are self-incompatible and outcrossing. Therefore, they will be genetically heterogeneous at some loci. However, phenotypically the accessions are very uniform, based on visual observations of 4 replications, which have also been rescored and confirmed in recent experiments (coefficient of variation within accessions: leaf number, 0.07; leaf edge shape, 0.07; flowering time, 0.10; and leaf phytate, 0.19).

Phenotyping

The collection of 160 accessions was grown 3 times.

[2] Supplementary data for this article are available on the journal Web site (http://genome.nrc.ca) or may be purchased from the Depository of Unpublished Data, Document Delivery, CISTI, National Research Council Canada, Building M-55, 1200 Montreal Road, Ottawa, ON K1A 0R6, Canada. DUD 5224. For more information on obtaining material refer to http://cisti-icist.nrc-cnrc.gc.ca/irm/unpub_e.shtml.

Table 1. List of traits analyzed.

Trait type	Trait name	Code	Trait description	Scale
Leafy trait	Leaf edge shape	LES	1, entire; 2, slightly serrated; 3, moderately serrated; 4, fully serrated	1–4
	Leaf trichomes	LT	Density of leaf trichomes: 0, none; 1, few; 2, many	0–2
	Leaf number	LN	Number of leaves at 4-week stage	Number
Flowering time	Flowering time without vernalization	NDF	Days from sowing to appearance of first open flower	Days
	Flowering time after vernalization	VDF	Days from sowing to appearance of first open flower	Days
Phytate	Phytate level in leaves	LPHY	Phytate content in lyophilized leaves	mg/g
	Phosphate level in leaves	LPHO	Phosphate content in lyophilized leaves	mg/g
	Phytate level in seeds	SPHO	Phytate content in mature dry seeds	mg/g
	Phosphate level in seeds	SPHY	Phosphate content in mature dry seeds	mg/g

First, from December 2002 till March 2003, accessions after seedling vernalization were grown to produce selfed seeds for phytate and phosphate analysis. Second, from November 2002 to February 2003, accessions were grown and leaves were harvested for phytate and phosphate analysis. The third set of plants was grown from March to May 2003 to measure a number of morphological traits. In total, 9 traits were measured in this study (Table 1).

For the first experiment, 3 plants per accession were vernalized after germination for 2 weeks in a dark cold room (4–6 °C) and thereafter seedlings were grown in a greenhouse with supplementary light (16 h day length) from December 2002 to March 2003 in Wageningen. The number of days to flowering of vernalized plants (VDF) was recorded, from sowing to the appearance of the first open flower. One plant from each accession was selected and its young leaves or flower buds were collected for DNA isolation and AFLP genotyping as described in a previous study (Zhao et al. 2005). Since many accessions are not homozygous for all markers and we analysed DNA from only 1 plant per accession, we could not assay all allelic information. Three batches of mature seeds from 1 plant of each accession were used for phytate and phosphate analysis (SPHY and SPHO, respectively).

A second set of 4 non-vernalized plants per accession was grown under similar soil and light conditions in the greenhouse from November 2002 to February 2003 in Wageningen and was used to score the number of days to flowering (NDF). For some very late turnip and Chinese cabbage types, which did not flower within the experimental period, NDF was set to 120 days. One whole leaf was collected and lyophilized from each of four 5-week-old plants of each accession to measure phytate and phosphate contents (LPHY and LPHO, respectively) (mg/g dry mass). Leaves of 2 plants were ground together to represent 1 biological replication. The phytate and phosphate levels were determined using HPLC as described by Bentsink et al. (2003) with minor modifications.

A third set of 3 non-vernalized plants per accession was grown under similar soil and light conditions in the greenhouse from March to May 2003 in Wageningen and was used to measure leaf number (LN) at 4 weeks after sowing. Additionally, the leaf edge shape (LES: 1, entire; 2, slightly serrated; 3, moderately serrated; 4, fully serrated) and density of leaf trichomes (LT: 0, no trichomes; 1, few trichomes; 2, many trichomes) were scored.

Data analysis for summary statistics, one-way analysis of variance (ANOVA), and correlation analysis were performed in GenStat release 8.1.

Genotyping

The *B. rapa* accessions in this study have been genotyped using AFLP fingerprinting (Zhao et al. 2005). In total, 437 scorable amplification products ranging from 50 bp to 500 bp were generated with 4 primer combinations (pAT/mCCA, pGG/mCAA, pAG/mCAC, and pTA/mCAT). Of the 437 AFLP bands, 389 were polymorphic.

Map positions of markers were derived from an integrated map with AFLP and SSR markers that was based on 2 double haploid (DH) populations, DH-38 (PC-175 × YS-143) and DH-30 (VT-115 × YS-143), sharing the common parent YS-143 (Zhao 2007 and unpublished data). The three parental lines (Yellow sarson, YS-143; Pak choi, PC-175; and vegetable turnip, VT-115) were included in the AFLP fingerprinting study, allowing comparison of the AFLP markers based on band size. Of the AFLP markers used to detect association among the 160 accessions, 76 were mapped on the integrated map, with 3 to 11 markers per chromosome (= R-linkage group). The nearly 300 remaining AFLPs could not be mapped because most of them were either not polymorphic or were absent in parental lines of the DH mapping populations.

Population structure

The program *structure* version 2.1 was used to identify groups in the population using a Bayesian approach (Pritchard et al. 2000; Falush et al. 2003). The accessions are classified into a pre-set number of clusters such that LD does not occur within the groups, whereas LD is present between groups.

Within *structure*, we tested with a model allowing population admixture, implying that genotypes can have a mixed ancestry, and assumed independent allele frequencies between subpopulations. The number of subpopulations (K) was set to vary between 1 and 10, and for each fixed number of subpopulations, 2 independent Markov Chain Monte Carlo analyses were run using 600 000 iterations for each; the first 100 000 iterations were discarded as burn-in. The likelihood of the different K values is calculated and the value of K with the highest likelihood can be interpreted to correspond to the number of clusters in the sample. Moreover, under the model with admixture, *structure* estimates the membership probabilities, i.e., it assigns every genome proportionally to each cluster and thus allows identification

of groups of accessions that form a population. The average likelihood values of 2 runs for a given K value increased gradually until $K = 4$. Additional clusters for $K > 4$ did not result in groups containing the majority of any single genome.

To test the homogeneity of AFLP band frequencies across the 4 subpopulations, Fisher's exact test was applied to contingency tables of marker presence/absence versus subpopulation using SAS® software (SAS Institute Inc. 1999).

Analysis of LD between markers

To investigate LD between pairs of markers, LD was calculated within subpopulations, restricting the analysis to mapped markers whose absolute band frequency was between 5 and (N_i – 5), where N_i is the size of the subpopulation i (i = 1, …, 4). Chi-square tests of association were calculated for the 2 × 2 contingency tables formed from the band presence/absence scores for the pairs of loci that fulfilled the frequency condition in a particular subpopulation. The p values of these tests were combined across subpopulations for pairs of loci to form the statistic $-2\sum_{i}^{k}\ln(p_i)$, which should be distributed as $\chi^2_{(2k)}$ (Fisher 1954), with k being the number of subpopulations for which the association could be calculated ($k \leq 4$). The overall amount of LD between pairs of marker loci used the p value of the combination statistic expressed as $-\log_{10}(p\text{ value})$. This latter quantity was plotted against genetic distance (in cM) to investigate LD decay.

Association analysis of quantitative traits

Association analysis of quantitative traits (LN, NDF, VDF, LPHO, LPHY, SPHO, and SPHY) was performed in 3 steps with a series of increasingly complex mixed models and was carried out in Genstat (Payne and Arnold 2002) using restricted maximum likelihood.

Model 1: phenotypic response = marker + error

This model corresponds to a series of simple t tests, without correction for population substructure and additional QTLs present elsewhere in the genome. The design matrix corresponding to the fixed effects (marker) is a vector corresponding to the marker scores, i.e., a vector having the value 1 if a band is present and 0 otherwise.

Model 2: phenotypic response = structure + marker + error

This model corrects for population substructure by adding a random term to model 1, labeled "structure", containing the subgroup membership probabilities (Q matrix) obtained from *structure* (Pritchard et al. 2000; Falush et al. 2003). The design matrix for the random term (structure) contains the membership probabilities for each subgroup. This model is very similar to that described in Yu et al. (2006), with the difference that here we use the membership probabilities matrix instead of 0/1 scores as described in Yu et al. (2006). Therefore, our design matrix measures not only the differences among subgroups but also differences among accessions.

Model 3: phenotypic response = structure + selected marker subset + error

This model is a multi-QTL model with a correction for population substructure, like model 2, obtained from the set of putative QTLs (markers) identified by model 2 after a cleaning-up by backward selection.

Another set of traits, such as LES and LT, were measured as ordered categories, and for those traits association analysis was based on ordinal regression (Dobson 2002), using the analogue of the 3 models above.

Correction for multiple testing

The previous association tests (marker–trait and marker–marker) were subjected to correction aimed at controlling the false discovery rate (FDR). The FDR is the expected proportion of false positives in the set of all rejected tests, i.e., the proportion of false associations in the total set of significant associations. The p values obtained in the different analysis were adjusted according to the two-stage linear step-up procedure (Benjamini and Yekutieli 2005) from which q values were obtained. These q values attribute to each test a measure of significance in terms of the FDR. Here we set the significance level to 0.05.

Results

Population structure

In a previous study (Zhao et al. 2005), cluster analysis using the unweighted pair group method with arithmetic averages (UPGMA) produced a phenetic tree that suggested 2 main groups based on the origin of the material (Asia versus Europe), plus a small group of spring oil types from India and Bangladesh. The model-based approach of *structure* suggests the presence of 4 subpopulations consisting of 3 large groups — S1, consisting of 60 accessions; S2, consisting of 40 accessions; and S4, consisting of 51 accessions — and 1 small group, S3, with 9 accessions (Fig. 1; Table 2; Table S1[2]). Most Oriental accessions are grouped into S1 and S4, and most Western accessions are grouped into S2, while the spring oil types from India and Bangladesh group into S3.

The S1 subpopulation of accessions encompasses all Pak choi types (including 2 Caixin and 1 Wutacai accession) and most of the turnip rape accessions from China, 4 winter turnip rape accessions from Pakistan, 8 Japanese turnips, 1 turnip from Pakistan, and 1 turnip from Russia, and 3 Japanese Neep greens (2 Komatsuna and 1 turnip green accession). Two Mizuna accessions from Japan and the Netherlands are also grouped into S1 with an admixture of S2, S3, and S4. S4 is another subpopulation of Oriental origin and encompasses 41 accessions of Chinese cabbage cultivars from Asia (China, Korea, and Japan) and 7 from Western countries. The S4 group also includes 1 French turnip accession with S2 admixture plus 1 Chinese turnip rape and 1 Japanese turnip green admixed with S1 and S2. The accessions of subpopulation S2 originate mainly from Western countries, comprising 25 European turnips and 3 turnips from Uzbekistan and India, 6 Italian Broccolettos, and 6 oil types from the USA, Canada, Sweden, and Pakistan. The small but distinct S3 subpopulation is formed by spring oil types comprising 2 Yellow sarsons originating

Fig. 1. Results from the program *structure* under the assumption of cluster number $K = 4$. *Brassica rapa* accessions are represented by a bar, which is partitioned into several segments with different shades of gray according to the accession's estimated membership fractions of the 4 clusters (S1, S2, S3, S4).

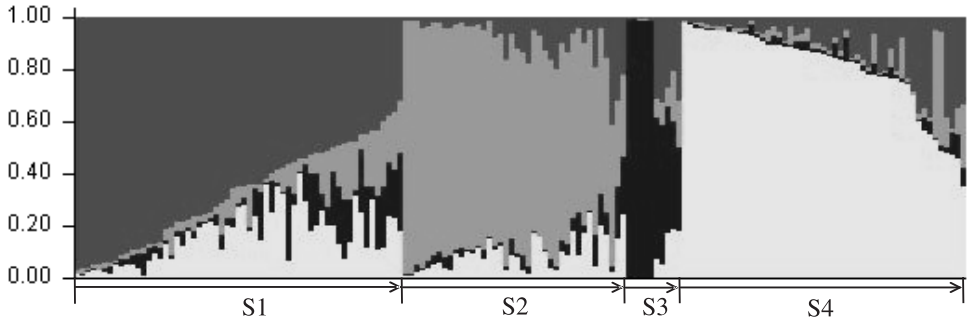


Table 2. Composition of subpopulations S1–S4, with number of accessions per cultivar group and their origin.

		Cultivar group											Origin	
Subpopulation	Total no. of accessions	T	BRO	CC	PC	NG	MIZ	YS	RC	WO	SO	OR	East	West
S1	60	10	0	0	32	3	2	0	0	4	0	9	56	4
S2	40	28	6	0	0	0	0	0	0	4	2	0	5	35
S3	9	0	0	0	0	0	0	2	1	0	6	0	5	4
S4	51	1	0	48	0	1	0	0	0	0	0	1	43	8

Note: T, turnip; BRO, Broccoletto; CC, Chinese cabbage; PC, Pak choi, Caixin, and Wutacai; NG, Neep green; MIZ, Mizuna; YS, Yellow sarson; RC, rapid cycling; WO, winter turnip rape; SO, spring turnip rape; OR, Chinese turnip rape; East, Oriental country; West, Western country.

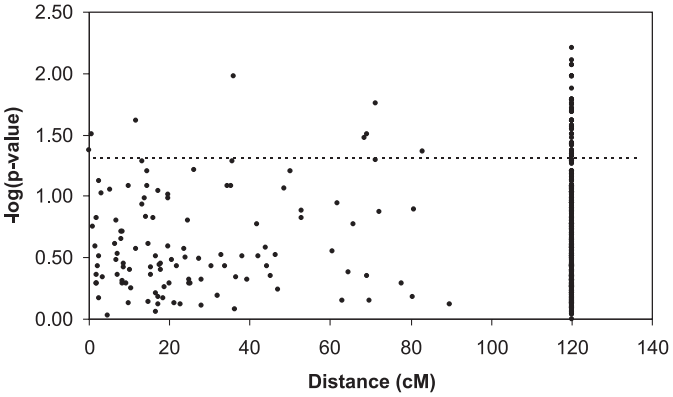
from India, 6 spring turnip rape cultivars from Bangladesh, and 1 rapid cycling line from the USA, which probably is derived from the Yellow sarson types from the Indian subcontinent.

Of the 233 polymorphic markers, 119 were associated with the 4 structured subgroups ($p < 0.0001$). Marker band frequency differences between the groups may be responsible for phenotypic differences, but not necessarily so.

Analysis of marker–marker associations

The analysis of marker–marker associations within subpopulations showed a few cases of significant LD up to around 80 cM (Fig. 2), but there was no clear LD decay with genetic distance. Furthermore, LD between markers on different chromosomes, concentrated at 120 cM in Fig. 2, was sometimes higher than the LD between markers on the same chromosome. A reason for the absence of LD decay may be the relatively small number of marker pairs located on the same linkage group (115).

Fig. 2. Linkage disequilibrium (LD) decay for a set of 76 mapped AFLP markers. Here LD = $-\log_{10}(p$ value) and the genetic distance is expressed in centimorgans. Pairs of markers located on different linkage groups are shown at a distance of 120 cM. The remaining data points represent LD values between pairs of markers on the same chromosome. The points above the horizontal line indicate significant LD ($p < 0.05$).

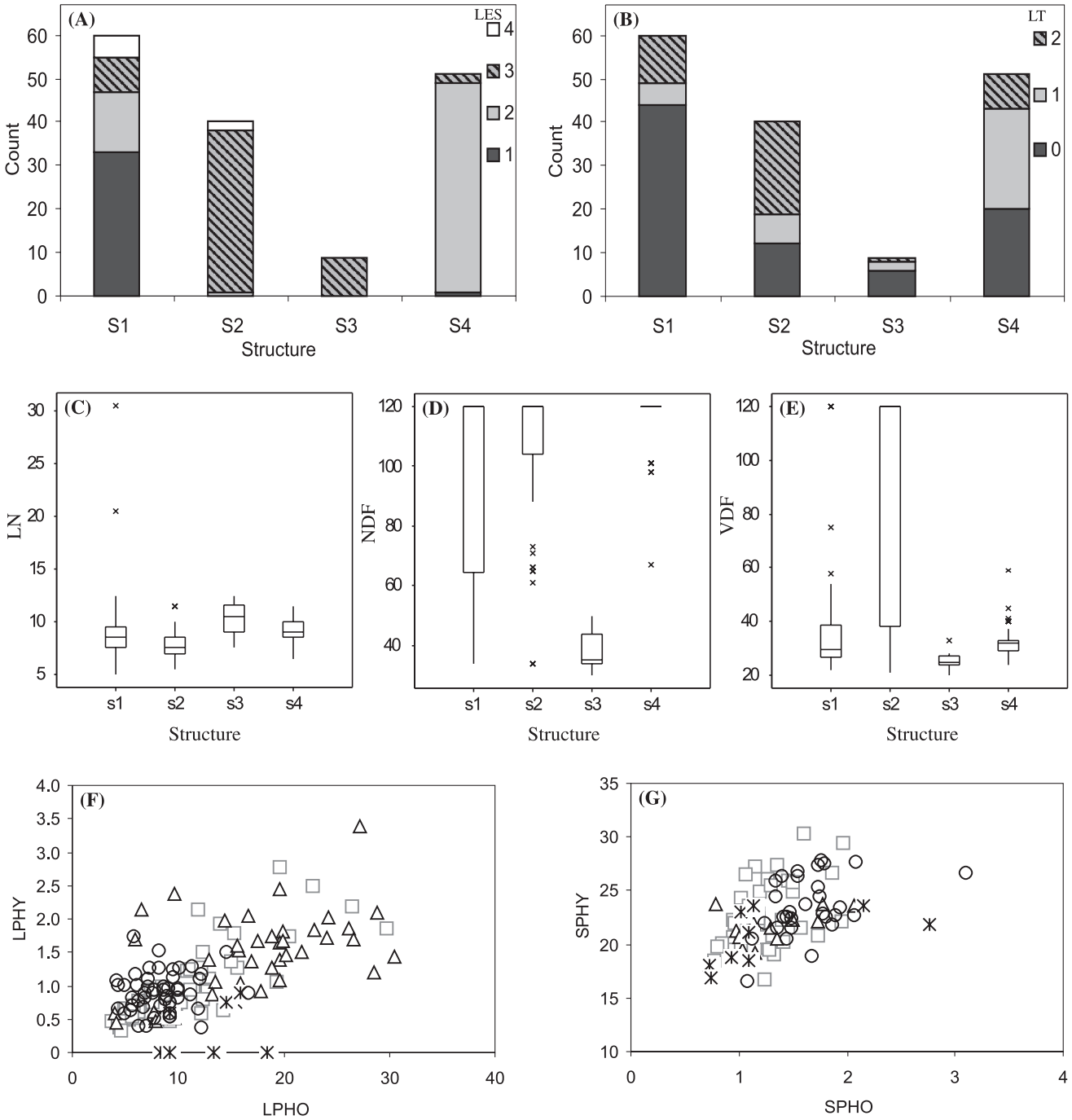


Variation in observed traits

The distributions of traits and the correlations between traits are shown in Fig. 3, separately for the subgroups S1 to S4. Statistics for all observed traits, organized in different subpopulations, are summarized in Table 3.

Leaf characteristics including leaf edge shape (LES), leaf trichomes (LT), and leaf number (LN) are important morphological traits distinguishing vegetable *B. rapa* types. The distribution of LES and LT is related to the structured subpopulations (Figs. 3A, 3B; Table 3). More than 50% of the accessions in S1 have entire leaves (LES = 1), most of the accessions in S2 and S3 have moderately serrated leaves (LES = 3), most of the accessions in S4 have very lightly serrated leaves (LES = 2), and 1 Mizuna and some winter oil accessions in S1 and S2 have severely serrated leaves (LES = 4). The different classes of leaf trichome frequency (0, 1, and 2) are distributed within each subpopulation, but most of the accessions in S1 (mainly Pak choi) have a hairless leaf surface and most of the accessions in S4 (Chinese cabbage) have few or no trichomes. The variation of LES and LT was significantly different ($p < 0.01$) between different subpopulations. For LN, the range was similar in S2, S3, and S4, from 6 to 13 leaves (Fig. 3C; Table 3). Within S1, the variation of LN is higher (5–31) because of 2 Mizuna and 1 Wutacai accession with many leaves. The mean value

Fig. 3. Natural variation of (A) leaf edge shape (LES: 1, entire; 2, slightly serrated; 3, moderately serrated; 4, fully serrated), (B) leaf trichomes (LT: 0, no trichomes; 1, few trichomes; 2, many trichomes), and (C) leaf number (LN); (D) flowering time without vernalization (NDF, days) and flowering time after vernalization (VDF, days); and phytate and phosphate concentrations in (E) leaves (LPHY and LPHO, mg/g dry mass) and (F) seeds (SPHY and SPHO, mg/g dry mass) of 160 *B. rapa* accessions. The different symbols refer to the different subgroups as illustrated in Fig. 1. □, S1; △, S2; ×, S3; ○, S4.



of LN in S2 is lower although 2 accessions (1 spring and 1 winter oil type) had a high value of 11.5. The variation of LN differed ($p = 0.01$) between subpopulations.

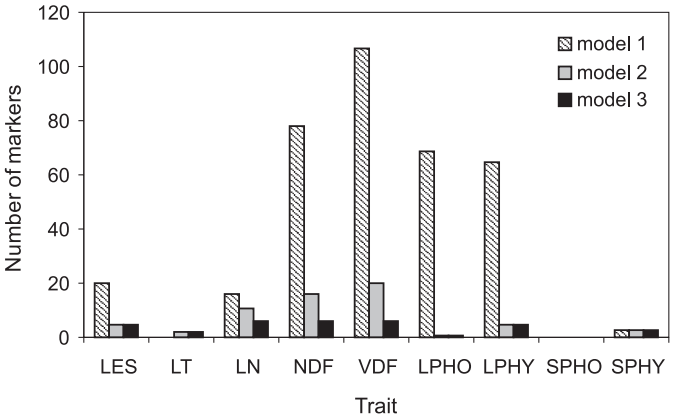
Flowering time is a very important developmental trait in *B. rapa*, and wide variation in days to flowering was observed within the collection; more than a 3-fold difference was found between accessions without vernalization (NDF, 30–103 days) and those with vernalization (VDF, 20–75 days) when non-flowering plants were excluded (Figs. 3D, 3E; Table 3). Under non-vernalization conditions, only 52 accessions flowered, including 26 accessions in S1 (15 Pak choi, 2 Caixin, 2 Komatsuna, 1 vegetable turnip, 4 winter turnip rape, and 2 Chinese oil type cultivars), 6 Broccoletto and 4 oil type accessions in S2, all spring oil accessions in S3, and 6 Chinese cabbage cultivars and 1 Chinese oil type in S4. Late forms of Chinese cabbage in S4 re-

Table 3. Statistical description of observed traits in the subpopulations S1–S4.

		S1			S2			S3			S4		
Trait	Scale	Mean	SD	Range	Mean	SD	Range	Mean	SD	Range	Mean	SD	Range
LES**	1–4	1.8	1.0	1–4	3.0	0.3	2–4	3.0	0.0	3–3	2.0	0.2	1–3
LT**	0–2	0.5	0.8	0–2	0.2	0.9	0–2	0.4	0.7	0–2	0.8	0.7	0–2
LN*	Number	9.1	3.6	5–31	7.9	1.3	6–12	10.2	1.6	8–13	9.2	1.1	7–12
NDF**	Days	96.3	30.7	34–120	105.6	26.5	34–120	38.2	7.4	30–50	116.6	9.6	67–120
VDF**	Days	36.2	19.1	22–120	88.0	42.2	21–120	25.7	1.2	20–33	32.6	5.7	24–59
LPHY**	mg/g	1.0	0.6	0.3–2.8	1.5	0.6	0.4–3.4	0.3	0.4	0.0–0.9	0.9	0.3	0.4–1.74
LPHO**	mg/g	10.7	5.9	3.7–29.8	17.0	7.2	4.0–30.5	13.1	3.5	8.3–18.5	11.2	6.1	4.2–16.7
SPHO**	mg/g	1.3	0.3	0.8–1.9	1.3	0.4	0.8–2.1	1.3	0.7	0.7–2.8	1.6	0.4	1.1–3.1
SPHY*	mg/g	22.9	3.3	16.7–30.4	21.7	1.6	19.2–23.8	20.6	2.6	16.9–23.6	23.7	2.8	16.6–27.9

Note: One-way analysis of variance (ANOVA) between subpopulations: *, significant at $p < 0.05$; **, significant at $p < 0.01$. SD, standard deviation. See Table 1 for definitions of trait abbreviations.

Fig. 4. Number of markers associated with traits, as resulting from three different models.



sponded strongly to vernalization, whereas only 8 Japanese turnips in S1 and 2 other turnips in S2 flowered upon vernalization. Most turnip accessions in S2 and 2 turnips in S1 did not flower after vernalization, which may indicate that these accessions require a longer period of cold or vernalization at a later stage of development to induce flowering. The differences in flowering time are also associated with population structure, as illustrated in Table 3. The variation of NDF and VDF was significantly different ($p < 0.01$) between different subpopulations.

The levels of phytate in seeds were 10 times higher than those in leaves. However, phosphate levels in leaves were 10 times higher than those in seeds. Across the accessions, a positive correlation between the two compounds was detected at $p < 0.01$, with a correlation coefficient of $r = 0.52$ in leaves and of $r = 0.44$ in seeds (Figs. 3F, 3G). The correlation between phytate in seeds and phosphate in leaves was low ($r = -0.21$) and not significant. In leaves, the variation of phytate (LPHY, 0–3.4 mg/g dry mass) and phosphate (LPHO, 3.7–30.5 mg/g dry mass) was large. In some oil type accessions, phytate levels were below detection level in leaves. Variation in seeds was less, being 16.7–30.4 mg/g (dry mass) for phytate and 0.7–3.1 mg/g (dry mass) for phosphate.

Variation of phytate and phosphate levels was observed within each subpopulation (Figs. 3E, 3F; Table 3). The variation was significantly different between subpopulations at $p < 0.05$ for SPHY and $p < 0.01$ for SPHO, LPHY, and LPHO. However, for SPHO and SPHY, only 87 accessions were evaluated because many accessions (mainly turnips in S2) did not produce enough seeds. Although the phytate and phosphate concentrations in Chinese cabbages in S4 and spring oil accessions in S3 were lower than those in Pak choi accessions in S1 and turnip accessions in S2, the range of variation, for both seeds and leaves in each subgroup, is overlapping within this *B. rapa* collection.

Association mapping

Association between markers and quantitative traits was examined using 3 models: a simple *t* test, a model correcting for population structure, and a multi-QTL procedure (Table S2[2]).

Fig. 4 illustrates that the 3 different models detect quite different numbers of markers that are strongly associated with the traits analyzed. Using the *t* test (model 1), 170 markers (from 0 to 107 markers/trait) were found to be associated with observed traits, including many markers for NDF, VDF, LPHO, and LPHY. When population structure was used in the mixed models (models 2 and 3) to correct for spurious associations, the number of markers associated with the different traits was much lower. However, the associations identified without taking population structure into account in model 1 can be either associated with the trait or spurious and related to the population structure.

The statistically preferable multi-QTL procedure (model 3), with correction for population structure and genetic background noise, detected only a few significant markers (0–6) per trait. Table 4 shows an overview of all significant associations between measured traits and AFLP markers using the multi-QTL procedure. The linkage group of the associated marker is listed together with previously identified QTLs in the same group. In total, 27 markers, of which 6 had known map positions, were associated with the 8 traits at $p < 0.05$.

For leaf traits (LES, LT, and LN), 13 associated AFLP markers were detected with $p < 0.05$ using the multi-QTL model. Of the 5 markers associated with LES, only 1 (pAG/mCAC0090.5) was mapped, namely at the bottom of linkage group R08. One marker associated with LT was mapped on R05. Leaf number was associated with 6 markers with unknown map position (Table 4). One unmapped AFLP marker (pAG/mCAC073.0) was associated with both LN

Table 4. An overview of all significant associations ($p < 0.05$) between measured traits and AFLP markers using the multi-QTL model (model 3).

			Significant p value in association with various traits								
Marker	Linkage group (cM)	Previous QTL*	LES	LT	LN	NDF	VDF	LPHO	LPHY	SPHO	SPHY
pAG/mCAC0154.7	R02 (63.4)						0.035				
pTA/mCAT0230.2	R02 (84.5)	SPHY, SPHO									0.017
pAT/mCCA0135.0	R03 (15.8)	FLQTL-4					0.050				
pGG/mCAA0339.8	R05 (50.1)			0.000			0.000				
pAG/mCAC0316.7	R07 (34.0)	LPHY							0.001		
pAG/mCAC090.5	R08 (96.5)		0.014								
pAG/mCAC0465.9	Unmapped			0.018							
pAG/mCAC0317.4	Unmapped		0.003						0.014		
pAG/mCAC0255.5	Unmapped								0.012		
pAG/mCAC0253.3	Unmapped					0.002					
pAG/mCAC0247.5	Unmapped					0.009					
pAG/mCAC0245.8	Unmapped						0.000				
pAG/mCAC0205.1	Unmapped		0.005								
pAG/mCAC073.0	Unmapped				0.002	0.001	0.018				
pGG/mCAA0329.6	Unmapped					0.001					
pGG/mCAA0181.3	Unmapped		0.000								
pTA/mCAT0419.3	Unmapped				0.001						
pTA/mCAT0362.5	Unmapped								0.001		
pTA/mCAT0241.3	Unmapped										0.001
pTA/mCAT0152.1	Unmapped				0.001						
pTA/mCAT077.7	Unmapped						0.017				
pAT/mCCA0408.7	Unmapped					0.006					
pAT/mCCA0267.7	Unmapped				0.013						
pAT/mCCA0264.7	Unmapped		0.019								
pAT/mCCA0184.2	Unmapped				0.001						
pAT/mCCA089.6	Unmapped				0.001						
pAT/mCCA076.9	Unmapped					0.001		0.001	0.020		

Note: QTLs were identified in QTL analysis in 4 double haploid populations and one F_2 population as described by Zhao (2007) and Lou et al. (2007). See Table 1 for definitions of trait abbreviations.

*QTLs identified in previous studies in a similar genomic region (FLQTL-4, flowering time QTL).

and flowering time (NDF and VDF), which will contribute to a correlation between the two traits. Eleven markers were associated with days to flowering. However, only 1 AFLP marker (pAG/mCAC073.0) was associated with both NDF and VDF, which explains why the two traits were not that strongly correlated. VDF was associated with 6 markers, 3 of which had a known map position and were distributed over R02, R03, and R05. One associated marker (pAT/mCCA0135.0) was located on R03, close to the position of a flowering time QTL (FLQTL-4) that had been identified in our previous study (Zhao 2007). The 6 markers that were correlated with NDF had no known map position.

For phytate and phosphate levels in seeds and leaves, 7 associated AFLP markers were detected with $p < 0.05$. One unmapped marker (pTA/mCCA076.9) was associated with both LPHY and LPHO, which illustrates close linkage or pleiotropy of the two traits. However, association with the same markers was not detected for SPHY and SPHO. For 2 markers associated with these 4 traits, the map positions were known and QTLs were identified at those positions in mapping studies (Zhao 2007). The marker pAG/mCAC0316.7, related to LPHY, co-localized with a strong QTL region related to LPHY on R07 based on 3 DH populations analyzed in a previous study (Zhao 2007). Furthermore, for a marker (pTA/CAT0230.2) mapped on R02, an association to SPHY was detected with $p = 0.017$.

The smallest detected differences in trait means between marker genotypes for the models used give some indication of the power of the present study. The present association analysis (model 1) was able to detect differences between marker genotypes of 1.1 for the mean number of leaves, 10.6 days for flowering time (with and without vernalization), 2.4 mg/g for leaf phosphate, 0.2 mg/g for leaf phytate, and 2.3 mg/g for seed phytate. For seed phosphate no marker–trait associations were found, whereas a maximum difference of 0.6 mg/g in seed phosphate means was observed.

Discussion

In the present study we analysed a number of traits in a set of *B. rapa* genotypes representing the various cultivar types in the germplasm of this species. For these genotypes AFLP analyses had been performed, which indicated a loose population structure based on UPGMA (Zhao et al. 2005). Although for each accession the phenotypes were assessed on plants that were different from the plants that were genotyped, we assume that for most relevant loci the genotypes

were identical, since we noticed hardly any phenotypic variation within accessions. Based on allele frequencies, the analysis of the same AFLP data set with the program *structure* confirmed the above result and suggested 4 subgroups (Pak choi, group S1; turnip, group S2; spring oil, group S3; and Chinese cabbage, group S4). The structured subgroups S1 and S4 belonged mainly to the UPGMA group 1, and the structured subgroup S2 belonged mainly to the UPGMA group 2. From the *structure* results, the admixture between accessions could also be detected (Table S1[2]). For example, several Pak choi accessions and Chinese oil types in S1 share a genetic background with S4, which is possibly related to their breeding history.

This population structure was taken into account in the analysis of variation of a number of traits determined in these materials. The number of leaves was evaluated at the 4-week-old plant stage, which does not reflect the whole vegetable development process because some vegetable types form more leaves during later development. Flowering time of *B. rapa* species used as vegetables or turnips is agronomically important because it relates to yield and quality. Flowering time was assessed with and without vernalization in the present study. Chinese cabbage and turnip types displayed a different vernalization requirement compared with other cultivar groups, which suggests that different genes affecting flowering time are present in these groups. Vernalization greatly reduced the range of variation in flowering time when non-flowering plants were excluded.

The extensive variation of leaf phosphate might be used to breed for better phosphate use efficiency. Phytate content is relevant for oilseed types, because the meal is used as feed, and a 2-fold range of variation in seed phytate levels exists. We also observed a positive correlation of phytate and phosphate, as has been reported in *Arabidopsis* (Bentsink et al. 2003) and corn (Raboy et al. 2001). Despite this general correlation, a few accessions were identified with relatively high phosphate and low phytate levels in seeds, compared with the mean values of 22.8 mg/g for phytate and 1.4 mg/g for phosphate. Examples are the spring oil accession SO-032 (phytate, 21.8 mg/g; phosphate, 2.8 mg/g) and the winter oil accession WO-082 (phytate, 18.3 mg/g; phosphate, 3.6 mg/g). For leaf content also some genotypes with a strongly altered relationship between phosphate and phytate levels were found. The vegetable turnip accessions VT-015 and WO-024 have a higher phosphate level (30.5 mg/g and 28.6 mg/g) but a low phytate level (1.43 mg/g and 1.19 mg/g), and all spring oil accessions in S3 have lower or non-detectable phytate levels (from 0 to 0.9 mg/g). In future *Brassica* breeding programs, it will be possible to combine high phosphate and low phytate levels and to select ideal genotypes as parents of mapping populations for QTL identification.

Naturally occurring genetic variation is a useful resource for the genetic mapping of complex phenotypic traits (Koornneef et al. 2004). We applied association mapping in *B. rapa* for identification of genetic markers associated with leaf traits, flowering time, and phosphate levels and to compare the outcome of association mapping with QTLs detected in DH populations that we developed for this purpose. The presence of population structure may affect LD and produce false positives. The associations among markers themselves were also examined; markers that differ in allele frequency between subpopulations provide an example of LD due to population structure. Some of these markers may be responsible for observed phenotypic differences between the groups. However, marker frequencies between groups can also differ because of chance processes. For example, the frequency of one band, pAG/mCAC085.6, clearly differed between subgroups (71.7% for S1, 15.0% for S2, 90.0% for S3, and 72% for S4). When this band was present, the mean leaf phosphate level (LPHO; 14.1 mg/g for absence and 10.0 mg/g for presence) decreased 29.1%. We cannot discriminate between chance and non-chance associations at the level of the phenotypic differences between groups.

It is difficult to estimate the marker density needed to maintain power in association analysis because for most markers used in this study the map position is not known. Although LD among a subset of mapped markers could be observed up to a distance of 80 cM, too few marker pairs were used to assess LD decay with sufficient accuracy. Based on the limited amount of mapping data, we could not obtain sufficient information on LD decay. Therefore, it is difficult to compare our data on the extent of LD with that in *Arabidopsis* or other plant species.

Since trait values differed significantly between subgroups, an association between population structure and these trait values is suggested even for traits such as phytate and phosphate levels, for which we assumed no selection had occurred. Markers associated with the traits analyzed are shown for models 1 to 3 in Table S2[2]. Therefore, a number of associations were identified without correction for population structure, of which some can still point to "causal" QTLs. Those markers are responsible for the phenotypic differences but are also related to structure, as also suggested for *Arabidopsis* (Aranzana et al. 2005). For example, one marker, pAT/mCCA430.9 (mapped on R07), was significantly associated with leaf number (LN) before correction for population structure (Table S2), and a QTL in this region was also detected for LN in our previous study (Lou et al. 2007). This association might indicate true linkage, although it disappeared after correction for population structure. Markers that show association after correction for substructure can more reliably be interpreted as being linked to QTLs. In the present study, the outcomes of association analysis were partly similar for models 2 and 3. The associations for LES, LT, LPHO, LPHY, SPHO, and SPHY identified by model 2 are also apparent in model 3. However, the number of markers associated with LN, NDF, and VDF in model 3 is reduced compared with that in model 2 (Fig. 4) because correlations between markers are taken into account in model 3 and redundant markers are removed.

The traits studied in this paper have also been analysed in a set of mapping populations including 8 parents (Lou et al. 2007). These analyses identified a number of QTLs related to the different traits. Multiple QTLs can underlie trait variation. Since most of the AFLPs in this study were not mapped, we could not conclude that multiple associated markers referred to an equal number of genome positions. One objective of future studies is to use mapped markers for association mapping. Many SSR markers and increasing sequencing information for *B. rapa* are already available

(http://www.brassica.info), which makes it possible to profile SSR and gene-target markers across all accessions, allowing determination of the LD level across the genome and facilitating the identification of QTLs in *B. rapa*.

Confirmation of some of the marker–trait associations by QTL analysis indicated that association mapping allows the detection of linkage with moderately frequent alleles, which thereafter can be confirmed by linkage analysis in mapping populations designed on the basis of phenotype and marker contrasts identified in association mapping studies.

Acknowledgements

We thank the Wageningen Plant Sciences Experimental Centre of Wageningen University for taking care of the plants. We thank Noortje Bas from the Centre for Genetic Resources, the Netherlands (CGN); Li Xixiang from the Institute of Vegetables and Flowers, Chinese Academy of Agricultural Sciences; Liu Shengyi, Oil Crop Research Institute, Chinese Academy of Agricultural Sciences; and T.C. Osborn, University of Wisconsin, Madison, USA, for kindly supplying the accessions used in this study. We thank Dunia Pino Del Carpio and Christina Requena Robles for additional phenotypic scoring. The project is sponsored by the Asian Facility (project AF01/CH/8 "Sino–Dutch Genomic Lab and Vegetable Research Center") and a fellowship from a student exchange programme (Huygens) between the Chinese Scholarship Council and Nuffic (Netherlands Organisation for International Cooperation in Higher Education).

References

Aranzana, M.J., Kim, S., Zhao, K., Bakker, E., Horton, M., Jakob, K., et al. 2005. Genome-wide association mapping in *Arabidopsis* identifies previously known flowering time and pathogen resistance genes. PLoS Genet. **1**(5): e60. doi:10.1371/journal.pgen.0010060.

Benjamini, Y., and Yekutieli, D. 2005. Quantitative trait loci analysis using the false discovery rate. Genetics, **171**: 783–790. doi:10.1534/genetics.104.036699. PMID:15956674.

Bentsink, L., Yuan, K., Koornneef, M., and Vreugdenhil, D. 2003. The genetics of phytate and phosphate accumulation in seeds and leaves of *Arabidopsis thaliana* using natural variation. Theor. Appl. Genet. **106**: 1234–1243. PMID:12748774.

Breseghello, F., and Sorrells, M.E. 2006. Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. Genetics, **172**: 1165–1177. doi:10.1534/genetics.105.044586. PMID:16079235.

Dobson, A.J. 2002. An introduction to generalized linear models. 2nd ed. Chapman & Hall/CRC, Boca Raton, Fla.

Falush, D., Stephens, M., and Pritchard, J.K. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics, **164**: 1567–1587. PMID:12930761.

Farnir, F., Coppieters, W., Arranz, J.J., Berzi, P., Cambisano, N., Grisart, B., et al. 2000. Extensive genome-wide linkage disequilibrium in cattle. Genome Res. **10**: 220–227. doi:10.1101/gr.10.2.220. PMID:10673279.

Fisher, R.A. 1954. Statistical methods for research workers. 12th ed. Hafner Publishing Company Inc., New York.

Flint-Garcia, S.A., Thornsberry, J.M., and Buckler, E.S. 2003. Structure of linkage disequilibrium in plants. Annu. Rev. Plant Biol. **54**: 357–374. doi:10.1146/annurev.arplant.54.031902.134907. PMID:14502995.

Garris, A.J., McCouch, S.R., and Kresovich, S. 2003. Population structure and its effect on haplotype diversity and linkage disequilibrium surrounding the *xa5* locus of rice (*Oryza sativa* L.). Genetics, **165**: 759–769. PMID:14573486.

Gupta, P.K., Rustgi, S., and Kulwal, P.L. 2005. Linkage disequilibrium and association studies in higher plants: present status and future prospects. Plant Mol. Biol. **57**: 461–485. doi:10.1007/s11103-005-0257-z. PMID:15821975.

Jorde, L.B. 2000. Linkage disequilibrium and the search for complex disease genes. Genome Res. **10**: 1435–1444. doi:10.1101/gr.144500. PMID:11042143.

Koornneef, M., Alonso-Blanco, C., and Vreugdenhil, D. 2004. Naturally occurring genetic variation in *Arabidopsis thaliana*. Annu. Rev. Plant Biol. **55**: 141–172. doi:10.1146/annurev.arplant.55.031903.141605. PMID:15377217.

Kraakman, A.T.W., Niks, R.E., van den Berg, P.M.M.M., Stam, P., and van Eeuwijk, F.A. 2004. Linkage disequilibrium mapping of yield and yield stability in modern spring barley cultivars. Genetics, **168**: 435–446. doi:10.1534/genetics.104.026831. PMID:15454555.

Kraakman, A.T.W., Martinez, F., Mussiraliev, B., van Eeuwijk, F.A., and Niks, R.E. 2006. Linkage disequilibrium mapping of morphological, resistance, and other agronomically relevant traits in modern spring barley cultivars. Mol. Breed. **17**: 41–58. doi:10.1007/s11032-005-1119-8.

Kruglyak, L. 1999. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. Nat. Genet. **22**: 139–144. doi:10.1038/9642. PMID:10369254.

Lander, E.S., and Schork, N.J. 1994. Genetic dissection of complex traits. Science (Washington, D.C.), **265**: 2037–2048. doi:10.1126/science.8091226. PMID:8091226.

Lou, P., Zhao, J., Kim, J.S., Shen, S., Pino Del Carpio, D., Song, X., et al. 2007. Quantitative trait loci for flowering time and morphological traits in multiple populations of *Brassica rapa*. J. Exp. Bot. In press.

Mackay, I., and Powell, W. 2007. Methods for linkage disequilibrium mapping in crops. Trends Plant Sci. **12**: 57–63. doi:10.1016/j.tplants.2006.12.001. PMID:17224302.

Malosetti, M., Linden, C.G., Vosman, B., and Eeuwijk, F.A. 2007. A mixed model approach to association mapping using pedigree information with an illustration to resistance for *Phytophthora infestans* in potato. Genetics, **175**: 879–889. doi:10.1534/genetics.105.054932. PMID:17151263.

Mollers, C., Lickfett, T., Matthaus, B., and Velasco, L. 1999. Influence of P-fertilizer on phytic acid content in seeds of *Brassica napus* L. and development of a NIRS calibration. *In* Proceedings of the 10th International Rapeseed Congress, Canberra, Australia, 26–29 September 1999. *Edited by* N. Wratten and P.A. Salisbury. pp. 26–29.

Nordborg, M., Borevitz, J.O., Bergelson, J., Berry, C.C., Chory, J., Hagenblad, J., et al. 2002. The extent of linkage disequilibrium in *Arabidopsis thaliana.* Nat. Genet. **30**: 190–193. doi:10.1038/ng813. PMID:11780140.

Nordborg, M., Hu, T.T., Ishino, Y., Jhaveri, J., Toomajian, C., Zheng, H., et al. 2005. The pattern of polymorphism in *Arabidopsis thaliana*. PLoS Biol. **3**(7): e196. doi:10.1371/journal.pbio.0030196.

Nsengimana, J., Baret, P., Haley, C.S., and Visscher, P.M. 2004. Linkage disequilibrium in the domesticated pig. Genetics, **166**: 1395–1404. doi:10.1534/genetics.166.3.1395. PMID:15082558.

Palaisa, K.A., Morgante, M., Williams, M., and Rafalski, A. 2003. Contrasting effects of selection on sequence diversity and link-

age disequilibrium at two phytoene synthase loci. Plant Cell, **15**: 1795–1806. doi:10.1105/tpc.012526. PMID:12897253.

Payne, R.W., and Arnold, G.M. 2002. Genstat Release 6.1. Reference manual. Part 3. Procedure Library PL14. VSN International, Oxford, U.K.

Peng, J., Slominsiki, B.A., Guenter, W., Campbell, L.D., and Xiong, Y.Z. 2001. The anti-nutritional factors in Chinese double-low rapeseed meal. J Chinese Cereals and Oil Association, **16**: 6–10.

Pritchard, J.K., Stephens, M., and Donnelly, P. 2000. Inference of population structure using multilocus genotype data. Genetics, **155**: 945–959. PMID:10835412.

Raboy, V., Young, K.A., Dorsch, J.A., and Cook, A. 2001. Genetics and breeding of seed phosphorus and phytic acid. J. Plant Physiol. **158**: 489–497. doi:10.1078/0176-1617-00361.

Remington, D.L., Thornsberry, J.M., Matsuoka, Y., Wilson, L.M., Whitt, S.R., Doebley, J., et al. 2001. Structure of linkage disequilibrium and phenotypic associations in the maize genome. Proc. Natl. Acad. Sci. U.S.A. **98**: 11479–11484. doi:10.1073/pnas.201394398. PMID:11562485.

SAS Institute Inc. 1999. SAS/STAT®. Version 9.1 [computer program]. SAS Institute Inc., Cary, N.C.

Tenaillon, M.I., Sawkins, M.C., Long, A.D., Gaut, R.L., Doebley, J.F., and Gaut, B.S. 2001. Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). Proc. Natl. Acad. Sci. U.S.A. **98**: 9161–9166. doi:10.1073/pnas.151244298. PMID:11470895.

Varshney, R.K., Graner, A., and Sorrells, M.E. 2005. Genomics-assisted breeding for crop improvement. Trends Plant Sci. **10**: 621–630. doi:10.1016/j.tplants.2005.10.004. PMID:16290213.

White, P.J., and Broadley, M.R. 2005. Biofortifying crops with essential mineral elements. Trends Plant Sci. **10**: 586–593. doi:10.1016/j.tplants.2005.10.001. PMID:16271501.

Yu, J., Pressoir, G., Briggs, W.H., Bi, I.V., Yamasaki, M., Doebley, J.F., et al. 2006. A unified mixed-model method for association mapping accounting for multiple levels of relatedness. Nat. Genet. **38**: 203–208. doi:10.1038/ng1702. PMID:16380716.

Zhao, J. 2007. The genetics of phytate content and morphological traits in *Brassica rapa*. Ph.D. thesis. Wageningen University, the Netherlands. ISBN 90-8504-588-6. pp. 61–107.

Zhao, J., Wang, X., Deng, B., Lou, P., Wu, J., Sun, R., et al. 2005. Genetic relationships within *Brassica rapa* as inferred from AFLP fingerprints. Theor. Appl. Genet. **110**: 1301–1314. doi:10.1007/s00122-005-1967-y. PMID:15806345.