Institute of Agronomy and Plant Breeding I

Department of Plant Breeding

Justus Liebig University Giessen

Professor Dr. Rod Snowdon

# Functional markers, sub-genomic selection patterns and haplotype regions associated with seed glucosinolates and chlorophyll content in rapeseed (*Brassica napus* L.)

A thesis submitted for the requirement of the doctoral degree in

Agricultural Sciences from the Faculty of Agricultural and

Nutritional Sciences and Environmental Management

Justus Liebig University Giessen, Germany

Submitted by

**Lunwen Qian**

Giessen 2016

Institute of Agronomy and Plant Breeding I

Department of Plant Breeding

Justus Liebig University Giessen

# Functional markers, sub-genomic selection patterns and haplotype regions associated with seed glucosinolates and chlorophyll content in rapeseed (*Brassica napus* L.)

A thesis submitted for the requirement of the doctoral degree in

Agricultural Sciences from the Faculty of Agricultural and

Nutritional Sciences and Environmental Management

Justus Liebig University Giessen, Germany

Submitted by

**Lunwen Qian**

**Examiners**

Supervisor: Prof. Dr. Rod Snowdon

Co-supervisor: Prof. Dr. Matthias Frisch

Giessen 2016

**Table of Contents**

# 1 General Introduction

This chapter containing sections that have been previously published by the author in the following publications:

**Lunwen Qian**, Wei Qian and Rod J Snowdon (2014) Sub-genomic selection patterns as a signature of breeding in the allopolyploid *Brassica napus* genome. *BMC Genomics* 15, 1170.

**Lunwen Qian**, Wei Qian and Rod J Snowdon (2016) Haplotype hitchhiking promotes trait coselection in *Brassica napus*. *Plant Biotechnology Journal* 14, 1578-88.

Ying Fu, Kun Lu, **Lunwen Qian**, Jiaqin Mei, Dayong Wei, Xuhui Peng, Xinfu Xu, Jiana Li, Martin Frauen, Felix Dreyer, Rod J Snowdon and Wei Qian (2015) Development of genic cleavage markers in association with seed glucosinolate content in canola. *Theoretical and Applied Genetics* 128, 1029-37.

## 1.1 Rapeseed (*Brassica napus* L.)

The relationships between the six major cultivated *Brassica* species were originally described by U (1935), who associated the diploid species *B. rapa* (AA, 2n=20), *B. oleracea* (CC, 2n=18) and *B. nigra* (BB, 2n=16) with the amphiploids *B. juncea* (AABB, 2n=36), B. carinata (BBCC, 2n=34) and *B. napus* (AACC, 2n=38; Fig. 1.1). *Brassica napus* is derived from interspecific hybridisation events between *B. rapa* and *B. oleracea* (Allender and King 2010). Besides artificially synthesised *B. napus*, only cultivated forms are known, and genetic diversity analyses have revealed only a few eco-geographically and genetically distinct gene pools among cultivated *B. napus* (Bus et al. 2011; Hasan et al. 2006). These suggest that the species may have derived by independent interspecific hybridisation events in Europe and Asia.

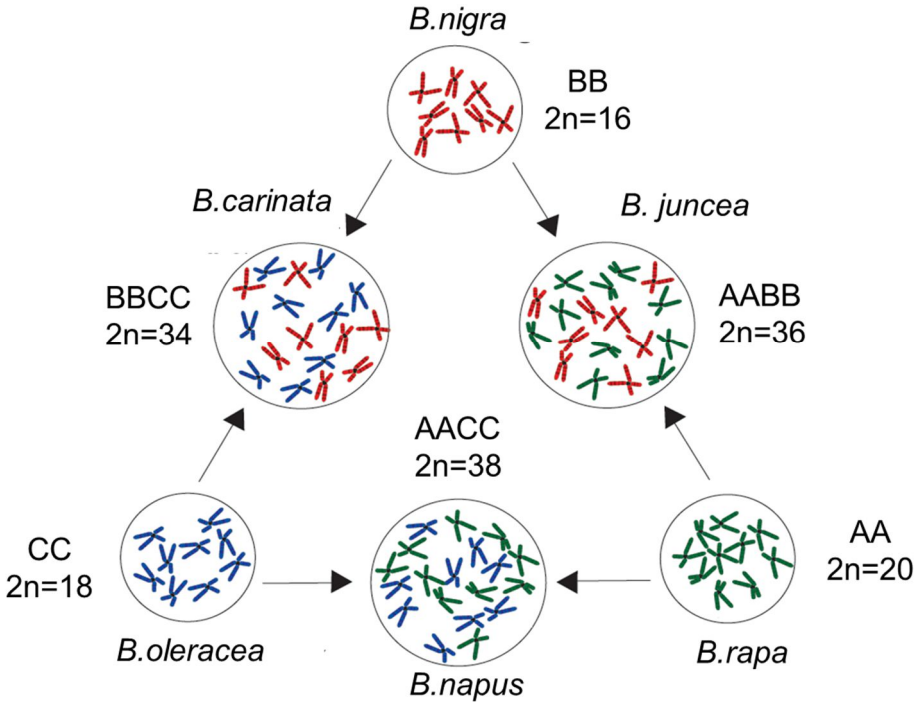



**Figure 1.1** The Brassica triangle of Nagahara U showing that genetic relationship between six species. n = chromosome number. A, B and C represent three different genomes.

*Brassica napus* (rapeseed, oilseed rape, Canola) is the second most important oilseed crop in the world after soybean. Worldwide production of rapeseed was 61 million metric tons (MT) in 2011 (http://faostat.fao.org/). China was the top rapeseed producing country, producing

14.7 million MT followed by India, producing about 7.3 million MT. The 27 countries of the European Union (EU) accounted for 23 million MT. Nowadays, soybeans are the major oilseed used for biodiesel production in United States while rapeseed is the most common oilseed used for biodiesel in Europe (http://agmrc.org/). Biodiesel made from rapeseed gels at a lower temperature than biodiesel produced from other feedstocks, making canola biodiesel a more suitable fuel for colder regions. Research at the University of Idaho showed that canola biodiesel had a "cloud point" of 1°C and a "pour point" of -9°C (Peterson et al. 1997). However, comparatively little of this crop is grown because the market for canola and edible rapeseed is much larger than the market for industrial rapeseed. There is often confusion between the use of the terms “rapeseed” and “canola.” Rapeseed is the traditional name for the group of oilseed crops in the *Brassicaceae* family. It can be divided into two types — industrial rapeseed or canola. Visually, the seeds of the two types are identical. The distinguishing difference between the two types is their individual chemical or fatty acid profiles. Generally, “industrial rapeseed” refers to any rapeseed with a high content (at least 45 percent) of erucic acid in the oil. Canola however refers to the edible oil crop that is characterized by low erucic acid (less than 2 percent; http://agmrc.org/).

In addition both tuberous (swede or rutabaga) and leafy forms (fodder rape and kale) of the species are grown as vegetables for human consumption and animal fodder. Oilseed *B. napus* has only achieved economic importance in the past forty years following an intensive breeding programme to minimise nutritionally undesirable components of the oil and meal, and to increase yield production. In some areas, rapeseed, which contains more than 40 percent oil content, becomes more profitable than soybeans, with only 18 percent oil content. Initially attention was focused on reducing levels of erucic acid in the seed oil, and then reducing levels of aliphatic glucosinolates (GSL) levels in the meal to make it more palatable and safer for livestock. As with other crops, ongoing breeding programmes aim to increase overall harvestable yield and quality, with resistance to crop pests and pathogens as major targets. Whilst successful, the collateral effect of these improvements has been the production of elite varieties that possess only a fraction of the genetic diversity available in

the wider *Brassica* genepools. This will cause an increase concern, particularly with respect to lack of resistance to pests and fungus. Sources of new alleles from closely related species that can be transferred into elite breeding lines are required in order to increase heterosis, to provide new functional disease resistance loci, and refine oil qualities to serve a variety of nutritional and industrial purposes (Mba et al. 2012).

### 1.2 Genome diversity improvement of rapeseed

The genetic basis of commonly used rapeseed is quite narrow which is limiting the breeding progress for this species (Gómez-Campo 1999). One important reason could be the sort of cultivation in comparison to its two parental species (Becker et al. 1995). Genetic diversity in Chinese *B. napus* has been further improved by introgressions from Asian *B. rapa* (Chen et al. 2008; Qian et al. 2006), the diploid donor of the *B. napus* A-subgenome. According to Liu (1985) and Shiga (1970), more than 50% of *B. napus* cultivars in China and Japan are derived from *B. napus* × *B. rapa* crossings. Qian et al. (2014) suggested that A genome have more genetic diversity than C genome and increases in genetic diversity and recombination within the C-genome is particularly important for breeding. According to Mei et al. (2011), natural *B. napus* has very low genetic diversity compared with its diploid progenitors, therefore intercrossing with the parental species could be an effective way to broaden genetic diversity in rapeseed. To achieve this, it may be necessary to overcome sexual compatibility barriers by using embryo rescue techniques. For example, in recent years considerable progress has been made in introducing novel C-genome donors to European winter oilseed rape, in order to improve genetic diversity particularly for disease resistance (Rygulla et al. 2007; Ding et al. 2013) or heterosis (Jesske et al. 2013).

Today's Asian semi-winter type rapeseed represents a major intermediate gene pool between European winter-type oilseed rape and spring-sown canola, grown primarily in North America. China's most important oilseed crop, grown on over 13 million hectare (ha), is also a potentially rich source of genetic variation to bring diversify within the narrow gene pools. Chinese rapeseed breeding has extensively used diploid *Brassica* species, particularly

*B. rapa*, to enrich the genetic potential of the local gene pool for resistance traits and to improve heterosis. Different *B. napus* gene pools have undergone strict selection for flowering-related traits, including vernalisation requirements, winter survival and photoperiod-dependant flowering, and for essential seed quality traits (primarily low erucic acid and GSL contents). Together with its recent alloploidisation, this makes *B. napus* an interesting model for investigating genome-wide and subgenome-specific patterns of genomic and allelic diversification, in the face of broad selective sweeps, during crop domestication.

### 1.3 Improvement of seed glucosinolate content

After the identification of the Polish spring rapeseed variety “Bronowski” as a source of low seed GSL in 1969 (Kondra and Stefansson 1970), great efforts were made to introgress this trait into breeding materials worldwide. Accessions with low seed glucosinolate (GSL) content (less than 30 µmoles/g meal), were initially released in Canada and became known as ‘canola’ (https://en.wikipedia.org/). European winter oilseed rape production was subsequently also converted to low GSL varieties having less than 18 µmoles/g in whole seeds. GSL biosynthesis in *Brassicas* has quantitative and sporophytic inheritance and is regulated by complex genetic factors, climate, and moisture availability during the growing season (Kondra and Stefansson, 1970; Halkier and Gershenzon, 2006). These influences are obvious in the reported values of GSLs in canola meals from different countries (Table 1). The maximum level of GSLs shown in Australian canola meal (Table 1.3) is important as it shows the upper level close to the maximum allowable level for canola (30 µmoles/g in oil free meal; www.dpi.nsw.gov.au).

**Table 1.3** GSL content of canola meat of different origin.

| Country of Origin | GSL range | Reference |
|---|---|---|
| Auctralia | 7.4-25.3 | (Mullan et al. 2000) |
| Canada | 11.4-21.7 | (Slominski et al., 1999) |
| Belgium | 13.8-33.0 | (Derycke et al., 1999) |
| Denmark | 10.7-18.3 | (Jensen et al., 1995) |

Note: GSLs reported as µmoles/gram in oil free meal.

Selection for low erucic acid and low GSL content has narrowed down the genetic pool and produced a genetic bottleneck in 00-rapeseed breeding material (Friedt and Snowdon, 2009; Hasan et al., 2006). The orthologues of HIGH ALIPHATIC GSL1 (HAG1), which controls aliphatic GSL biosynthesis in *A. thaliana*, were suggested as candidates for major QTL on A09, C02, C07 and C09 of rapeseed (Li et al. 2014; Hirai et al. 2007; Howell et al. 2003; Harperet al. 2012; Zhao and Meng 2003). These QTL form the basis of the major reduction in seed GSL content which has been achieved in worldwide canola breeding during the past three decades. Qian et al. (2014) confirmed that stronger artificial and natural selection caused expansive blocks of conserved linkage disquilibrium surrounding major C-subgenome QTL for GSL content in modern *B. napus* breeding materials, whereas the corresponding homoeologous regions in the A-subgenome were considerably more fragmented and diverse. This implies that an increase in genetic diversity and recombination within the C-genome will contribute to GSL content improvement. Mei et al. (2011) showed that wild *B.rapa* and *B.oleracea* can be an effective way to broaden A and C genome genetic diversity in rapeseed. C-genome donors to European winter oilseed rape improved genetic diversity particularly for disease resistance (Rygulla et al. 2007; Ding et al. 2013). Meanwhile other methods, for example gene identification, cloning and transformation combine with half-seed techniques and efficient analytical methods are being used to expedite the conventional breeding methods for ultra-low seed GSL content.

### 1.4 SNP array development in crops

A Single Nucleotide Polymorphism (SNP) is a DNA sequence variation occurring commonly within a population (e.g. 1%) in which a single nucleotide — A, T, C or G — in the genome (or other shared sequence) differs between members of a biological species or paired chromosomes. Almost all common SNPs have only two alleles. In molecular biology and bioinformatics, SNP array is a type of DNA microarray which is used to detect polymorphisms within a population. Recently SNP genotyping arrays are considered as one of the most important tools forfine mapping in crops (Chagné et al. 2007). There is great potential for tremendously fine genetic mapping using SNPs, which are the most abundant

form of DNA polymorphism. Local haplotype patterns surrounding densely-spaced SNP markers with significant trait associations can reveal information on selective sweeps and genome diversity associated with important crop traits.

For crop development, a large SNP array which defines its functionality over a diverse germplasm gene pool as well as its utility for fine linkage mapping and genome-wide association study (GWAS) is important. For example, more than four million high-quality SNPs identified from re-sequencing of 47 soybean accessions were used to select 180 961 SNPs for creation of the Axiom SoyaSNP array (Lee et al. 2015). A genotyping array including about 90,000 gene-associated SNPs was developed and used to characterise genetic variation in allohexaploid and allotetraploid wheat populations (Wang et al. 2014). Four cultivars of the important amphidiploid oilseed species *Brassica napus* were used to develop a *B. napus* Infinium™ array containing 5,306 SNPs (Dalton-Morgan et al. 2014). These SNP arrays have been used to identity correlations between genes and haplotype regions for differerent traits and to reveal information on regions with selective sweeps. In 2012 an international consortium, in collaboration with Illumina Inc. (San Diego, CA, USA) developed and released a Brassica SNP array containing assays for 60,000 SNPs, mainly from *B. napus* (Edwards et al. 2013). In the meantime this array has become the standard tool for high-resolution genetic analysis in rapeseed.

## 1.5 Population structure, linkage disequilibrium (LD) and haplotype block in diverse population

A well-known problem with genome-wide association studies (GWAS) is the presence of undetected population structure, which can lead to both false-positive results and a failure to detect genuine associations (Marchini et al. 2004). Meanwhile, population size also affects power of test correlation between marker and phenotype traits. So, a diversified germplasm collection plays a key role in genomic research for any crop species.

The concept of LD describes the non-random association of alleles at two or more loci caused by genetic linkage. Many evolutionary and genetic factors can influence LD. In

particular it can reflect the history of natural and artificial selection, mutation, segmental recombination rates, gene conversion and other forces that cause selective sweeps in a genome. Estimates of the extent of LD decay in crop genomes vary depending on the particular species, gene pool or population under investigation. For example, in a diversity set of different sorghum collections a decay of LD was reported to occur within 15-20 Kb (Hamblin et al. 2005), 50-100 Kb (Bouchet et al. 2012) and 400 Kb (Bekele et al. 2013). Less variation was observed in different populations of maize, with estimates between 0.5-7.0 Kb (Remington et al. 2001; Ching et al. 2002; Palaisa et al. 2003) and 1-10 Kb (Yan et al. 2009), and rice, with estimates of 20-50 cM (Jin et al. 2010) and 75-150 Kb (Mather et al. 2007). In *Arabidopsis,* LD has been estimated from 50 Kb (Nordborg et al. 2005) to over 250 Kb (Hagenblad et al. 2002). In different *B. napus* populations average LD estimates based on genetic distance measurements were also estimated to vary greatly, from 1-2 cM (Ecke et al. 2010) to more than 20 cM (Zou et al. 2010).

Some studies have demonstrated that SNPs in strong LD are organised into discrete haplotype blocks that are possibly separated by hotspots of recombination. Genetic variation across the genome is defined by these haplotype blocks, while species-specific block structure is defined by the differential contribution of population history effects in combination with mutation and recombination events. Conservation of haplotype structure may therefore be used for the identification and characterization of functionally important genomic regions during evolution and/or selection. For example, high-resolution analysis of human Y-chromosome haplotypes suggested that a large component of a present-day Asian gene pool originates from Eastern Africa and that Asia was the source of a back-migration to sub-Saharan Africa (Cruciani et al. 2002). Haplotype map analysis in maize found hundreds of selective sweeps and highly differentiated regions that probably contain loci that are keys to geographic adaptation (Gore et al. 2009). High-throughput SNP genotyping technologies today enable the use of large numbers of SNPs to construct high resolution LD and haplotype block maps. This is crucial for accurate understanding of associations between

markers, genes and phenotypic traits, and at the same time can give more in-depth understanding with regard to species evolution.

Natural and artificial selection can cause conservation of haplotype blocks, comprising specific combinations of nucleotides on the same chromosome, in genome regions carrying genes under positive or negative selection. Haplotypes can therefore provide more information than any single SNP regarding the complex relationship between DNA variation and quantitative phenotypes (Stephens et al. 2001). Elucidating the evolutionary relationships among local haplotypes can further improve the detection power of GWAS scans (Buntjer et al. 2005). Detailed analysis of LD surrounding major quantitative trait loci (QTL) revealed strong signatures of artificial selection associated with important traits in different breeding pools of rapeseed and bread wheat (Qian et al. 2014; Voss-Fels et al. 2015; Voss-Fels and Snowdon 2015). In such cases, haplotypes reveal the extent to which genetic variation in a given chromosome region is described by clustering markers. Comparing haplotype diversity can help to understand the effects of natural and artificial selection on genome-scale and single-gene variation, as shown recently in grapevine, maize and *Arabidposis* (Fernandez et al. 2014; Yang et al. 2013; Li et al. 2014). New high-density genome screening tools provide an unprecedented level of insight into local LD patterns in even complex crop genomes (Edwards et al. 2013). For example, clear haplotype patterns detected in high-density population genomic studies could be associated with domestication, adaptation and breeding in sorghum (Mace et al. 2013), rapeseed (Qian et al. 2014) and bread wheat (Voss-Fels et al. 2015), respectively.

### 1.6 Genome wide association studies (GWAS) in crops

Linkage mapping is a key tool for identifying the genetic basis of quantitative traits in plants. Most agronomic traits in crops are controlled by complex quantitative trait loci (QTL) and their genetic basis is frequently dissected using QTL mapping. In regards to the world's second most important oilseed crop, a vast number of studies have reported QTL for various agronomic, developmental, seed quality and resistance traits since the first genetic mapping

of QTL in this species by Uzunova et al. (1995). Bi-parental genetic mapping populations can be limited by low polymorphism or small population size, however. In addition, only two alleles per locus and few recombination events are considered to estimate the genetic distance between marker loci and to identify the causative genomic regions underlying QTL, thereby limiting the mapping resolution. Although the power of QTL detection in bi-parental mapping populations is generally high, the value of the detected QTL for breeding is often underscored by unpredictable effects in different genetic backgrounds.

Association genetics approaches, which utilise genetically unrelated collections or populations of varieties and breeding lines, are a useful alternative for QTL localisation (Flint-Garcia *et al*. 2003). In contrast to conventional QTL mapping, association mapping is based on linkage disequilibrium (LD). Utilisation of the higher number of historical recombinations in less related populations can greatly improve the mapping resolution compared to a segregating bi-parental population (Ersoz *et al*., 2007). With the rapid developments in genomics and dramatically decreasing cost of genotyping technology, association mapping has become a direct and efficient approach to dissect for complex traits. So far, genome-wide association mapping has contributed to reveal genetic architecture of complex traits in rapeseed, rice, maize, *Arabidopsis* and so on (Li et al. 2014; Zhao et al. 2011; Cook et al. 2012; Atwell et al. 2010). A major prerequisite for association mapping is the availability of densely-spaced, molecular markers spanning the entire genome. The discovery and implementation of genome-wide screening for SNP markers, even in complex polyploid crop species like *B. napus*, has advanced extremely rapidly in recent years since the introduction of ultrafast DNA sequencing technologies (Edwards et al. 2013).

### 1.7 Scope and Aims

Low GSL content accession selection has narrowed down the genetic pool and created a genetic bottleneck, and genetic diversity analysis in major QTL regions of GLS content is very important for development of cultivars with ultra-low GSL content. This dissertation aimed to investigate diversity for GSL content and related chlorophyll traits in populations

with Asian rapeseed background. Quantitative trait loci (QTL) for seed glucosinolate (GSL) content in a low-GSL genetic background were mapped over seven environments in Germany and China in a doubled haploid population from a cross between two low-GSL oilseed rape parents with transgressive segregation. Sequence polymorphisms between the corresponding coding regions of the parental lines were used to develop cleaved amplified polymorphic site markers for two QTL linked genes, ISOPROPYLMALATE DEHYDROGENASE1 and ADENOSINE 5′-PHOSPHOSULFATE REDUCTASE3 (Chapter 2). High-density single-nucleotide polymorphism (SNP) genotyping arrays, a powerful tool for GWAS, can give valuable insight into patterns of population structure, linkage disequilibrium (LD) and haplotype analysis in significantly trait-associated genome regions. Here, high density-SNP markers were used to analyze A and C genome genetic diversity in 203 Chinese semi-winter rapeseed inbred lines (Chapter 3). Local haplotype patterns surrounding SNP markers with significant associations with chlorophyll and GSL content were found, as well as correlation between haplogroup-related phenotypes in in different chromosome haplotype regions (Chapter 4).

This work set out with following overall aims:

- Development of CAPS markers to facilitate breeding for ultra-low GSL in rapeseed.
- Use of the 60k SNP Illumina Brassica consortium genotyping array to assess the influence of selection and breeding for the major QTL region of important qulity traits (GSL and erucic acid content) on LD and haplotype structure in a diverse panel of 203 Chinese semi-winter rapeseed breeding lines, and disclose correlations between genome diversity and trait improvement.
- Correlations of haplotype diversity to phenotype groups, and analysis of gene content in conserved haplotype blocks.
- Detection of co-selection for chlorophyll and GSL content in different regions of the genome.

# 2 Development of genic cleavage markers in association with seed glucosinolate content in canola

Ying Fu, Kun Lu, Lunwen Qian, Jiaqin Mei, Dayong Wei,

Xuhui Peng, Xinfu Xu, Jiana Li, Martin Frauen,

Felix Dreyer, Rod J Snowdon and Wei Qian

*Theoretical and Applied Genetics* 2015

Vol 128, pp. 1029-1037

Theor Appl Genet (2015) 128:1029–1037
DOI 10.1007/s00122-015-2487-z

ORIGINAL PAPER

# Development of genic cleavage markers in association with seed glucosinolate content in canola

**Ying Fu · Kun Lu · Lunwen Qian · Jiaqin Mei · Dayong Wei · Xuhui Peng · Xinfu Xu · Jiana Li · Martin Frauen · Felix Dreyer · Rod J. Snowdon · Wei Qian**

Received: 18 July 2014 / Accepted: 24 February 2015 / Published online: 8 March 2015


**Abstract**

***Key message*** **The orthologues of *Arabidopsis* involved in seed glucosinolates metabolism within QTL confidence intervals were identified, and functional markers were developed to facilitate breeding for ultra-low glucosinolates in canola.**

*Abstract* Further reducing the content of seed glucosinolates will have a positive impact on the seed quality of canola (*Brassica napus*). In this study 43 quantitative trait loci (QTL) for seed glucosinolate (GSL) content in a low-GSL genetic background were mapped over seven environments in Germany and China in a doubled haploid population from a cross between two low-GSL oilseed rape parents with transgressive segregation. By anchoring these QTL to the reference genomes of *B. rapa* and *B. oleracea,* we identified 23 orthologues of Arabidopsis involved in GSL metabolism within the QTL confidence intervals. Sequence polymorphisms between the corresponding coding regions of the parental lines were used to develop cleaved amplified polymorphic site markers for two QTL-linked genes, *ISOPROPYLMALATE DEHYDROGENASE1* and *ADENOSINE 5′-PHOSPHOSULFATE REDUCTASE 3*. The genic cleavage markers were mapped in the DH population into the corresponding intervals of QTL explaining 3.36–6.88 and 4.55–8.67 % of the phenotypic variation for seed GSL, respectively. The markers will facilitate breeding for ultra-low seed GSL content in canola.

Communicated by M. L. Federico.

**Electronic supplementary material** The online version of this article (doi:10.1007/s00122-015-2487-z) contains supplementary material, which is available to authorized users.

Y. Fu · K. Lu · L. Qian · J. Mei · D. Wei · X. Peng · X. Xu · J. Li · W. Qian (✉)
College of Agronomy and Biotechnology, Southwest University, 400716 Chongqing, China
e-mail: qianwei666@hotmail.com

Y. Fu · L. Qian · J. Mei · R. J. Snowdon (✉)
Department of Plant Breeding, IFZ Research Centre for Biosystems, Land Use and Nutrition, Justus Liebig University, Heinrich-Buff-Ring 26-32, 35392 Giessen, Germany
e-mail: Rod.Snowdon@agrar.uni-giessen.de

M. Frauen · F. Dreyer
Norddeutsche Pflanzenzucht Hans-Georg Lembke KG, 24363 Hohenlieth, Germany

## Introduction

*Brassica napus* (AACC, $2n = 38$) is an allopolyploid that originated from spontaneous hybridisations between *B. rapa* (AA, $2n = 20$) and *B. oleracea* (CC, $2n = 18$) and is today one of the world's most important oil crops. The meal obtained after oil extraction contains 35–40 % of high-quality protein and is a valuable animal feed (Dimov et al. 2012; Leckband et al. 2002; Wanasundara 2011). However, high quantities of glucosinolates (GSL) and their degradation products, which occur naturally in the tissues of all *Brassica* species, can cause problems with palatability, along with thyroid, liver, and kidney abnormalities (Walker and Booth 2001) and especially limit the use of meal as a feed supplement for monogastric livestock.

After the identification of the Polish spring rapeseed variety "Bronowski" as a source of low seed GSL in 1969 (Kondra and Stefansson 1970), great efforts were made to introgress this trait into breeding materials worldwide. Accessions with low seed GSL content (less than 30 μmol $g^{-1}$ meal), were initially released in Canada and became known as 'canola'. European winter oilseed rape

production was subsequently also converted to low GSL varieties having less than 18 μmol $g^{-1}$ in whole seeds. Further reductions in GSL content would increase the ratio of oilseed rape/canola meal that can be used particularly for feeding of monogastric livestock.

The core pathway of GSL biosynthesis is well-known from the model crucifer *Arabidopsis*, a relative of *B. napus*. The genes involved in amino acid chain elongation, core structure and side-chain formation have been identified, along with the structural genes responsible for most biosynthetic steps (Wittstock and Halkier 2000; Bak and Feyereisen 2001; Grubb and Abel 2006; Mikkelsen et al. 2004; Piotrowski et al. 2004; Halkier and Gershenzon 2006). For instance, *ISOPROPYLMALATE DEHYDROGENASE1* (*IPMDH1*) is co-expressed with nearly all the genes known to be involved in aliphatic GSL biosynthesis (Gigolashvili et al. 2009; Binder et al. 2007; He et al. 2009), and the disruption of *IPMDH1* in *Arabidopsis* causes a dramatic decrease in the concentrations of GSL with side chains of four or more carbons (He et al. 2009). Another gene, *ADENOSINE 5′-PHOSPHOSULFATE REDUCTASE 3* (*APR3*), is involved in the synthesis of activated sulfate in the GSL biosynthesis network (Lee et al. 2011).

A large number of studies have identified quantitative trait loci (QTL) for seed GSL content in crucifers (Uzunova et al. 1995; Howell et al. 2003; Mahmood et al. 2003; Quijada et al. 2006; Sharpe and Lydiate 2003; Zhao and Meng 2003; Basunanda et al. 2007; Hasan et al. 2008; Harper et al. 2012; Javidfar and Cheng 2013; Li et al. 2014). However, previous efforts to determine the genetic basis of the low seed GSL trait in *B. napus* have focused on a small number of major QTL with large effects on the phenotypic variance. All of the above studies utilized crosses involving at least one parent with high seed GSL content, however. Reliable detection of minor QTL that segregate between different low-GSL materials was therefore masked by the strong effects of a few major QTL. Recently, the orthologues of *HIGH ALIPHATIC GLUCOSINOLATE 1* (*HAG1*), which controls aliphatic GSL biosynthesis in *A. thaliana*, were suggested as candidates for major QTL on A09, C02, C07 and C09 of rapeseed (Li et al. 2014; Hirai et al. 2007; Howell et al. 2003; Harper et al. 2012; Zhao and Meng 2003). These QTL form the basis of the major reduction in seed GSL content which has been achieved in worldwide canola breeding during the past three decades. On the other hand, further reductions in seed GSL require combination of these main-effect loci with additional QTL that have less prominent effects. Such loci are often overlooked in mapping studies that use parents carrying main-effect loci, making it difficult to implement them into breeding programmes. In this study, 43 QTL for seed GSL content were dissected over multiple environments in a doubled haploid (DH) population derived from two low-GSL parents. Among these QTL, cleaved amplified polymorphic site markers were developed for *B. napus* orthologues of the glucosinolate biosynthesis genes *IPMDH1* and *APR3*, both located within the confidence intervals of interesting QTL.

## Materials and methods

### Plant materials and phenotypic evaluation

A mapping population consisting of 261 DH lines was developed by microspore culture, using a single F1 plant derived from a cross between an inbred line of the European winter oilseed rape cultivar 'Express' (female) and the Chinese semi-winter breeding line 'SWU07' (male). The parents were selected because both have low seed GSL content (<30 μmol $g^{-1}$ meal) when grown in their respective countries of origin, but previous QTL mapping studies in crosses with high-GSL parents (Basunanda et al. 2007; Fu et al. unpublished data) suggested they may carry different minor-effect QTL facilitating transgressive segregation.

The DH population was grown at Chongqing, China, for 5 years (from 2009 to 2013) and at Hohenlieth, Germany for 2 years (2008, 2012). These locations represent extremely different environments, ranging from the subtropical continental basin of the Yangtze River to the cool maritime climate of the Baltic Sea. A randomised complete block design with two replications was employed. Seeds were harvested from ten self-pollinated plants per genotype for quality analysis. Total seed GSL content from each seed sample was determined by near-infrared reflectance spectroscopy (NIRS), with two technical replicates. The seed GSL content (μmol $g^{-1}$ meal) was calculated after subtracting oil content in seed, which also was determined by NIRS.

### Statistical analysis

Analysis of variance (ANOVA) was performed for GSL over multiple environments using the GLM procedure of SAS (SAS Institute, SAS and Institute 2000). The broad-sense heritability ($H^2$) was calculated as follows: $H^2 = \sigma_g^2/(\sigma_g^2 + \sigma_{ge}^2/n + \sigma_e^2/nr)$, where $\sigma_g^2$, $\sigma_{ge}^2$ and $\sigma_e^2$ are estimates of the variances of genotype, genotype × environment interactions and error, respectively, $n$ is the number of environments and $r$ is the number of replications per environment (Hallauer and Miranda 1988). Pearson's product–moment correlation coefficient between traits of interest was calculated using the CORR procedure of SAS (SAS and Institute 2000).

### Construction of linkage groups and QTL mapping

Genomic DNA was extracted from pooled leaves of ten plants of each DH line. Simple sequence repeat (SSR) primers were either obtained from public datasets (Lowe et al. 2004; Piquemal et al. 2005; http://ukcrop.net/ace/search/BrassicaDB; http://www.brassica.info/ssr/SSRinfo.htm; http://www.osbornlab.agronomy.wisc.edu/research/maps/ssrs.html) or designed according to the genome sequences of *Brassica* species and *Arabidopsis* (markers with prefix CEN, FM, POD, SWUC and YD).

Genetic linkage groups were constructed using the software JOINMAP 3.0 (Stam 1993) and assigned to chromosomes using published positions of common SSR markers (Suwabe et al. 2002; Gao et al. 2006; Long et al. 2007; Shi et al. 2009). QTL detection was performed with the composite interval mapping (CIM) procedure of the software WinQTL Cartographer 2.5 (Wang et al. 2005). A 1000-permutation test was performed to estimate a significance threshold of the test statistic for a QTL based upon a 5 % experiment-wise error rate (Churchill and Doerge 1994).

### Alignment of QTL into reference genomes of *B. rapa* and *B. oleracea*

The QTL intervals of GSL were aligned into the reference genome of *B. rapa* (http://brassicadb.org) and *B. oleracea* (http://ocri-genomics.org) by BLAST analysis of the sequences of SSR markers linked with QTL or their primers when the full sequences were not available. The top significant hits returned by 'BLASTn' (*e* value $\leq$0.005) were used to infer the putative physical positions of these markers on the *B. oleracea* and *B. rapa* genomes. When a marker had multiple amplification loci on a same chromosome, the accurate position for a particular locus was determined manually by referring to the physical positions of its upstream and downstream markers. Potential candidate genes for GSL were identified by annotation analyses of genes within the physical boundaries of the QTL confidence intervals.

### Gene sequencing and development of CAPS markers

The standard molecular cloning procedure of Sambrook and Russell (2001) was followed to isolate the genomic sequence of the candidate genes between the two parents. The sequences of *BnaIPMDH1-A02* and *BnaAPR3-A03* were aligned among clones using the software VectorNTI (www.invitrogen.com/VectorNTI).

The primer pairs for amplifying the complete open reading frame (ORF) were BnaIPMDH1-A02F (5′-ATGGCGG CAGCTTTACAAACGAAC-3′) and BnaIPMDH1-A02R (5′-TTAAACAGTAGCTGTAACTTTGGAGTCCAC-3′) for *BnaIPMDH1-A02*, and BnaAPR3-A03F (5′-ATGGC ACTAGCAATCAACGTTTCTTCATC-3′) and BnaAPR3-A 03R (5′-TTACCTAACAAGATTCAAGAAAGATGTCAA AGAATCA-3′) for *BnaAPR3-A03*. Based on the results of identified sequence polymorphisms, the amplified ORFs of *BnaIPMDH1-A02* and *BnaAPR3-A03* were digested by the restriction enzymes *Sac*I and *Hin*dIII, respectively, in each of the 261 DH lines and the two parental lines. The PCR products were digested for 1 h at 37 °C in a total volume of 20 µl with 5 units of *Sac*I or *Hin*dIII (Thermo Scientific). The digested PCR products were separated in 1.2 % agarose gels, stained with ethidium bromide and visualized under UV light.

## Results

### Variation for seed GSL content

Averaged over all years, the winter oilseed rape parental line 'Express' exhibited 29.02 and 40.03 µmol total GSL $g^{-1}$ meal in selfed seeds from the field trials in Germany and China, respectively, while the semi-winter parental line 'SWU07' exhibited 16.02 and 28.87 µmol total GSL $g^{-1}$ meal in Germany and China, respectively. A transgressive segregation of GSL was detected in the DH population, with normal distribution of values ranging from 10.56 to 88.64 µmol $g^{-1}$ meal (Fig. 1). This indicates that the two parental lines carry different QTL for GSL. Although the overall values of GSL in Germany were lower than those of China, significant and positive correlations were detected between environments, with correlation coefficients ranging from 0.43 to 0.85 ($P < 0.01$) (Table S1).

The results of ANOVA revealed significant differences among genotypes, environments and genotype-by-environment interactions for GSL ($P < 0.01$) (Table 1), which was in accordance with the moderate heritability of GSL across environments ($H^2 = 68.89$ %).

### Microsynteny analysis of QTL and identification of candidate genes

A total of 316 SSR markers were placed into 19 linkage groups. Via common markers these were subsequently assigned to the 19 *B. napus* chromosomes, designated A01–A10 and C01–C09. The genetic map spanned a genetic distance of 1198 cm, with an average distance of 3.79 cm between adjacent markers.

The QTL analysis via CIM procedure in the software WinQTL Cartographer 2.5 revealed a total of 43 individual QTL for seed GSL content in individual environments, located across seven chromosomes (A02, A03, A04, A07, A09, C03 and C08) and each explaining between 3.35

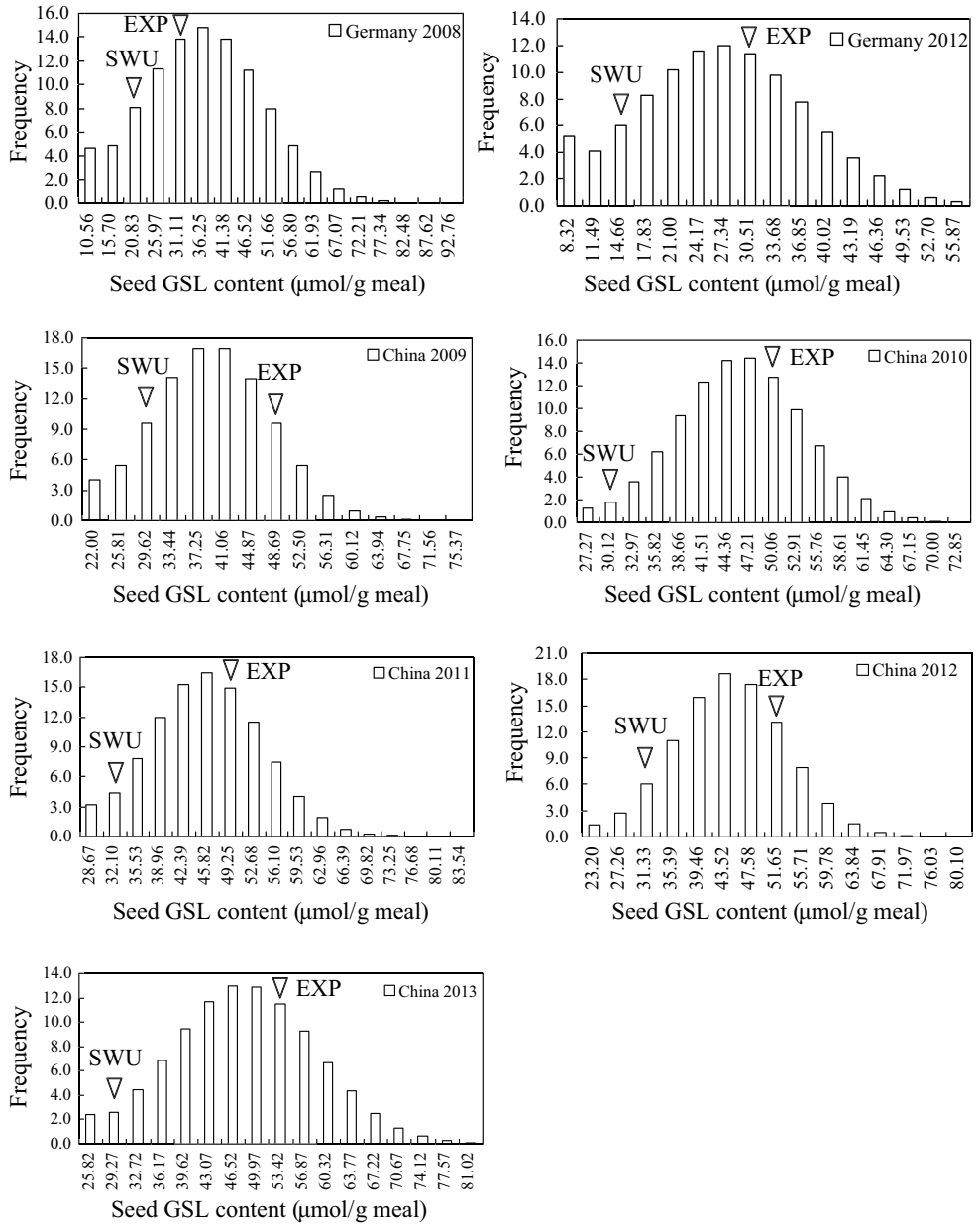



**Fig. 1** Frequency distributions for seed GSL content in the DH population Express*SWU07, grown in seven environments in Germany and China from 2008 to 2013

**Table 1** Analysis of variance and heritability for GSL content in DH population

| Source | Df | Mean square |
|---|---|---|
| Genotype (*G*) | 260 | 650.68* |
| Environment (*E*) | 6 | 21,969.71* |
| $G \times E$ | 1287 | 67.55* |

* Significance at $P = 0.01$

and 31.97 % of the phenotypic variation (Table S2). The confidence intervals for 31 single-environment QTL overlapped between Chinese and German environments (Table S2), suggesting that at these genome positions the same loci were influencing the GSL content under different environments.

In the physical genome regions of *B. rapa* and *B. oleracea,* delineated by SSR markers or SSR primers flanking the QTL peaks, a total of 23 orthologues

of Arabidopsis genes whose annotations suggest an involvement in GSL metabolism were found in 28 QTL regions (Tables S2, S3). In order to develop functional markers linked with GSL, these orthologues were sequenced to screen polymorphisms between two parents. Based on the discovered sequence polymorphisms, cleaved amplified polymorphic site (CAPS) markers were successfully developed within two orthologues, *BnaIPMDH1-A02* and *BnaAPR3-A03* (Fig. 2). The open reading frame of *BnaIPMDH1-A02* was represented by a component of 1902 and 1893 bp in 'Express' and 'SWU07', respectively, exhibiting more than 90 % sequence similarity against the sequence of *AtIPMDH1* (*AT5G14200*) from *Arabidopsis* and *BrIPMDH1-A02* (Bra023450) from *B. rapa*. The alignment of the *BnaIPMDH1-A02* sequence revealed 25 SNP polymorphisms and 1 insertion/deletion polymorphism between 'Express' and 'SWU07' across the entire coding region (Fig. S2). Among these, the presence of an SNP (T → C) at 871 bp of 'Express' and 873 bp of 'SWU07' resulted in the creation of a *Sac*I restriction site (5′… GAGCT ▼ C… 3′) in 'SWU07' (Fig. 2). This means that *Sac*I can digest the genomic sequence of *BnaIPMDH1-A02* from SWU07 into two fractions (873 and 1021 bp), whereas the *BnaIPMDH1-A02* sequence from 'Express' remains uncleaved (Fig. 3).

The *BnaAPR3-A03* sequences from 'Express' (1587 bp) and 'SWU07' (1579 bp) exhibited more than 85 and 90 % sequence similarity to *A. thaliana AtAPR3* (*AT4G21990*) and *B. rapa BrAPR3* (Bra019406), respectively. Amongst a number of SNP polymorphisms and 5 insertion/deletion polymorphisms between 'Express' and 'SWU07' (Fig. S3), the presence of one SNP (T → C) within the exon at 733 bp of the 'Express' sequence and 725 bp of the 'SWU07' sequence resulted in the creation of a *Hin*dIII restriction site (5′…A▼AGCTT…3′) in 'Express' but not in 'SWU07' (Fig. 2). This enabled *Hin*dIII to cleave the genomic sequence of *BnaAPR3-A03* in 'Express' into

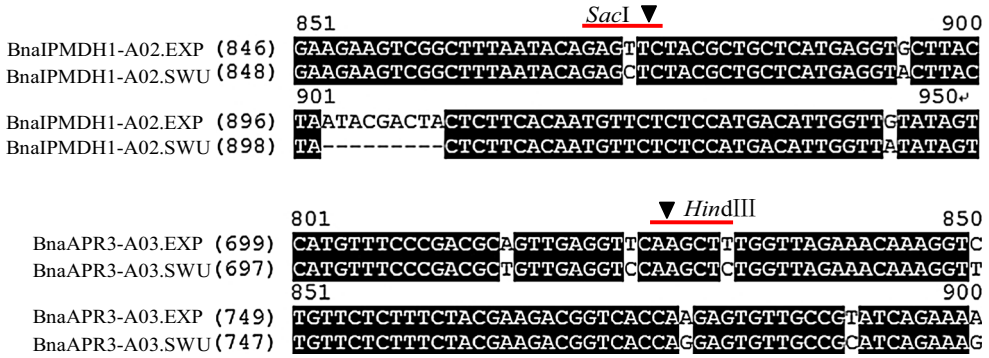



**Fig. 2** Alignment of the partial sequence of *BnaIPMDH1-A02* and *BnaAPR3-A03* containing single nucleotide polymorphisms (SNPs) which resulted in the creation of the restriction sites for *Sac*I and *Hin*dIII, respectively, between Express and SWU07. *Lines* represent the restriction sites, while *arrows* indicate the positions of the cleavage sites in the respective sequences

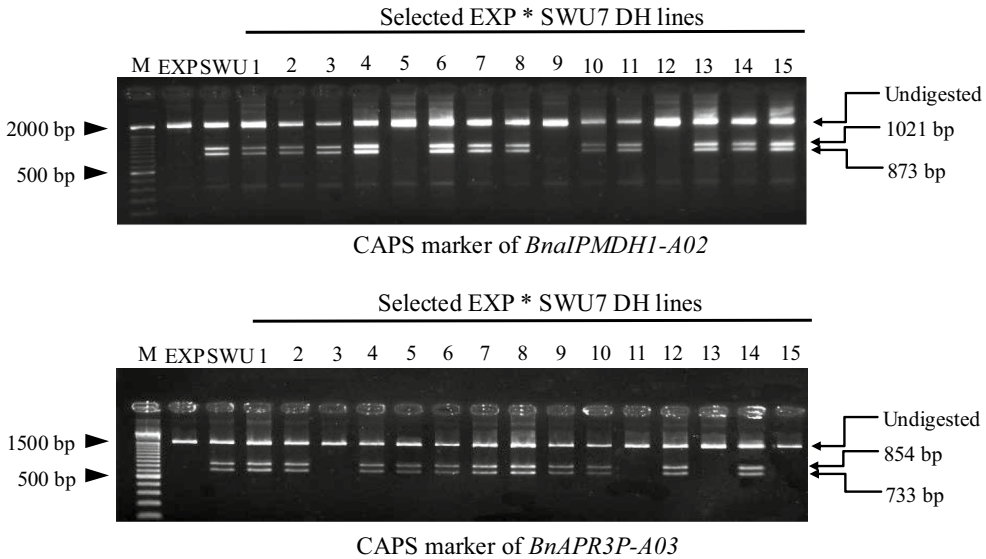



**Fig. 3** Digestion of genomic sequence for *BnaIPMDH1-A02* and *BnaAPR3-A03* by *Sac*I and *Hin*dIII in the parental lines 'Express' and 'SWU07' along with DH *lines*

**Table 2** QTL for GSL content on chromosomes A02 and A03 before and after integrating CAPS markers within *BnaIPMDH1-A02* and *BnaAPR3-A03*, respectively

| QTL | Chr[a] | Env[b] | Add[c] | QTL mapping | | QTL re-mapping with CAPS markers | | |
|---|---|---|---|---|---|---|---|---|
| | | | | Pos[d] | $R^2$ (%)[e] | Pos. | CAPS pos[f] | $R^2$ (%) |
| qGSLC09A02 | A02 | C2009 | – | 3.2–11.2 | 4.69 | 0–11.7 | 10.43 | 4.07 |
| qGSLC10A02 | A02 | C2010 | – | | | 3.2–12.2 | 10.43 | 3.36 |
| qGSLC11A02 | A02 | C2011 | – | 3.2–10 | 5.87 | 4.7–12.5 | 10.43 | 6.9 |
| qGSLG08A02 | A02 | G2008 | – | 6–11 | 5.99 | 5.1–12.2 | 10.43 | 5.33 |
| qGSLG12A02 | A02 | G2012 | – | 5.5–11 | 5.01 | 5.1–12.2 | 10.43 | 5.84 |
| qGSLC13A02 | A02 | C2013 | – | | | 6.1–11.6 | 10.43 | 3.89 |
| qGSLG08A03 | A03 | G2008 | + | 34.6–41.6 | 4.1 | 34.9–39.8 | 38.02 | 4.55 |
| qGSLG12A03 | A03 | G2012 | + | 35–40.4 | 8.92 | 35.3–39.4 | 38.02 | 8.78 |
| qGSLC09A03 | A03 | C2009 | + | 34.8–41.2 | 13.5 | 34.4–38.5 | 38.02 | 4.67 |
| qGSLC10A03 | A03 | C2010 | + | | | 35.8–39.5 | 38.02 | 6.7 |

[a] Chromosome

[b] Environments for field trials: C, China; G, Germany; the suffix number represents the year

[c] Additive effect. The direction of additive effect is from the allele of 'Express', while a negative additive effect indicates an allelic contribution from 'SWU07'

[d] Length of 2-LOD score confidence interval

[e] Percentage of the phenotypic variance explained by each QTL

[f] Genetic map position of CAPS marker in linkage group

two fractions (733 and 854 bp), whereas the *BnaAPR3-A03* sequence from 'SWU07' remains undigested (Fig. 3).

In order to test the phenotypic effects on seed GSL content estimated by the two CAPS markers, we genotyped the DH population using both markers (Fig. 3). For the marker BnaIPMDH1-A02.CAPS, the GSL content in the DH group with the allele from 'Express' was significantly lower than that of 'SWU07' in all environments ($P < 0.01$), with the GSL reduced by an average of 1.65 μmol $g^{-1}$ meal. For the marker BnaAPR3-A03.CAPS, the GSL content of the group containing the allele from 'SWU07' was significantly lower than that of the group carrying the 'Express' allele ($P < 0.01$), with the GSL reduced by 3.33 μmol $g^{-1}$ meal. Furthermore, mapping of the CAPS markers back to the genetic map confirmed the localisation of genes. *BnaIPMDH1-A02* was located within a QTL cluster from 3.2 cm to 11.2 cm on A02 (between markers 'C2.141' and 'A2.246') (Table 2). The corresponding region harbours four overlapping QTL for seed GSL content, explaining 4.69–5.99 % of the phenotypic variation over environments (Table 2). *BnaAPR3-A03* was mapped within the QTL region from 34.6 to 41.6 cm on chromosome A03 (flanking markers, 'CNU146' and 'A3.1487b') (Table 2). This chromosome segment carries three overlapping QTL for seed GSL content with 4.1–13.5 % genetic effects of individual QTL over environments (Table 2). These findings show the association of both *BnaIPMDH1-A02* and *BnaAPR3-A03* with GSL content.

## Discussion

A total of 43 GSL QTL were detected across seven environments in this study. Among these QTL, 41 were detected in the *B. napus* A-subgenome and only 2 in the C-subgenome. This bias possibly relates to the broader genetic variation within the A-subgenome of modern *B. napus* cultivars, which has been achieved through implementation of *B. rapa* germplasm via interspecific hybridisation (Wang et al. 2014). In accordance with this assumption, Qian et al. (2014) confirmed the presence of expansive blocks of conserved linkage disquilibrium surrounding major C-subgenome QTL for glucosinolate content in modern *B. napus* breeding materials, whereas the corresponding homoeologous regions in the A-subgenome were considerably more fragmented and diverse. A similar bias towards A-subgenome QTL for agronomic traits in *B. napus* was also reported by Shi et al. (2009). The majority of the detected QTL in our study were confirmed in highly distinct environments in China and Germany (Table S2). This result is in accordance with the moderate heritability and the correlation for GSL detected across environments.

Plant comparative mapping is a powerful tool which not only reveals the processes and rates of genome evolution, but also allows the transfer of genetic knowledge between species (Parkin et al. 2005). These advantages are particularly important in *Brassicaceae*, where a huge quantity of functional genomic information is available for the model

crucifer *A. thaliana*. The release of the genome sequences of *B. rapa* (Wang et al. 2011) and *B. oleracea* (Liu et al. 2014), the progenitor species of *B. napus*, opened the possibility to directly translate knowledge on important genes and pathways from Arabidopsis to oilseed rape, even before the availability of the recently released *B. napus* reference sequence (Chalhoub et al. 2014).

In this study, we identified candidate genes from the GSL biosynthesis pathway within the confidence intervals of QTL that segregate transgressively in low-GSL (double-low, canola quality) oilseed rape. This represents a first step towards potential cloning of genes having minor but positive effects on this valuable seed quality trait and provides useful CAPS markers for breeding of ultra-low GSL content in oilseed rape and canola meals. Although confirmation of the candidate genes will require additional functional analyses, for breeding purposes the availability of effective markers tightly linked to interesting QTL can be of considerable assistance both for identification of useful new diversity and for its introgression into elite materials by marker-assisted backcrossing. This is particularly true for traits like seed GSL content, where small-effect loci are often masked by a few large-effect QTL and field testing in multiple environments is necessary for accurate selection of ultra-low phenotypes.

Defatted rapeseed meal is enriched with a high-quality protein with a desirable amino acid composition similar to soybean protein. Reducing GSL content will further improve the nutritive value of meal in oilseed rape and canola meals. Although the low GSL trait in all current *B. napus* cultivars derives from the same founder accession, Bronowski (Krzymanski 1970), wide variances for the composition of GSL were detected in canola (Li et al. 2005). This suggests the presence of additional genetic factors besides the well-described major QTL. Hutcheson et al. (2000) developed an ultra-low GSL spring-type turnip rape (*B. rapa*) with a seed GSL content of only 4.2 μmol $g^{-1}$ meal. This material derived from a cross between members of a low aliphatic GSL *B. rapa* breeding population (BC86-18) and the low indole GSL *B. rapa* breeding population DLY (Hutcheson et al. 2000). We also detected numerous minor-effects QTL for GSL content in a DH population derived from two lines of oilseed rape that each have low seed GSL. Those findings indicated variant alleles for seed GSL content in canola.

We performed a literature review for QTL reported in rapeseed in linkage mapping studies or by association approaches (Uzunova et al. 1995; Howell et al. 2003; Zhao and Meng 2003; Li et al. 2014) and compared the published QTL with those detected in this study. Four major QTL on chromosomes A09, C02, C07 and C09, which were detected independently in different studies (Uzunova et al. 1995; Howell et al. 2003; Zhao and Meng 2003), were positioned at four common regions at 3.2, 50.0, 39.9 and 2.8 Mb of A09, C02, C07 and C09, respectively, which were proved to be associated with homologues of the key gene controlling aliphatic glucosinolate biosynthesis (*HAG1*, At5g61420) (Li et al. 2014). Nevertheless, none of the 43 QTL detected in this study overlap or were nearby these major-effect loci. Routine phenotypic selection for low GSL content is generally performed in advanced, homogeneous breeding generations by use of high-throughput near-infrared spectroscopy (NIRS; Wittkop et al. 2009). Although this approach has been highly successful in the breeding of canola-quality *B. napus*, environmentally induced variation among low-GSL materials makes it challenging to implement new loci with additional small effects using conventional selection methods. The markers identified in this work and the plant materials that carry them, therefore, represent important resources to develop ultra-low GSL canola by pyramiding low GSL alleles with both major and minor effects. A previous study by Hasan et al. (2008) confirmed marker–trait associations of gene-linked SSR markers to seed glucosinolate content in genetically diverse *B. napus* germplasm. These included loci associated with homologues of *CYP79A2* and *MAM1/MAM3*, which we also identified within QTL intervals in the present study. Due to the narrow genetic background of the low GSL trait in current canola and oilseed rape cultivars, most of which carry common major-effect QTL from the same origin, we expect the novel minor QTL we describe in this work to be generally effective across a broad range Asian, North American and European breeding materials.

**Author contribution statement** Y.F. conducted all experiments and wrote the manuscript, K.L. designed primers, L.Q., J.M., D.W, X.P., X.X. participated in the field experiments and seed quality anlysis, W.Q. designed the experiment, J.L., M.F., F.D., W.Q. and R.S. directed the project and contributed to the writing.

**Acknowledgments** We thank Norddeutsche Pflanzenzucht Hans-Georg Lembke KG, Germany, for developing the DH population. This study was supported by grant 973 (2015CB150201), the Key Projects in the National Science and Technology (2014BAD01B07), the Fundamental Research Funds for the Central Universities (XDJK2013A013, XDJK2014C148, XDJK2014B036, XDJK2014A015, SWU113106, SWU113065) and NSFC (31171585, 31471529).

**Conflict of interest** The authors declare that they have no conflict of interest.

## References

Bak S, Feyereisen R (2001) The involvement of two P450 enzymes, CYP83B1 and CYP83A1, in auxin homeostasis and glucosinolate biosynthesis. Plant Physiol 127:108–118

Basunanda P, Spiller TH, Hasan M, Gehringer A, Schondelmaier J, Lühs W, Friedt W, Snowdon RJ (2007) Marker-assisted increase of genetic diversity in a double-low seed quality winter oilseed rape genetic background. Plant Breed 126:581–587

Binder S, Knill T, Schuster J (2007) Branched-chain amino acid metabolism in higher plants. Physiol Plant 129:68–78

Chalhoub B, Denoeud F, Liu S, Parkin IAP, Tang H, Wang X, Chiquet J, Belcram H, Tong C, Samans B, Corréa M, Da Silva C, Just J, Falentin C, Koh CS, Le Clainche I, Bernard M, Bento P, Noel B, Labadie K, Alberti A, Charles M, Arnaud D, Guo H, Daviaud C, Alamery S, Jabbari K, Zhao M, Edger PP, Chelaifa H, Tack D, Lassalle G, Mestiri I, Schnel N, Le Paslier M-C, Fan G, Renault V, Bayer PE, Golicz AA, Manoli S, Lee T-H, Thi VHD, Chalabi S, Hu Q, Fan C, Tollenaere R, Lu Y, Battail C, Shen J, Sidebottom CHD, Wang X, Canaguier A, Chauveau A, Bérard A, Deniot G, Guan M, Liu Z, Sun F, Lim YP, Lyons E, Town CD, Bancroft I, Wang X, Meng J, Ma J, Pires JC, King GJ, Brunel D, Delourme R, Renard M, Aury J-M, Adams KL, Batley J, Snowdon RJ, Tost J, Edwards D, Zhou Y, Hua W, Sharpe AG, Paterson AH, Guan C, Wincker P (2014) Early allopolyploid evolution in the post-neolithic *Brassica napus* oilseed genome. Science 346:950–953

Churchill GA, Doerge RW (1994) Empirical threshold values for quantitative trait mapping. Genetics 138:963–971

Dimov Z, Suprianto E, Hermann F, Möllers C (2012) Genetic variation for seed hull and fibre content in a collection of European winter oilseed rape material (*Brassica napus* L.) and development of NIRS calibrations. Plant Breed 131:361–368

Gao M, Li G, Potter D, McCombie WR, Quiros CF (2006) Comparative analysis of methylthio alkylmalate synthase (MAM) gene family and flanking DNA sequences in *Brassica oleracea* and *Arabidopsis thaliana*. Plant Cell Rep 25:592–598

Gigolashvili T, Yatusevich R, Rollwitz I, Humphry M, Gershenzon J, Flügge UI (2009) The plastidic bile acid transporter 5 is required for the biosynthesis of methionine-derived glucosinolates in *Arabidopsis thaliana*. Plant Cell 21:1813–1829

Grubb CD, Abel S (2006) Glucosinolate metabolism and its control. Trends Plant Sci 11:89–100

Halkier BA, Gershenzon J (2006) Biology and biochemistry of glucosinolates. Annu Rev Plant Biol 57:303–333

Hallauer AR, Miranda J (1988) Quantitative genetics in maize breeding. Iowa State University, Ames

Harper AL, Trick M, Higgins J, Fraser F, Clissold L, Wells R, Hattori C, Werner P, Bancroft I (2012) Associative transcriptomics of traits in the polyploid crop species *B. napus*. Nat Biotechnol 30:798–802

Hasan M, Friedt W, Pons-Kühnemann J, Freitag NM, Link K, Snowdon RJ (2008) Association of gene-linked SSR markers to seed glucosinolate content in oilseed rape (*Brassica napus* ssp. *napus*). Theor Appl Genet 116:1035–1049

He Y, Mawhinney TP, Preuss ML, Schroeder AC, Chen B, Abraham L, Jez JM, Chen S (2009) A redox-active isopropylmalate dehydrogenase functions in the biosynthesis of glucosinolates and leucine in *Arabidopsis*. Plant J 60:679–690

Hirai MY, Sugiyama K, Sawada Y, Tohge T, Obayashi T, Suzuki A, Araki R, Sakurai N, Suzuki H, Aoki K, Goda H, Nishizawa OI, Shibata D, Saito K (2007) Omics-based identification of *Arabidopsis* Myb transcription factors regulating aliphatic glucosinolate biosynthesis. Proc Natl Acad Sci 104:6478–6483

Howell PM, Sharpe AG, Lydiate DJ (2003) Homoeologous loci control the accumulation of seed glucosinolates in oilseed rape (*B. napus*). Genome 46:454–460

Hutcheson DS, Falk KC, Rakow GFW (2000) TR4 summer turnip rape. Can J Plant Sci 80:837–838

Javidfar F, Cheng B (2013) Construction of a genetic linkage map and QTL analysis of erucic acid content and glucosinolate components in yellow mustard (*Sinapis alba* L.). BMC Plant Biol 13:142

Kondra ZP, Stefansson BR (1970) Inheritance of the major glucosinolates of rapeseed (*Brassica napus*) meal. Can J Plant Sci 50:643–647

Krzymanski J (1970) Inheritance of thioglucoside content by rapeseed (*B. napus*). J Int Sur le Colza Paris 37:213–218

Leckband G, Frauen M, Friedt W (2002) NAPUS 2000. Rapeseed (*B. napus*) breeding for improved human nutrition. Food Res Int 35:273–278

Lee BR, Koprivova A, Kopriva S (2011) The key enzyme of sulfate assimilation, adenosine 5′-phosphosulfate reductase, is regulated by HY5 in *Arabidopsis*. Plant J 67:1042–1054

Li PW, Zhao YG, Zhang W, Ding XX, Yang M, Wang XF, Xie CH, Fu TD (2005) Analysis of glucosinolate components and profiles in *Brassica napus*. Scientia Agricultura Sinica 38:1346–1352

Li F, Chen B, Xu K, , Wu J, Song W, , Bancroft I, Harper AL, Trick M, , Liu S, , Gao G, Wang N, Yan G, , Li J, , Qiao J, , Xiao X, Zhang T, Wu X (2014) Genome-wide association study dissects the genetic architecture of seed weight and seed quality in rapeseed (*Brassica napus* L.). DNA Res 21:355–367

Liu S, Liu Y, Yang X et al (2014) The *Brassica oleracea* genome reveals the asymmetrical evolution of polyploid genomes. Nat Commun 5:3930

Long Y, Shi J, Qiu D, Li R, Zhang C, Wang J, Hou J, Zhao J, Shi L, Beom-Seok Park, Choi SR, Lim YP, Meng J (2007) Flowering time quantitative trait loci analysis of oilseed *Brassica* in multiple environments and genomewide alignment with *Arabidopsis*. Genetics 177:2433–2444

Lowe AJ, Moule C, Trick M, Edwards KJ (2004) Efficient large-scale development of microsatellites for marker and mapping applications in Brassica crop species. Theor Appl Genet 108:1103–1112

Mahmood T, Ekuere U, Yeh F, Good AG, Stringam GR (2003) Molecular mapping of seed aliphatic glucosinolates in *Brassica juncea*. Genome 46:753–760

Mikkelsen MD, Naur P, Halkier BA (2004) *Arabidopsis* mutants in the C-S lyase of glucosinolate biosynthesis establish a critical role for indole-3-acetaldoxime in auxin homeostasis. Plant J 37:770–777

Parkin IAP, Gulden SM, Sharpe AG, Lukens L, Trick M, Osborn TC, Lydiate DJ (2005) Segmental structure of the *Brassica napus* genome based on comparative analysis with *Arabidopsis thaliana*. Genetics 171:765–781

Piotrowski M, Schemenewitz A, Lopukhina A, Müller A, Janowitz T, Weiler EW, Oecking C (2004) Desulfoglucosinolate sulfotransferase from *Arabidopsis thaliana* catalyze the final step in the biosynthesis of the glucosinolate core structure. J Biol Chem 279:50717–50725

Piquemal J, Cinquin E, Couton F, Rondeau C, Seignoret E, Doucet I, Perret D, Villeger MJ, Vincourt P, Blanchard P (2005) Construction of an oilseed rape (*Brassica napus* L.) genetic map with SSR markers. Theor Appl Genet 111:1514–1523

Qian L, Qian W, Snowdon RJ (2014) Sub-genomic selection patterns as a signature of breeding in the allopolyploid *Brassica napus* genome. BMC Genomics in press

Quijada PA, Udall JA, Lambert B, Osborn TC (2006) Quantitative trait analysis of seed yield and other complex traits in hybrid spring rapeseed (*Brassica napus* L.):1. Identification of genomic regions from winter germplasm. Theor Appl Genet 113:549–561

Sambrook J, Russell DW (2001) Molecular cloning: a laboratory manual (3-volume set). Cold Spring Harbor, USA

SAS and Institute (2000) SAS/STAT user's guide, version 8. SAS Institute, Cary

Sharpe AG, Lydiate DJ (2003) Mapping the mosaic of ancestral genotypes in a cultivar of oilseed rape (*Brassica napus*) selected via pedigree breeding. Genome 46:461–468

Shi J, Li R, Qiu D, Jiang CC, Long Y, Morgan C, Bancroft I, Zhao J, Meng JL (2009) Unraveling the complex trait of crop yield with quantitative trait loci mapping in *Brassica napus*. Genetics 182:851–861

Stam P (1993) Construction of integrated genetic linkage maps by means of a new computer package: join Map. Plant J 3:739–744

Suwabe K, Iketani H, Nunome T, Kage T, Hirai M (2002) Isolation and characterization of microsatellites in *Brassica rapa* L. Theor Appl Genet 104:1092–1098

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

Walker KC, Booth EJ (2001) Agricultural aspects of rape and other *Brassica* products. Eur J Lipid Sci Technol 103:441–446

Wanasundara JPD (2011) Proteins of Brassicaceae oilseeds and their potential as a plant protein source. Crit Rev Food Sci Nutr 51:635–677

Wang S, Basten CJ, Basten CJ, Zeng ZB (2005) Windows QTL cartographer version 2.5. Statistical genetics. North Carolina State University, Raleigh

Wang X, Wang H, Wang J et al (2011) The genome of the mesopolyploid crop species *Brassica rapa*. Nat Genet 43:1035–1039

Wang N, Li F, Chen B, Xu K, Yan G, Qian J, Li J, Gao G, Bancroft I, Meng J, King GJ, Wu X (2014) Genome-wide investigation of genetic changes during modern breeding of *Brassica napus*. Theor Appl Genet 127:1817–1829

Wittkop B, Snowdon RJ, Friedt W (2009) Status and perspectives of breeding for enhanced yield and quality of oilseed crops for Europe. Euphytica 170:131–140

Wittstock U, Halkier BA (2000) Cytochrome P450 CYP79A2 from *Arabidopsis thaliana* L. catalyzes the conversion of L-phenylalanine to phenylacetaldoxime in the biosynthesis of benzylglucosinolate. J Biol Chem 275:14659–14666

Zhao J, Meng J (2003) Detection of loci controlling seed glucosinolate content and their association with Sclerotinia resistance in *Brassica napus*. Plant Breed 122:19–23

# 3 Sub-genomic selection patterns as a signature of breeding in the allopolyploid *Brassica napus* genome

Lunwen Qian, Wei Qian and Rod J Snowdon

*BMC genomics* 2014

Vol 15, pp.1170

Qian et al. BMC Genomics 2014, 15:1170
http://www.biomedcentral.com/1471-2164/15/1170

BMC Genomics

RESEARCH ARTICLE Open Access

# Sub-genomic selection patterns as a signature of breeding in the allopolyploid *Brassica napus* genome

Lunwen Qian[1], Wei Qian[2] and Rod J Snowdon[1*]

## Abstract

**Background:** High-density single-nucleotide polymorphism (SNP) genotyping arrays are a powerful tool for genome-wide association studies and can give valuable insight into patterns of population structure and linkage disequilibrium (LD). In this study we used the Brassica 60kSNP Illumina consortium genotyping array to assess the influence of selection and breeding for important agronomic traits on LD and haplotype structure in a diverse panel of 203 Chinese semi-winter rapeseed (*Brassica napus*) breeding lines.

**Results:** Population structure and principal coordinate analysis, using a subset of the SNPs, revealed diversification into three subpopulations and one mixed population, reflecting targeted introgressions from external gene pools during breeding. Pairwise LD analysis within the A- and C-subgenomes of allopolyploid *B. napus* revealed that mean LD, at a threshold of $r^2 = 0.1$, decayed on average around ten times more rapidly in the A-subgenome (0.25-0.30 Mb) than in the C-subgenome (2.00-2.50 Mb). A total of 3,097 conserved haplotype blocks were detected over a total length of 182.49 Mb (15.17% of the genome). The mean size of haplotype blocks was considerably longer in the C-subgenome (102.85 Kb) than in the A-subgenome (33.51 Kb), and extremely large conserved haplotype blocks were found on a number of C-genome chromosomes. Comparative sequence analysis revealed conserved blocks containing homoloeogous quantitative trait loci (QTL) for seed erucic acid and glucosinolate content, two key seed quality traits under strong agronomic selection. Interestingly, C-subgenome QTL were associated with considerably greater conservation of LD than their corresponding A-subgenome homoeologues.

**Conclusions:** The data we present in this paper provide evidence for strong selection of large chromosome regions associated with important rapeseed seed quality traits conferred by C-subgenome QTL. This implies that an increase in genetic diversity and recombination within the C-genome is particularly important for breeding. The resolution of genome-wide association studies is also expected to vary greatly across different genome regions.

**Keywords:** Rapeseed, Population genomics, LD, Haplotype, Selection

## Background

Linkage mapping is a key tool for identifying the genetic basis of quantitative traits in plants. Most agronomic traits in crops are controlled by complex quantitative trait loci (QTL) and their genetic basis is frequently dissected using QTL mapping. In rapeseed (oilseed rape, canola: *Brassica napus* L.), the world's second most important oilseed crop, a vast number of studies have reported QTL for various agronomic, developmental, seed quality and resistance traits since the first genetic mapping of QTL in this species by [1]. Bi-parental genetic mapping populations can be limited by low polymorphism or small population size, however. In addition, only two alleles per locus and few recombination events are considered to estimate the genetic distance between marker loci and to identify the causative genomic regions underlying QTL, thereby limiting the mapping resolution. Although the power of QTL detection in bi-parental mapping populations is generally high, the value of the detected QTL for breeding is often underscored by unpredictable effects in different genetic backgrounds.

* Correspondence: rod.snowdon@agrar.uni-giessen.de
[1]Department of Plant Breeding, IFZ Research Centre for Biosystems, Land Use and Nutrition, Justus Liebig University, Heinrich-Buff-Ring 26-32, 35392 Giessen, Germany
Full list of author information is available at the end of the article

Association genetics approaches, utilising genetically unrelated collections or populations of varieties and breeding lines, are a useful alternative for QTL localisation [2]. In contrast to conventional QTL mapping, association mapping is based on linkage disequilibrium (LD). Utilisation of the higher number of historical recombinations in less related populations can greatly improve the mapping resolution compared to a segregating bi-parental population [3]. In recent years association mapping has been broadly adopted for quantitative genetic analyses in crop species [4-6]. A major prerequisite for association mapping is the availability of densely-spaced, molecular markers spanning the entire genome. The discovery and implementation of genome-wide screening for single-nucleotide polymorphism (SNP) markers, even in complex polyploid crop species like *B. napus*, has advanced extremely rapidly in recent years since the introduction of ultrafast DNA sequencing technologies [7]. High-density SNP arrays like the *Brassica* 60 k SNP Illumina consortium array (Illumina, San Diego, CA, USA) have opened the way for high-resolution QTL analyses based on linkage disequilibrium (LD) in both major and minor crops.

A well-known problem with genome-wide association studies (GWAS) is the presence of undetected population structure, which can lead to both false-positive results and a failure to detect genuine associations [8]. Because it also strongly influences LD patterns [2], an accurate estimate and understanding of population structure is critically important for association mapping. On the other hand, LD analyses also provide important insight into the history of both natural and artificial selection (breeding) and can give valuable guidance to breeders seeking to diversify crop gene pools. Recent studies of different *B. napus* ecotypes using collections of genome-wide simple-sequence repeat (SSR) markers gave first insight into genetic diversity and population structure in large collections of *B. napus* [9,10]. However studies with limited numbers of PCR-based markers are often unable to capture the full extent of LD in diverse populations, and conclusions are limited when data on LD cannot be accurately related to genomic positions of the markers.

The concept of LD describes the non-random association of alleles at two or more loci caused by genetic linkage. Many evolutionary and genetic factors can influence LD. In particular it can reflect the history of natural and artificial selection, mutation, segmental recombination rates, gene conversion and other forces that cause selective sweeps in a genome. Estimates of the extent of LD decay in crop genomes vary depending on the specific species, gene pool or population under investigation. For example, in different sorghum diversity collections a decay of LD was reported to occur within 15–20 Kb [11], 50–100 Kb [12] and 400 kb [13]. Less variation was observed in different populations of maize, with estimates between 0.5-7.0 kb [14-16] and 1–10 kb [17], and rice, with estimates of 20–50 cM [18] and 75–150 Kb [19]. In Arabidopsis LD has been estimated from 50 Kb [20] to over 250 kb [21]. In different *B. napus* populations average LD estimates based on genetic distance measurements were also estimated to vary greatly, from 1–2 cM [22] to more than 20 cM [23].

Some studies have demonstrated that SNPs in strong LD are organised into discrete haplotype blocks that are possibly separated by hotspots of recombination. Genetic variation across the genome is defined by these haplotype blocks, while species-specific block structure is defined by the differential contribution of population history effects in combination with mutation and recombination events. Conservation of haplotype structure may therefore be used for the identification and characterization of functionally important genomic regions during evolution and/or selection. For example, high-resolution analysis of human Y-chromosome haplotypes suggested that a large component of a present-day Asian gene pool originates from Eastern Africa and that Asia was the source of a back-migration to sub-Saharan Africa [24]. Haplotype map analysis in maize found hundreds of selective sweeps and highly differentiated regions that probably contain loci that are keys to geographic adaptation [25]. High-throughput SNP genotyping technologies today enable the use of large numbers of SNPs to construct high resolution LD and haplotype block maps. This is crucial for accurate understanding of associations between markers, genes and phenotypic traits, and at the same time can give more in-depth understanding with regard to species evolution.

Low seed glucosinolate and erucic acid concentrations are two of the most important traits for rapeseed breeding. Both traits have undergone intense purifying selection in elite varieties during the short history of this crop. Detailed analyses of LD and haplotype blocks surrounding major QTL for these two traits [26] will provide valuable new information about selective sweeps and potential linkage drag in the corresponding chromosome areas. At the same time these QTL provide interesting examples to study the dynamics of recent selection signatures at homoeologous trait loci in an important allopolyploid crop species.

*Brassica napus* is a very recent allopolyploid (genome AACC, 2n = 38), derived from only a small number of interspecific hybridisation events between *B. rapa* (AA, 2n = 20) and *B. oleracea* (CC, 2n = 18) within just the past few thousand years [27]. Besides artificially synthesised *B. napus*, only cultivated forms are known, and genetic diversity analyses have revealed only a few eco-geographically and genetically distinct gene pools among cultivated *B. napus* [9,28]. These suggest that the species may have derived by independent interspecific hybridisation events

in Europe and Asia. Today's Asian semi-winter type rapeseed represents a major intermediate gene pool between European winter-type oilseed rape and spring-sown canola, grown primarily in North America. China's most important oilseed crop, grown on over 13 million ha, is therefore also a potentially rich source of genetic variation to diversify these narrow gene pools. Chinese rapeseed breeding has extensively used diploid *Brassica* species, particularly *B. rapa*, to enrich the genetic potential of the local gene pool for resistance traits and to improve heterosis. Different *B. napus* gene pools have undergone strict selection for flowering-related traits, including vernalisation requirement, winter survival and photoperiod-dependant flowering, and for essential seed quality traits (primarily low erucic acid and glucosinolate contents). Together with its recent alloploidisation this makes *B. napus* an interesting model for investigating genome-wide and subgenome-specific patterns of genomic and allelic diversification, in the face of broad selective sweeps, during crop domestication.

With these aspects in mind the objectives of this study were: (1) to evaluate genome-specific patterns of population structure and genetic diversity in Chinese semi-winter rapeseed using densely spaced genome-wide SNP markers, (2) to study the extent of LD decay and variation in the distribution of haplotype block size within the A- and C-subgenomes, and (3) to study the effects of intense selection for major seed quality QTL on homoeologous genome regions.

## Results

### Genome-wide SNP polymorphism

From the total of 52,157 SNPs called by the cluster file to be polymorphic in the diversity panel, a stringent BLAST alignment (zero mismatches) of their flanking sequences to the draft *B. napus* reference genome identified 10,065 SNPs with potentially two or more loci in the *B. napus* genome, along with 6,930 SNPs showing no identical BLAST hit. A total of 35,162 single-locus SNPs, each mapping to a single physical genome position, were henceforth implemented for the downstream analyses. For the LD and population structure analysis, 10,168 SNPs with MAF <0.05 were also eliminated, leaving 24,994 high-quality, polymorphic, single-locus SNPs with MAF ≥0.05. The genotype data for these 24,994 SNPs in the diversity panel are provided in Additional file 1 along with their flanking sequence information and expected chromosome positions in the *B. napus* Darmor-*Bzh* reference genome [30].

### Population structure and diversity analysis in the A- and C-subgenomes

The results of the population structure analysis measured using the model-based software STRUCTURE are shown in Figure 1a. The LnP(D) value for each given *K* increased together with *K*, the most significant change being observed when *K* increased from 2 to 3. Over all iterations of the *ΔK* calculation a much higher likelihood was shown for *K* = 3 than for *K* = 4-10. This suggests the presence of 3 main subpopulations, hereinafter designated Q1, Q2 and Q3 (Figure 1b). Subpopulation Q1 includes 86 Chinese semi-winter, 1 spring and 2 winter rapeseed accessions, while subpopulation Q2 contains 32 Chinese semi-winter rapeseed lines. Q3 contains 27 semi-winter lines, 3 spring-type and 1 winter-type rapeseed. The remaining 60 accessions, including 1 spring-type and 1 winter rapeseed accession, were classified into a mixed subpopulation as they had membership probabilities lower than 0.60 in any given subpopulation (Additional file 2).

The PCA based on Nei's genetic distance analysis reflected the STRUCTURE results, with the mixed subpopulation clustering in the middle of the three defined subpopulations (Figure 2a). The first principal component (PC1) accounted for 14.0% of the genetic variation and roughly grouped the semi-winter rapeseed into the three main groups Q1, Q2 and Q3. The second principal component (PC2) accounted for 10.7% of the genetic variation and particularly reflected the differentiation between Q1 and Q2.

Comparative analysis of genetic diversity within the three subpopulations revealed higher average PIC and gene diversity in subpopulations Q1 and Q3 than in Q2

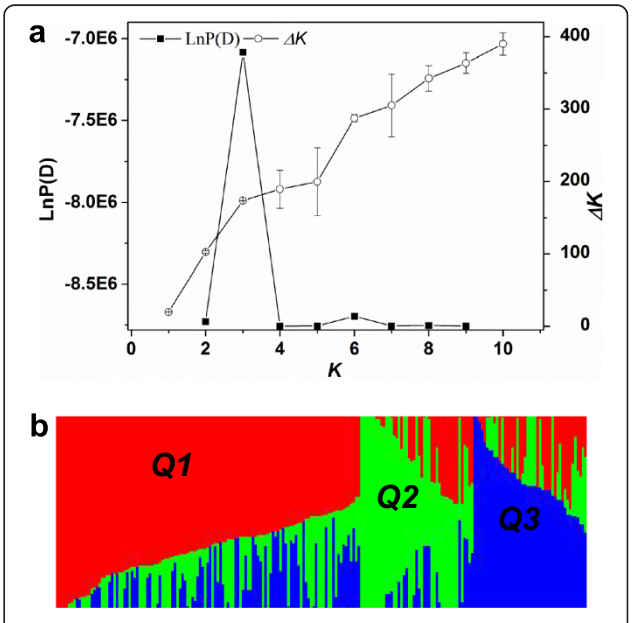



**Figure 1 Analysis of population structure by STRUCTURE in a total of 212 *Brassica napus* accessions genotyped with genome-wide SNP markers. a)** Results of estimated LnP(D) and *ΔK* analysis; **b)** Dissection of the genotypes into three distinct subpopulations using *K* = 3.

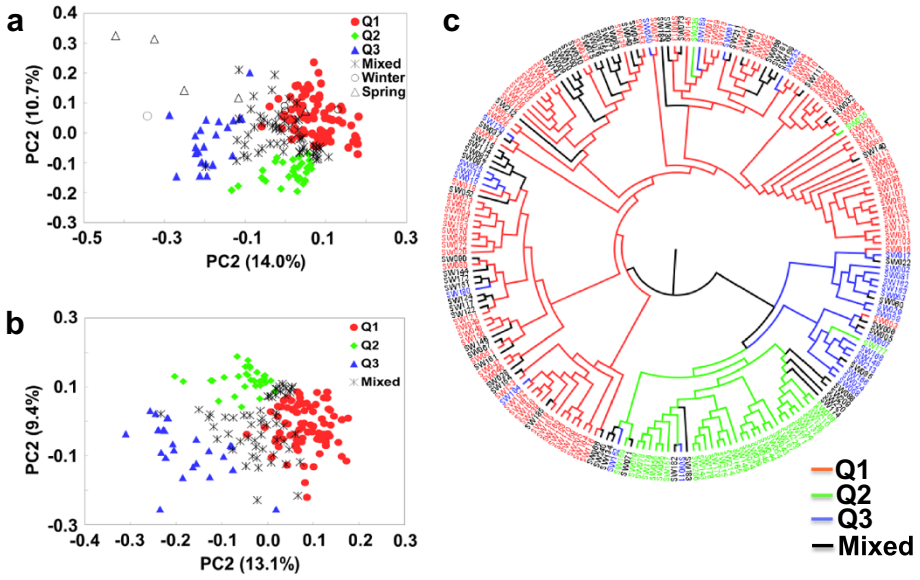



**Figure 2 Principal component analysis (PCA) and unweighted pair group matrix algorithm (UPGMA) tree describing genetic relationships among 212 winter, spring and semi-winter type *B. napus* accessions.** Q1, Q2 and Q3 are the three subgroups identified by STRUCTURE, assigned with the maximum membership probability, while the mixed subpopulation contains accessions that could not be specifically assigned by STRUCTURE to a single subpopulation. **a)** PCA analysis of 212 semi-winter, winter and spring-type accessions. **b)** PCA analysis of only the 203 Chinese semi-winter accessions. **c)** UPGMA analysis of only the 203 Chinese semi-winter accessions.

(Table 1). The average genetic distance among lines within Q1 (0.38) was very close to that among the winter rapeseed materials (0.36), whereas subpopulation Q3 showed the same average genetic distance as the five spring rapeseed accessions (0.41). Subpopulation Q2 had the lowest average genetic distance of 0.27 (Table 1). Collectively these results suggest introgressions of winter-type oilseed rape genetic background into subpopulation Q1 and spring-type genetic background into Q3, whereas subpopulation Q2 appears to represent a relatively pure genetic background of semi-winter *B. napus*.

PCA and UPGMA tree analysis, using only the 203 Chinese semi-winter genotypes to analyse population structure, resulted in subdivision into the same three subpopulations by PCA analysis, whereby the principal component accounting for genetic diversity was smaller than with inclusion of the outliers in the full set of 212 lines (Figure 2a and b). The results of the UPGMA tree analysis corresponded with around 91% similarity to the PCA (Figure 2c).

Detailed comparisons of population structure and genetic diversity in the A- and C-subgenomes, estimated using 10,750 randomly selected, unique SNPs with MAF ≥ 0.05, are shown in Figure 3 and Table 2. In the A-subgenome, the first and second principle components explained 13.7 and 10.0% of the genetic diversity. In the C-subgenome, the genetic diversity explained by the first and second principle components was more than double that in the A-subgenome, comprising 29.0 and 21.4%, respectively. However, gene diversity and PIC were higher in the A-subgenome (0.373 and 0.298, respectively) than in the C-subgenome (0.339 and 0.276, respectively; Table 2), suggesting that a small number of accessions had particularly high allelic diversity in some C-subgenome

**Table 1 Summary statistics for genetic diversity within a subset of 154 *Brassica napus* accessions, representing the semi-winter type oilseed rape subpopulations Q1, Q2 and Q3 along with spring (5 accessions) and winter rapeseed (4 accessions)**

| Type | Subpopulation | No. of accessions | Genetic distance | Gene diversity | PIC |
|---|---|---|---|---|---|
| Semi-winter | Q1 | 86 | 0.38 | 0.34 | 0.27 |
| | Q2 | 32 | 0.27 | 0.25 | 0.20 |
| | Q3 | 27 | 0.41 | 0.34 | 0.27 |
| Spring | | 5 | 0.41 | | |
| Winter | | 4 | 0.36 | | |

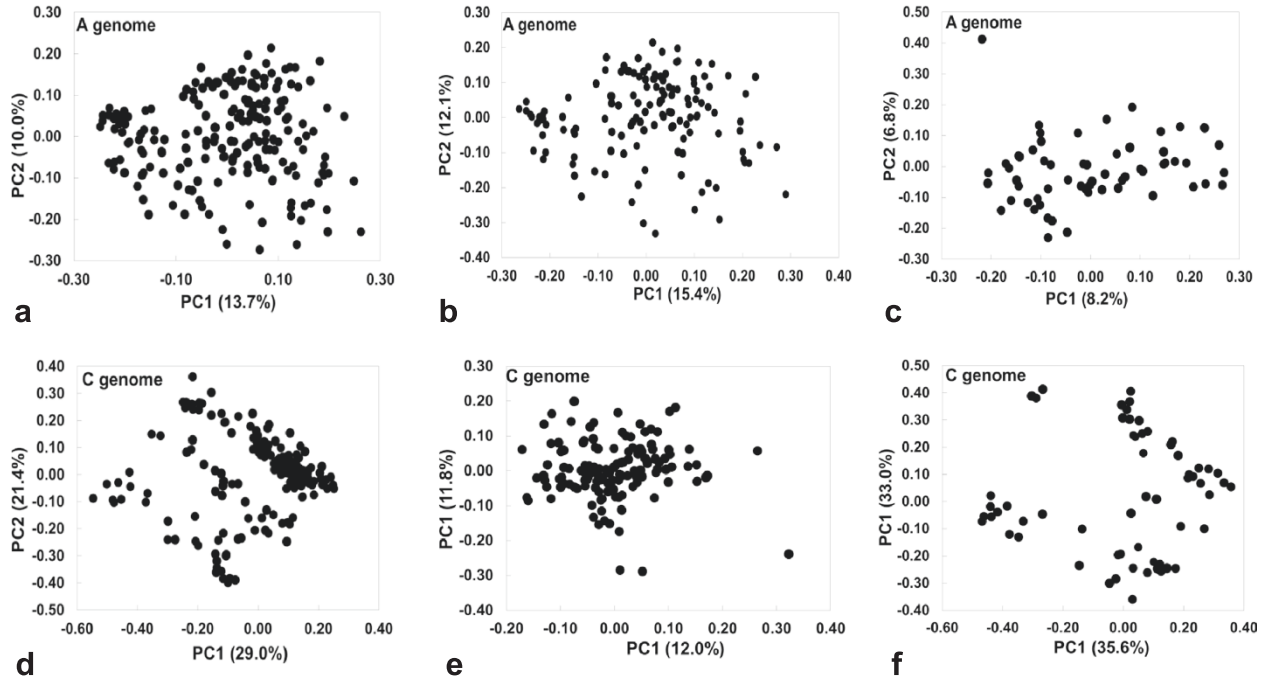



**Figure 3 Comparative principal coordinate analysis (PCA) of subgenomic genetic diversity, measured across 10,750 SNP markers per *B. napus* subgenome. a,b,c)** A-subgenome diversity; **d,e,f)** C-subgenome diversity; **a,d)** Analysis using all 203 semi-winter accessions**; b,e)** Analysis using 136 accessions with maximal A-subgenome diversity; **c,f)** Analysis using 67 accessions with maximal C-subgenome diversity.

chromosome regions. Hence, C-subgenome SNPs contributing to extreme PCA values were used to further subdivide the subpopulations based on allelic diversity.

In 135 of these accessions, the first and second principle component analysis accounted for 15.4 and 12.1% genetic diversity in the A-subgenome, compared to only 12.0 and 11.8% in the C-subgenome. Among these materials the gene diversity (0.365) and PIC (0.292) were also higher in the A-subgenome than the C-subgenome (0.277 and 0.225, respectively) (Table 2). In another group of 68 accessions, on the other hand, the first and second principle components explained only 8.2 and 6.8% when A-subgenome SNPs were used, but 35.6 and 33.0%, respectively, with C-subgenome SNPs. Accordingly, in these 67 accessions the gene diversity (0.377) and PIC (0.300) were also lower in the A-subgenome than the C-subgenome (0.392 and 0.310, respectively) (Table 2). Collectively these results suggest that the A-subgenome contributes more genetic diversity to Chinese semi-winter rapeseed than the C-subgenome, but also that a small group of materials appears to have benefited from targeted introgressions of C-subgenome diversity.

### Relative kinship

Analysis of kinship using 4000 unique SNPs each from the A- and C-subgenomes, all with MAF ≥ 0.05, supported the finding that the A-subgenome carries more overall genetic diversity than the C-subgenome (Figure 4). At the same time the kinship analysis showed only weak or no relationship among the materials. This might be attributed to the introgression of different rapeseed ecotypes and closely related species.

### Patterns of LD across the semi-winter rapeseed genome

To assess patterns of LD extent in more detail, we measured the physical distance at which the pair-wise

**Table 2 Comparative analysis of genetic diversity in the A- and C-subgenomes within subsets of Chinese semi-winter rapeseed inbred lines representing the total diversity (203 accessions), maximal A-subgenome diversity (135 accessions) and maximal C-subgenome diversity (68 accession), respectively**

| Genome | Subset of 203 accessions | | Subset of 135 accessions | | Subset of 68 accessions | |
|---|---|---|---|---|---|---|
| | Gene diversity | PIC | Gene diversity | PIC | Gene diversity | PIC |
| A | 0.373 | 0.298 | 0.365 | 0.292 | 0.377 | 0.300 |
| C | 0.339 | 0.276 | 0.277 | 0.225 | 0.392 | 0.310 |

PIC: Polymorphism information content.

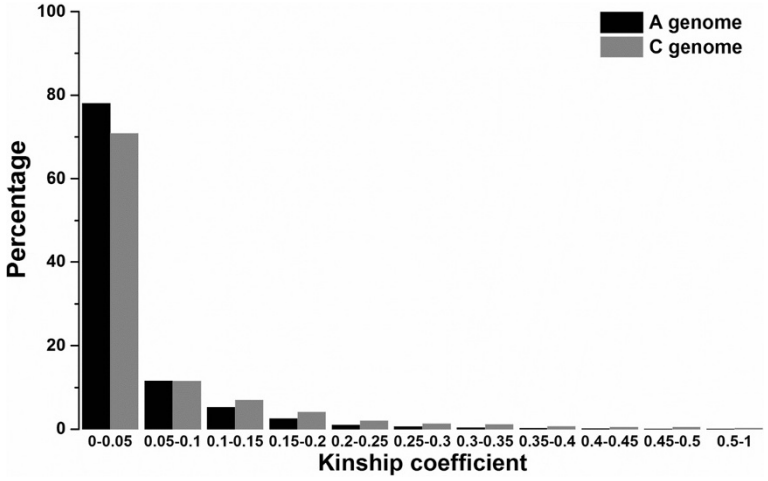



**Figure 4 Comparative analysis of kinship coefficients in the A-subgenome (black bars) and C-subgenome (grey bars) among 203 Chinese semi-winter rapeseed accessions.**

genotypic association in the filtered SNP dataset decays below a threshold of $r^2 = 0.1$. This revealed huge differences in LD decay between different chromosomes, with LD extending from 0.08-0.09 Mb (chromosome A02) up to 7.00-7.50 Mb (C01, C07 and C08) (Table 3). Figure 5 and Table 3 compare the distribution of $r^2$ with respect to the physical distance over the 19 chromosomes, as well as overall across each subgenome. Considerably faster mean LD decay was observed on A-subgenome chromosomes (0.25-0.30 Mb) than C-subgenome chromosomes (2.00-2.50 Mb; Table 3).

### Subgenome-specific haplotype block structure

The same markers used for LD estimation were employed to estimate haplotype blocks in the 203 semi-winter rapeseed accessions. In the A-subgenome, the mean MAF per chromosome varied from 0.20 (A08) to 0.31 (A09, A10), with a mean of 0.27 over all A-subgenome chromosomes. The mean $r^2$ per A-subgenome chromosome varied from 0.17 (A02, A03) to 0.36 (A09) with a mean of 0.23 over the whole A-subgenome (Table 4). On C-subgenome chromosomes mean MAF varied from 0.19 (C02) to 0.29 (C04, C05), with an average of 0.24 over the whole C-subgenome. The mean $r^2$ on C-subgenome chromosomes was considerably higher, ranging from 0.41 (C05, C07) to 0.78 (C04) with an average of 0.59 over the whole C-subgenome (Table 4). The higher MAF in the A-subgenome and stronger LD in the C-subgenome further indicate a higher genetic diversity of the A-subgenome than the C-subgenome.

A summary of the distribution, size and number of haplotype blocks per chromosome is presented in Table 4. A total of 3,097 conserved haplotype blocks were detected in the 203 Chinese semi-winter rapeseed accessions,

**Table 3 Average distance of linkage disequilibium (LD) decay ($r^2 = 0.1$) on A- and C-subgenome chromosomes, calculated using 24,994 unique, genome-wide SNP markers with minor allele frequency (MAF) ≥0.5, in a collection of 203 Chinese semi-winter *B. napus* accessions**

| Subgenome | Chromosome | LD decay (Mb) | No. of SNPs |
|---|---|---|---|
| A-subgenome | A01 | 0.11-0.12 | 1117 |
| | A02 | 0.08-0.09 | 891 |
| | A03 | 0.14-0.15 | 1646 |
| | A04 | 0.20-0.25 | 1139 |
| | A05 | 0.18-0.19 | 1249 |
| | A06 | 0.13-0.14 | 1153 |
| | A07 | 0.13-0.14 | 1412 |
| | A08 | 1.50-2.00 | 820 |
| | A09 | 1.00-1.50 | 1166 |
| | A10 | 0.45-0.50 | 1171 |
| | Mean | 0.25-0.30 | 1176 |
| C-subgenome | C01 | 7.00-7.50 | 2041 |
| | C02 | 5.00-5.50 | 1891 |
| | C03 | 0.60-0.65 | 2094 |
| | C04 | 3.50-4.00 | 2473 |
| | C05 | 0.40-0.45 | 718 |
| | C06 | 0.80-0.85 | 905 |
| | C07 | 7.00-7.50 | 1285 |
| | C08 | 7.00-7.50 | 1156 |
| | C09 | 1.00-1.50 | 667 |
| | Mean | 2.00-2.50 | 1581 |
| A + C | Mean | 0.85-0.90 | 1378 |

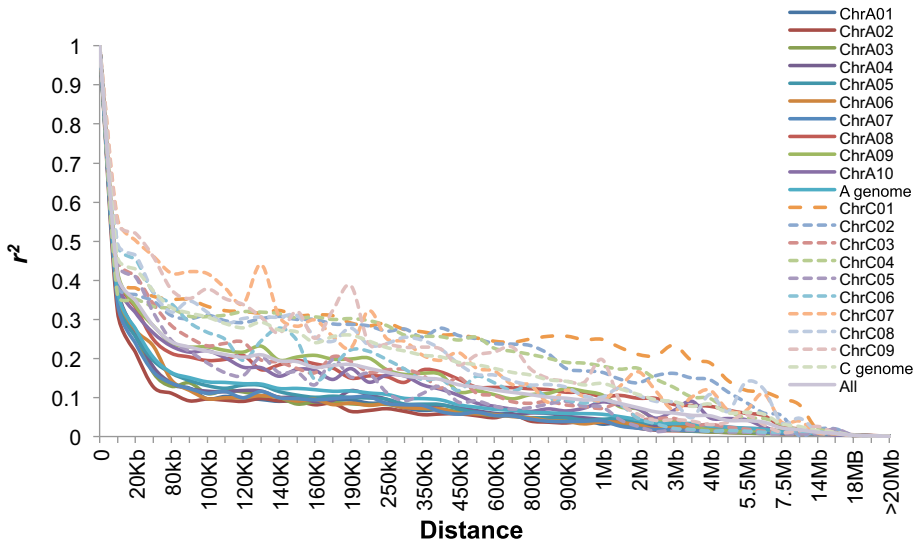



**Figure 5 Patterns of linkage disequilibrium (LD, $r^2 = 0.1$) across the 19 haploid chromosomes of semi-winter type *B. napus*, measured with 24,994 single-copy SNP markers.** The solid lines represent LD decay in A-subgenome chromosomes, while the dashed lines represent LD decay in C-subgenome chromosomes.

**Table 4 Chromosome-specific haplotype block structure analysed using in a collection of 203 Chinese semi-winter *B. napus* accessions**

| Chromosome | No.of SNPs | Chromosome length (Mb) | Mean MAF | Mean $r^2$ | Number of blocks | Mean block size (Kb) | Block coverage area per chromosome (Mb) | Block coverage percentage per chromosome (%) |
|---|---|---|---|---|---|---|---|---|
| A01 | 1117 | 23 | 0.27 | 0.19 | 190 | 27.74 | 5.27 | 0.23 |
| A02 | 891 | 25 | 0.25 | 0.17 | 145 | 23.30 | 3.38 | 0.14 |
| A03 | 1646 | 29 | 0.27 | 0.17 | 286 | 21.84 | 6.25 | 0.22 |
| A04 | 1139 | 20 | 0.24 | 0.19 | 175 | 26.46 | 4.63 | 0.23 |
| A05 | 1249 | 23 | 0.27 | 0.21 | 215 | 33.07 | 7.11 | 0.31 |
| A06 | 1153 | 24 | 0.28 | 0.20 | 213 | 32.07 | 6.83 | 0.28 |
| A07 | 1412 | 24 | 0.28 | 0.19 | 247 | 24.57 | 6.07 | 0.25 |
| A08 | 820 | 19 | 0.20 | 0.33 | 129 | 51.72 | 6.67 | 0.35 |
| A09 | 1166 | 33 | 0.31 | 0.36 | 157 | 60.57 | 9.51 | 0.29 |
| A10 | 1171 | 17 | 0.31 | 0.30 | 167 | 33.78 | 5.64 | 0.33 |
| A-subgenome mean | 1176 | 24 | 0.27 | 0.23 | 192 | 33.51 | 6.14 | 0.26 |
| C01 | 2041 | 39 | 0.20 | 0.73 | 135 | 119.92 | 16.19 | 0.42 |
| C02 | 1891 | 46 | 0.19 | 0.84 | 124 | 186.10 | 23.07 | 0.50 |
| C03 | 2094 | 60 | 0.23 | 0.46 | 209 | 89.810 | 18.77 | 0.31 |
| C04 | 2473 | 49 | 0.29 | 0.78 | 172 | 100.58 | 17.30 | 0.35 |
| C05 | 718 | 43 | 0.29 | 0.41 | 92 | 50.64 | 4.66 | 0.11 |
| C06 | 905 | 37 | 0.27 | 0.46 | 113 | 82.04 | 9.27 | 0.25 |
| C07 | 1285 | 45 | 0.25 | 0.41 | 138 | 95.46 | 13.17 | 0.29 |
| C08 | 1156 | 38 | 0.21 | 0.61 | 124 | 93.45 | 11.59 | 0.31 |
| C09 | 667 | 48 | 0.23 | 0.60 | 66 | 107.67 | 7.11 | 0.15 |
| C-subgenome mean | 1581 | 33 | 0.24 | 0.59 | 130 | 102.85 | 13.46 | 0.41 |
| Whole genome mean | 1378 | 29 | 0.25 | 0.41 | 161 | 68.18 | 9.80 | 0.34 |

MAF: Minor allele frequency.

spanning 182.49 Mb (15.17% of the assembled reference genome). In the A-subgenome chromosomes, mean haplotype block number ranged from 129 (A08) to 286 (A03) with an average of 192, while the mean haplotype size ranged from 21.84 (A03) to 60.57 Kb (A09) with an average of 33.51 Kb. The mean haplotype block number in C-subgenome chromosomes varied from 66 (C09) to 209 Kb (C03) with an average of 130 Kb, while mean haplotype size was considerably larger, ranging from 50.64 (C05) to 186.10 Kb (C02) with an average of 102.85 Kb (Table 4; Figure 6a and b). In the A-subgenome 53.85% and 30.72% of haplotype blocks ranged in size from 0–10 Kb and 10–50 Kb, respectively, whereas only 31.20% of C-subgenome haplotype blocks were in the 0–10 Kb size range and only 26.34% in the 10–50 Kb size range (Figure 6c). In contrast, much fewer regions with long-range haplotype conservation were observed in the A-subgenome, where haplotype blocks ranging in size from 50–100 Kb, 100–200 Kb and 200–500 Kb were present at frequencies of only 7.69, 4.89 and 3.27%, respectively. In the C-subgenome, on the one hand, the respective haplotype block sizes were observed at much higher frequencies of 13.04, 11.60 and 18.58%, respectively (Figure 6c), demonstrating that the higher mean haplotype block size in the C-subgenome is caused by retention of long-range LD.

Particularly high conservation of LD was observed on chromosomes C01, C02, C04 and C09, which showed very high average $r^2$ values of 0.73, 0.84, 0.78 and 0.60, respectively, and correspondingly large mean haplotype block sizes of 119.92, 186.10, 100.58 and 107.67 Kb, respectively (Table 4).

### Analysis of homoeologous QTL regions

Comparative sequence analysis revealed conserved haplotype blocks and LD corresponding to homoeologous QTL for seed glucosinolate content on homoeologous chromosomes A02/C02 and A09/C09, respectively, and for erucic acid content on chromosomes A08 and C03. The genomic positions of known QTL for seed glucosinolate content on chromosome A02/C02 (mapped in detail by [26]) were localised by a BioEdit local BLAST search [41] using 17 and 5 SNPs, respectively, within the QTL confidence intervals. These SNPs mapped to overlapping homoeologous regions from 19,680,403 – 23,996,416 bp on chromosome A02 and from 41,859,157 bp – 44,499,708 bp on chromosome C02, respectively (Additional file 3). Similarly, 17 SNPs spanning another major QTL for seed glucosinolate content on chromosome A09 [26] were localised to overlapping homoeologous regions from 775,293 – 3,831,394 bp and 290,810 – 5,109,219 bp on chromosomes A09 and C09, respectively (Additional file 3). Insufficient SNPs were

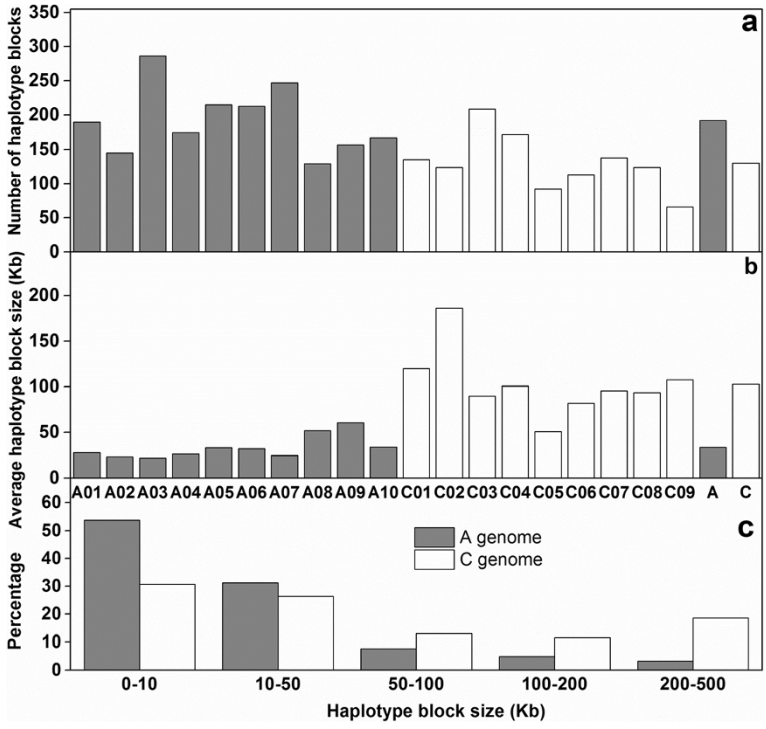



**Figure 6 Comparative analysis of haplotype block structure in the A-subgenome (grey bars) and C-subgenome (black bars) of semiwinter-type *Brassica napus*. a)** Comparison of the numbers of LD blocks on A- and C-subgenome chromosomes. **b)** Comparison of the average size of LD blocks on A- and C-subgenome chromosomes. **c)** Comparison of the size range distributions of haplotype blocks in the A- and C-subgenomes.

present in the QTL on chromosome C09 from [26], hence the physical region on C09 was predicted by a BLAST search of SNPs from the homoeologous QTL region on A09.

Homoeologous QTL for erucic acid content on chromosomes A08 and C03 [26] were physically localised using 17 and 5 SNPs, respectively, spanning these two loci. The corresponding QTL covered the regions from 9,513,648 – 12,196,483 bp and 54,259,136 – 57,154,658 bp on chromosomes A08 and C03, respectively (Additional file 3). As expected, these regions include the two *B. napus* homologues of the gene *FATTY ACID ELONGASE 1* (*Bna.FAE1*) that carry the agronomically essential low erucic acid mutations [44]. Results from matching of the physical positions were compared to BLAST alignments of sequences against each other with similar results (Additional file 3; Additional file 4; Figure 7).

### Comparative analysis of haplotype blocks within homoeologous QTL for key traits

Estimates of *D'* based on marker pairs lying within homoeologous QTL for seed glucosinolate (GLS, chromosomes A02/C02 and A09/C09) and erucic acid content (A08/C03) revealed large differences in recombination structure and extent of LD between the respective homoeologous chromosome regions (Figure 8). Two QTL for GLS described by Delourme *et al.* [26] were physically mapped to 4.32 and 3.06 Mb regions of chromosomes A02 and A09, whereas the corresponding homoeologous regions on chromosome C02 and C09 covered 7.39 and 4.83 Mb, respectively. On A02 and A09 the mean $r^2$ (0.23 and 0.12) and mean haplotype block sizes (20.67 and 12.46Kb) are both considerably smaller than in the homoeologous regions on C02 and C09 (0.45 and 0.21, 120.35 and 26.27 Kb, respectively) (Table 5; Figure 8). A similar observation was made for the eurcic acid content QTL region described by Delourme *et al.* [26], which mapped to a physical region covering 2.68 Mb of chromosome A08 and a homoeologous region of 4.89 Mb on chromosome C03. On A08 the erucic acid QTL region shows considerably lower LD (mean $r^2 = 0.35$) and mean haplotype block size (56.17 Kb) than the homoeologous QTL region on C03 (mean $r^2 = 0.45$, mean haplotype block size 181.29 Kb) (Table 5; Figure 8).

We further analyzed these homoeologous QTL regions for genes related to seed GLS and erucic acid content, respectively. Three GSL biosynthetic process genes and one GSL catabolic process gene were located in a 0.50 Mb region with disrupted LD that spanned the QTL for GSL on A02 (Additional file 5; Figure 8). In contrast the homoeologous QTL region on chromosome C02 contained seven GSL-related genes (including the *myb* transcription factor) that spanned a 5.7 Mb region with extensive LD. This region included four glucosinolate biosynthetic process genes, within three conserved haplotype blocks ranging in size from 38 to 326 Kb (Additional file 5; Figure 8). Similar patterns of LD conservation were observed between the homoeologous QTL for GSL on chromosomes A09 and C09. On A09, four GSL biosynthetic process genes (including a *myb34*) were located within a 1.12 Mb region with low LD. In contrast, seven genes implicated in GSL biosynthesis

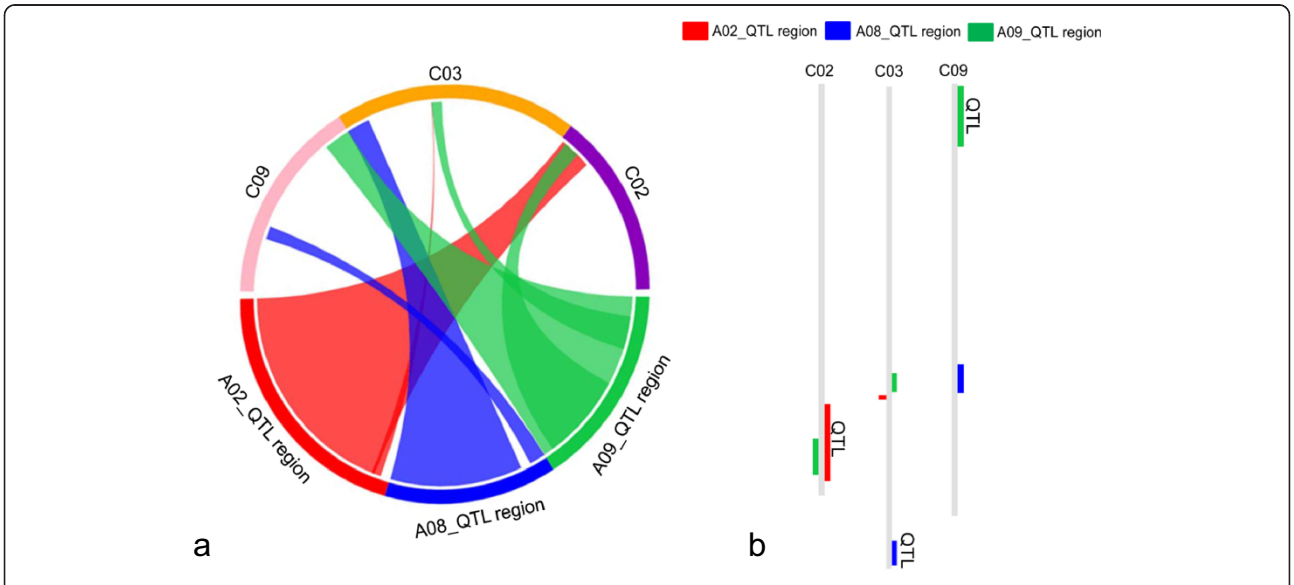



**Figure 7 Synteny alignments of QTL for seed glucosinolate (chromosomes A02 and A09) and erucic acid (A08) into homoeologous regions of chromosomes C02, C09 and C03, respectively. a)** Circular map and **b)** linear map, respectively, showing showing A-subgenome chromosomal positions corresponding to C-subgenome QTL.

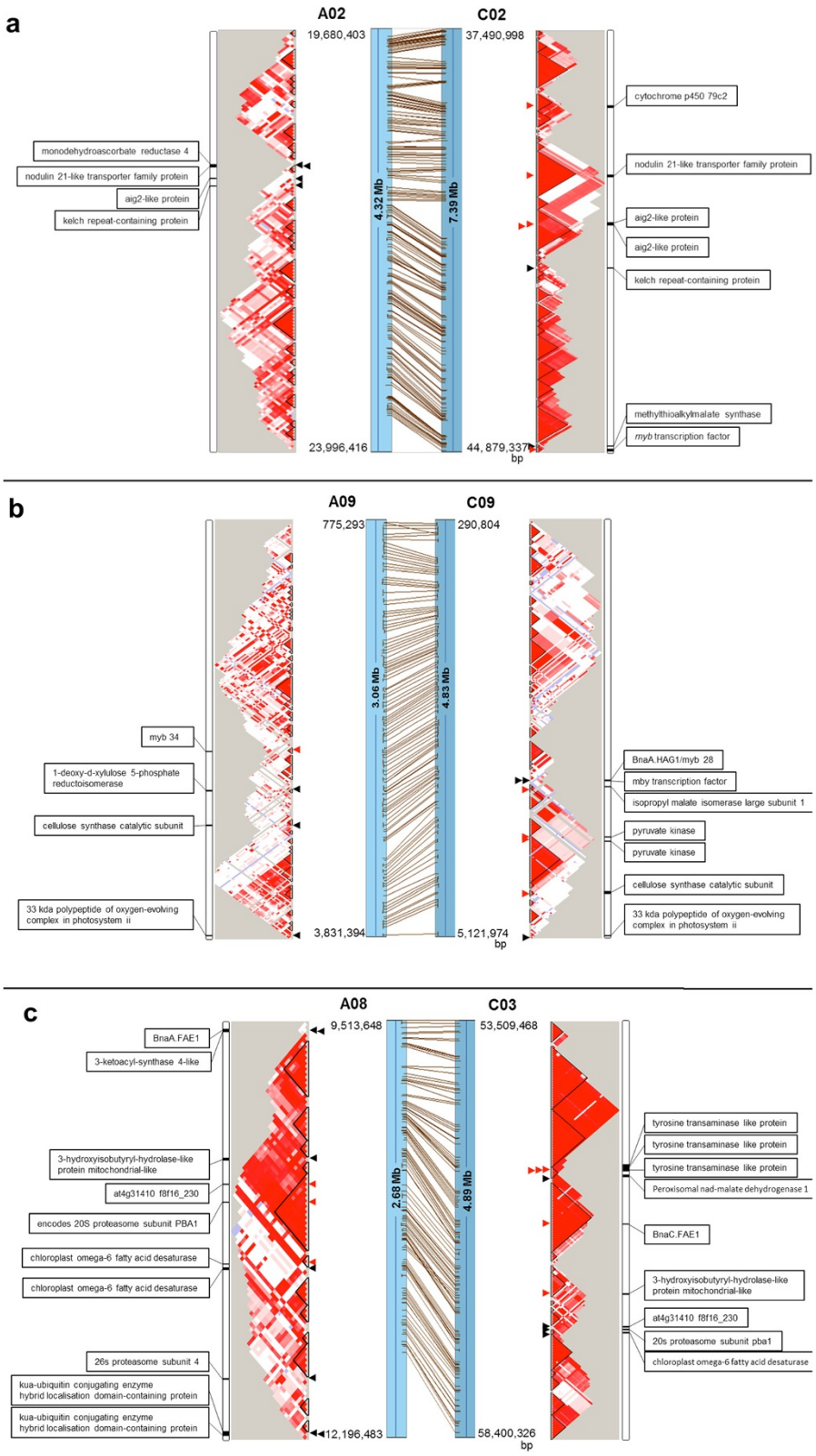



**Figure 8** (See legend on next page.)

(See figure on previous page.)
**Figure 8 Comparative sequence analysis showing differences in linkage disequilibrium (LD) and putative functional candidate gene content within haplotype blocks between homoeologous A-subgenome and C-subgenome QTL for (a,b) seed glucosinolate (GLS) and (c) erucic acid content on chromosomes (a) A02/C02, (b) A09/C09 and (c) A08/C03, respectively.** Regions with strong (D' = 1), high confidence LD (LOD > 2) are plotted bright red, lighter shades of red represent moderate LD (D' < 1) with high confidence (LOD > 2), while light blue blocks represent weak, low confidence LD (D' = 1, LOD < 2) and white blocks an absence of LD (D' < 1, LOD < 2). Connecting lines show syntenic sequence alignments between the homoeologous QTL regions. Arrows show putative function candidate genes annotated to **(a,b)** glucosinolate biosynthesis or catalysis or **(c)** fatty acid or oil biosynthesis. Red arrows indicate genes under strong selection within conserved LD/haplotype blocks, whereas black arrows indicate that the corresponding gene at the homoeologous locus is not within a regions under strong selection. Gene names are given opposite the arrow positions on the homoeologue on which they are present.

(including a *myb* transcription factor and the important GSL gene *BnaA.HAG1*/*myb28* [67]) were located in a 1.98 Mb region, including four GSL biosynthesis genes, within three conserved haplotype blocks ranging in size from 21 to 133 Kb (Additional file 5; Figure 8).

On chromosome A08, the major QTL for erucic acid content was found to contain five additional fatty acid biosynthetic process genes besides the causal gene *BnaA.FAE1*. These included a *3-ketoacyl-synthase 4-like* gene and two *chloroplast omega-6 fatty acid desaturase* orthologues, along with four fatty acid beta-oxidation genes. Within a total length of 1.93 Mb we found that two fatty acid beta-oxidation genes and one fatty acid biosynthetic process gene (*chloroplast omega-6 fatty acid desaturase*) were located in separate haplotype blocks, ranging in size from 13 to 137 Kb (Additional file 5; Figure 8). Within the corresponding homoeologous QTL region on chromosome C03 we localised six fatty acid biosynthetic process genes (including *BnaC.FAE1* and a *chloroplast omega-6 fatty acid desaturase*) and three fatty acid beta-oxidation genes, covering a total of 2.54 Mb. Four of the fatty acid biosynthesis genes (including *BnaC.FAE1*) and one fatty acid beta-oxidation gene were located in strongly conserved haplotype blocks ranging in size from 281 to 477 Kb (Additional file 5; Figure 8).

The different intensities of selection at A- and C-subgenome QTL for seed erucic acid and GSL content were confirmed by reanalyzing the extent of LD conservation based on $r^2$ with LDheatmap (Additional file 6). Again we found strong LD conservation and similar gene content within the C-subgenome QTL, whereas A-subgenome QTL showed less conserved LD and more fragmented haplotype structure (Additional file 5, Additional file 6). The results suggest considerably stronger retention of C-subgenome haplotype blocks than A-subgenome haplotypes within these important seed quality QTL.

## Discussion

### Population structure and genetic diversity

Rapeseed breeding materials in Australia and China have similar origins, with introductions from Europe, Canada and Japan in the mid-20th century and subsequent interchange of germplasm since that time [45]. Recently, hybrid breeding has received considerable attention, with development of genetically diverse gene pools through recurrent, reciprocal selection of genetic diversity from different *B. napus* ecotypes [9]. The three main population subgroups we observed in our diversity panel may reflect breeding efforts to diversify Chinese semi-winter rapeseed by introgressing genetically distant winter rapeseed (in the case of Q1) and spring canola (in the case of Q3) into different hybrid breeding pools.

Genetic diversity in Chinese *B. napus* has been further improved by introgressions from Asian *B. rapa* [45,46], the diploid donor of the *B. napus* A-subgenome. According to Liu [47] and Shiga [48], more than 50% of *B. napus* cultivars in China and Japan are derived from

**Table 5 Comparative sequence analysis among haplotype blocks showing conserved linkage disequilibrium (LD) covering homoeologous quantitative trait loci (QTL) for seed glucosinolate (GSL; chromosomes A02/C02 and A09/C09, respectively) and erucic acid content (chromosomes A08 and C03)**

| Chromosome | GLS QTL region (bp) | Erucic acid QTL region (bp) | Region size (Mb) | No. of blocks | Mean block size (Kb) | LD block coverage (Mb) | Mean $r^2$ |
|---|---|---|---|---|---|---|---|
| A02 | 19,680,403-23,996,416 | | 4.32 | 24 | 20.67 | 0.49 | 0.23 |
| C02 | 37,490,998-44, 879,337 | | 7.39 | 31 | 120.35 | 3.73 | 0.45 |
| A09 | 775,293-3,831,394 | | 3.06 | 37 | 12.46 | 0.46 | 0.12 |
| C09 | 290,804-5,121,974 | | 4.83 | 26 | 26.27 | 0.68 | 0.21 |
| A08 | | 9,513,648-12,196,483 | 2.68 | 12 | 56.17 | 0.67 | 0.35 |
| C03 | | 53,509,468-58,400,326 | 4.89 | 17 | 181.29 | 3.08 | 0.45 |

*B. napus* × *B. rapa* crossings. Correspondingly, we observed considerably more genetic diversity in the A-subgenome of Chinese rapeseed than the C-subgenome. interestingly, however, we found 67 accessions with a stronger diversity in the C-subgenome than the A-subgenome. These may derive from programs to introgress additional diversity and resistance alleles from C-genome donors.

### LD and haplotype block analysis

Around 15.17% of the assembled *B. napus* genome could be assigned to haplotype blocks, with large gaps between blocks (data not shown). With an average SNP density of only one SNP per 48.01 Kb, it is difficult to detect very small haplotype blocks [49]. Recent studies in maize [25] and Arabidopsis, [50] have demonstrated the power of comprehensive genome-wide SNP genotyping arrays for generation of detailed haplotype maps and high-resolution LD analysis. Whole-genome resequencing data provides the ultimate dimension to uncover LD in association with signatures of natural and artificial selection, but so far has been limited to species with relative small, diploid genomes, like sorghum [51]. Many of the problems with duplicated SNP loci on the Brassica 60 k SNP array result from the extensive recent genome duplications which make it quite challenging to design locus-specific SNP assays in many strongly homoeologous regions of the genome. These technical difficulties can cause a reduction in resolution in some genome regions. Nevertheless, use of a high-density SNP array to analyse and compare LD and selection in homoeologous QTL is a unique feature of this study in comparison to previous work in simple diploid species. As a recent allopolyploid *B. napus* thus provides interesting insight into the evolutionary processes of selection in an important crop [30].

The 24,994 unique, polymorphic SNPs we used in our analyses were nevertheless sufficient to perform a preliminary whole-genome analysis of haplotype block structure in *B. napus*. In particular we were able to demonstrate that some *B. napus* chromosomes carry extremely large segments of highly conserved LD, and that this phenomenon is a particular feature of C-subgenome chromosomes. This may indicate increased recombination rates of A-subgenome chromosomes after interspecific hybridisations with *B. rapa*. Boosts of homologous recombination among diploid chromosome pairs after interspecific hybridisation were documented in *Brassica* crosses by Leflon *et al.* [52]; this might have caused more rapid LD decay and subsequently shorter-range haplotype blocks in A-subgenome chromosomes in the present materials after hybridisations with *B. rapa*. Although breeders have used interspecific crosses to improve agronomic traits and increase C-subgenome genetic diversity in *B. napus*, it is extremely difficult to obtain viable hybrid seeds from *B. napus* × *B. oleracea* crosses [53,54], causing a constraint in the ability to diversify the C-subgenome genetic component. It is thought that *B. napus* arose only in post-neolithic times and from only a small number of independent hybridisation events [27], and that the Chinese rapeseed genepool may predominantly represent only one or a few of these events. Hence it is perhaps not surprising that recombination and diversity appear to be considerably lower in the C-subgenome of Chinese oilseed rape. An alternative explanation, which may also partly explain the great overall difference in LD between the A and C subgenomes, is the considerably greater expansion of transposable elements in the *B. napus* C-subgenome compared to the A-subgenome [30], since transposon-rich regions are often observed to be recombination-poor [55]. On the other hand, this fails to explain the great variation in the size of LD and long-range haplotype blocks we observed among different C-subgenome chromosomes. A more simple contributing factor is likely to be strong natural and artificial selection for key adaptation and seed quality traits, where specific variants seem to have been selected during the face of ecogeographical adaptation and human selection, for example for flowering time or quality traits. Strong selection at a locus is expected to reduce diversity and increase LD and haplotype block size in the surrounding region [56].

In particular, stronger LD and longer-range LD blocks on chromosomes C01, C02, C04 and C09 suggest particularly strong selection the corresponding region of these chromosomes. According to Liu [47] rapeseed was introduced into China from Europe in the 1930-1940s, although a later origin within the past few hundred years in Japan is also postulated [57,58]. Guryev *et al.* [59] showed that the evolutionary selection process drives conservation of long-range allele combinations, causing chromosome regions to retain a long-range haplotype block structure. Artificial selection can also have a profound effect on LD in crop plants, with selection for key agronomic traits like flowering behavior, resistances or essential quality parameters causing genetic bottlenecks that lead to extensive conserved haplotype blocks in chromosome regions carrying the responsible gene loci or major QTL for selected traits. Modern double-low quality oilseed rape has undergone selective sweeps for reduction of seed erucic acid and glucosinolate contents, along with flowering time, winter hardiness and vernalisation-related traits. Such selection tends to reduce allele diversity and increase haplotype block structure around the major responsible loci, however detailed studies of LD conservation in oilseed rape breeding pools on a DNA sequence level has only recently become possible since the availability of high-density genome-wide SNP markers [7] in combination with annotated *Brassica* genome sequences. Here we identified chromosome-scale LD

patterns in *B. napus* genome regions carrying important QTL for both a simple, bigenically inherited trait (erucic acid content) and for a complex quantitative trait (glucosinolate content).

The observed distance of LD decay in Chinese oilseed rape was 0.85-0.90 Mb, which is higher than maize with 0.5-10 kb [17] and Arabidopsis with 50–250 Kb [21]. This reflects the very recent domestication of *B. napus*, its exclusive use in cultivation, with no known wild forms, and the strong selection bottlenecks associated with cultivation and breeding. Previous studies (e.g. Wang et al. [60]) have shown that the A-subgenome has been successfully improved by closely related species, leading to more rapid decay of LD in the A-subgenome than the C-subgenome. Our results showed that the relatively low overall LD conservation in Chinese rapeseed is caused mainly by a lack of genetic diversity in the C-subgenome. According to Mei *et al*. [61], natural *B. napus* has very low genetic diversity compared with its diploid progenitors, therefore intercrossing with the parental species can be an effective way to broaden genetic diversity in rapeseed. To achieve this it may be necessary to overcome sexual compatibility barriers by using embryo rescue techniques, for example. In recent years considerable progress has been made in introducing novel C-genome donors to European winter oilseed rape, in order to improve genetic diversity particularly for disease resistance [62-64] or heterosis [65].

### Haplotype block and extent of LD of homologues QTL region

Conserved haplotype blocks with strong LD spanning major homoeologous QTL for seed GLS (chromosomes A02/C02 and A09/C09) and erucic acid (A08/C03) reflect the strong selection bottlenecks for these traits. On the other hand, the introgression of exotic A-subgenome diversity from *B. rapa* has apparently led to shorter-range haplotype blocks and lower LD in A-subgenome than C-subgenome QTL. Chinese *B. napus* originated from Europe [47], being introduced to China in the 1930-1940s and replacing the traditional oilseed crop *B. rapa*. Local adaptation to the new ecogeographical environment, and diversification of breeding pools, was achieved by introgressing local populations of the wild progenitors and closely related species, particularly *B. rapa* [66-68]. Our results show that this process resulted in substantial decay of LD surrounding important A-subgenome QTL, whereas longer-range haplotype blocks and higher LD are retained around C-subgenome QTL regions. Importantly, conserved haplotype blocks in C-subgenome QTL tend to retain multiple genes related to relevant biosynthetic processes, which can potentially cause linkage drag that slows breeding progress for the trait of interest.

Various forces have potentially contributed to haplotype conservation in C-subgenome QTL in *B. napus*, including genetic bottlenecks from artificial or natural selection or a simple lack of recombination and sequence diversity. We found rates of sequence polymorphism to be generally lower in C-subgeneome QTL regions than their corresponding A-subgenome homoeologues, suggesting that the former may be the dominant mechanism. On the other hand a suppression of recombination, due to the increased density of transposable elements in the C-subgenome [30], cannot be ruled out. Detailed haplotype block analysis of important QTL can help in the precise mapping of important genomic regions and location of favorable alleles. In association with genomic sequence data it can also help to more precisely predict quantitative trait-related genes (QTG) in QTL regions using targeted association mapping with high-density markers.

The strongly conserved LD we observed across the QTL on chromosomes C02 and C09 was found to be associated with a large number of functionally related genes in close genetic linkage. The corresponding homoeologous QTL on chromosomes A02 and A09 each contained fewer genes annotated to the QTL function. This result demonstrates the important role of gene loss during or after allopolyploidisation in natural and/or artificial selection of key traits like GSL content [30]. Natural evolution results in a positive and balancing selection within the genome, whereas artificial secletion can lead to partial separation of phenotypic traits. According to Harper *et al*. [69], deletions affecting homologues of the GSL biosynthesis gene *Bna.HAG1*/*myb28* resulted in selective sweeps affecting the QTL for GSL on A09 and C02. Both in this case, and in the case of homologous QTL erucic acid content on chromosomes A08 and C03, we demonstrate that selective sweeps can also incorporate additional, functionally-related genes for which alleles in strong LD may have either a positive or negative influence (linkage drag) on target traits (e.g. GSL content, fatty acid composition or oil content). Detailed analysis of LD structure and signatures of selection in important QTL can guide breeders towards a knowledge-based crop improvement by genome-based introgression of useful genetic diversity.

## Conclusions

Using densely-spaced genome-wide SNPs to analyse subgenomic genetic diversity in semi-winter *B. napus*, we found stronger LD and long-range haplotype conservation in C-subgenome chromosomes. Comparative sequence analysis revealed conserved blocks containing homoloeogous QTL for important seed quality traits under intense artificial selection. The results indicate strong selection for large chromosome regions associated with important seed quality traits conferred by C-subgenome QTL, suggesting that an increase in

genetic diversity and recombination within the C-genome is particularly important for breeding. The resolution of genome-wide association studies is also expected to vary greatly across different genome regions.

## Methods

### Plant germplasm and genotyping

A set of 203 homozygous *B. napus* inbred lines was collected to construct a diversity panel broadly representing variability in Chinese semi-winter rapeseed. The materials (Additional file 2) were obtained as self-pollinated seeds from Southwest University, Chongqing, China, where they represent part of a breeding program spanning genetic diversity from the broader Asian gene pool. In addition, five spring-type and four winter-type *B. napus* inbred lines were included as outliers to assess the extent and impact of introgressions from extant gene pools into the Asian semi-winter materials.

DNA was extracted by a modified CTAB procedure according to Murray and Thompson [29]. The Brassica SNP consortium 60 k Infinium genotyping array (Illumina Inc., San Diego, CA, USA) was used to obtain high-density genome-wide data from each accession, according to the manufacturer's protocol. DNA samples were analysed by a commercial genotyping service company (TraitGenetics, Gatersleben, Germany) and SNP calling was performed using a proprietary cluster file generated by the International Brassica SNP consortium which designed the array (Isobel Parkin, AAFC, Saskatoon, SK, Canada, personal communication). A pre-publication draft assembly of the *B. napus* 'Darmor-Bzh' reference genome assembly [30] was kindly provided by Boulos Chalhoub (INRA-UNRV, Évry, France) for assignment of physical genome positions of the SNPs.

### Genetic diversity and population structure analysis

Analyses of gene diversity, polymorphic information content (PIC) and genetic distance [31] were performed using the software PowerMarker version 3.25 [32]. The population structure among the 212 accessions in the panel was assessed using the model-based Bayesian clustering method implemented in STRUCTURE version 2.3.3 [33]. The number of subgroups ($K$) was set from 1 to 10. For each $K$, seven runs were performed separately with burn-in length and iterations set to 10000 and 50000, respectively. Lines with membership probabilities ≥0.6 were assigned to the corresponding subgroups and lines with membership probabilities <0.6 were assigned to a "mixed" subgroup.

The software Powermarker version 3.25 [32] was employed to calculate genetic distance among accessions according to Nei [31]. The double-centred genetic matrices thus created were used to obtain eigenvectors by implementing the modules DCENTER and EIGEN in the software NTSYSpc 2.1 [34]. In combination with the population structure result from STRUCTURE, the first and second principle components from these data were used to prepare 2D plots using Microsoft Office Excel 2010. An unweighted pair group matrix algorithm (UPGMA) tree was calculated by Powermarker version 3.25 and drawn using the software FigTree version 1.3.1 [35].

### Calculation of genome-specific relative kinship

Using the software package SPAGeDi [36], a selection of 8,000 SNPs, from the 24,994 with MAF ≥0.5, was used to calculate the relative kinship within the A- and C-subgenomes. For this purpose, 4,000 SNPs were randomly selected from *B. napus* A-subgenome chromosomes and 4000 from C-subgenome chromosomes. Negative values between two individuals, indicating that there was less relationship than that expected between two random individuals, were corrected to 0 as suggested by Yu *et al.* [37].

### Analysis of linkage disequilibrium

To investigate chromosome-wide and genome-specific patterns of linkage disequilibrium, the software package TASSEL 4.0 [38] was used to estimate LD ($r^2$) on each chromosome and across the A- and C-subgenomes, respectively, using the 24,994 unique SNPs with MAF >0.5 and set a cut-off value of $r^2 = 0.1$ to compare the extent of LD decay. We combined marker pairs into distance intervals, rather than considering them individually, to reduce the influence of outliers and to obtain a better visual description of the LD decay with distance. The genetic intervals of 44 regions were used in this study. As described by Yan *et al.* [17], the $r^2$ value for a marker distance of 0 Kb was assumed to be 1.

### Haplotype block structure

HAPLOVIEW v4.2 [39] was used to estimate haplotype block structure in the 203 Chinese semi-winter rapeseed accessions across the 24,994 unique SNPs. The method followed for block definition was previously described by Gabriel *et al.* [40], who defined 'strong LD' if the one-sided upper 95% confidence bound of D' is higher than 0.98 and if the lower bound is above 0.70.

### Haplotype block structure of homoeologous QTL regions

The genomic positions of known QTL for seed glucosinolate content on chromosome A02/C02 and A09/C09, and for erucic acid content on chromosomes A08/C03 (mapped in detail by Delourme *et al.* [26]) were localised by a BioEdit local BLAST search [41] using SNPs spanning the QTL confidence intervals.

A chromosome-scale alignment of the selected seed quality QTL regions was subsequently performed using the large-scale genome synteny tool SyMAP version 4.2 [42] (Additional file 4; Figure 7).

The genomic sequences of the overlapping homoeologous QTL regions for the two seed quality traits were used to search in the *A. thaliana* database (http://www.arabidopsis.org/Blast/) for genes annotated to seed glucosinolate, fatty acid or oil biosynthesis (Additional file 5). The haplotype block structure within the homoeologous QTL regions was studied in detail using HAPLOVIEW v4.2 to describe local LD around trait-relevant genes within in these regions. A heatmap comparing the LD structure across the QTL regions in the respective homoeologous A- and C-subgenome chromosomes were drawn using the R package LDheatmap [43].

## Additional files

**Additional file 1: Genotype matrix, flanking sequences and genomic positions (best BLAST hit against the Darmor-*Bzh B. napus* V4.1 reference genome) for the subset of 24,994 high-quality, polymorphic, single-locus *Brassica napus* SNP markers with MAF ≥0.05, as used for the LD analyses.** SNP allele calls were generated in 203 *B. napus* breeding lines using the Illumina 60kSNP Infinium Brassica Consortium Array (Illumina Inc., San Diego, USA.

**Additional file 2: Proportional memberships in subpopulations as defined by Structure.**

**Additional file 3: Integrated map showing genomic positions of SSR and SNP marker sequences from QTL for seed glucosinolate (GLS) and erucic acid content, identified by BLAST searches onto A- and C-subgenome chromosomes from the *Brassica napus* Darmor-Bzh reference genome.**

**Additional file 4: Details of synteny alignments for QTL positions for seed glucosinolate (chromosomes A02 and A09) and erucic acid (A08) into homoeologous regions of chromosomes C02, C09 and C03, respectively.**

**Additional file 5: Detailed information on putative functional candidate genes and LD (haplotype block) analysis within the investigated QTL intervals for seed glucosinolate (GLS) and erucic acid content.**

**Additional file 6: Comparative analysis of the extent of LD across homologous QTL for a,b) seed glucosinolate content (GLS) on homoeologous chromosomes a) A02/C02 and b) A09/C09, and c) erucic acid content on homoeologous chromosomes A08/C03.** The colored plots represent the pairwise LD across the respective homoeologous QTL regions, while the framed triangles represent regions with strongly conserved LD (LD blocks). The red and black small solid triangles represent positions of putative functional candidate genes, corresponding to Figure 8.

### Competing interests
The authors declare that they have no competing interests.

### Authors' contributions
WQ generated the plant population and performed field phenotyping. LQ performed field phenotyping and was responsible for the data analysis and interpretation. LQ and RS conceived the study, generated the genome-wide SNP data and drafted the manuscript. All authors read and approved the final manuscript.

### Acknowledgements
The authors thank Benjamin Wittkop for coordinating the field trials in Germany and the seed quality analysis, and Petra Degen, Malte Luh, Mechthild Schwarte and Stavros Tzigos for technical assistance. Funding was provided by DFG grant 14/16-1 to RS and a doctoral fellowship to LQ from the Chinese Scholarship Council.

### Author details
[1]Department of Plant Breeding, IFZ Research Centre for Biosystems, Land Use and Nutrition, Justus Liebig University, Heinrich-Buff-Ring 26-32, 35392 Giessen, Germany. [2]College of Agronomy and Biotechnology, Southwest University, 400716 Chongqing, China.

Received: 15 August 2014 Accepted: 18 December 2014
Published: 23 December 2014

### References
1. Uzunova M, Ecke W, Weißleder K, Röbbelen G: **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* 1935, **90:**194–204.
2. Flint-Garcia SA, Thornsberry JM, Buckler ES: **Structure of linkage disequilibrium in plants.** *Annu Rev Plant Biol* 2003, **54:**357–374.
3. Ersoz ES, Yu J, Buckler ES: *Applications of linkage disequilibrium and association mapping in crop plants.* Dordrecht: Genomics-assisted crop improvement Springer; 2007:97–120.
4. Li H, Peng ZY, Yang XH, Wang WD, Fu JJ, Wang JH, Han YJ, Chai YC, Guo TT, Yang N, Liu J, Warburton ML, Cheng YB, Hao XM, Zhang P, Zhao JY, Liu YJ GY, Li JS, Yan JB: **Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels.** *Nat Genetics* 2013, **45:**43–U72.
5. Zhao K, Tung CW, Eizenga GC, Wright MH, Ali ML, Price AH, Norton GJ, Islam MR, Reynolds A, Mezey J, McClung AM, Bustamante CD, McCouch SR: **Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*.** *Nat Commun* 2011, **2:**246.
6. Bhosale SU, Stich B, Rattunde HF, Weltzien E, Haussmann BI, Hash CT, Ramu P, Cuevas HE, Paterson AH, Melchinger AE, Parzies HK: **Association analysis of photoperiodic flowering time genes in west and central African sorghum [Sorghum bicolor (L.) Moench].** *BMC Plant Biol* 2012, **12:**32.
7. Edwards D, Batley J, Snowdon RJ: **Accessing complex crop genomes with next-generation sequencing.** *Theor Appl Genet* 2013, **126:**1–11.
8. Marchini J, Cardon LR, Phillips MS, Donnelly P: **The eVects of human population structure on large genetic association studies.** *Nat Genetics* 2004, **36:**512–517.
9. Bus A, Korber N, Snowdon RJ, Stich B: **Patterns of molecular variation in a species-wide germplasm set of *Brassica napus*.** *Theor Appl Genet* 2011, **123:**1413–1423.
10. Xiao Y, Cai D, Yang W, Ye W, Younas M, Wu J, Liu K: *Genetic structure and linkage disequilibrium pattern of a rapeseed (Brassica napus L.) association mapping association mapping in crop plants.* Dordrecht: Genomics-assisted crop improvement Springer; 2012:97–120.
11. Hamblin MT, Salas Fernandez MG, Casa AM, Mitchell SE, Paterson AH, Kresovich S: **Equilibrium processes cannot explain high levels of short- and medium-range linkage disequilibrium in the domesticated grass *Sorghum bicolor*.** *Genetics* 2005, **171:**1247–1256.
12. Bouchet S, Pot D, Deu M, Rami JF, Billot C, Perrier X, Rivallan R, Gardes L, Xia L, Wenzl P, Kilian A, Glaszmann JC: **Genetic structure, linkage disequilibrium and signature of selection in *Sorghum*: lessons from physically anchored DArT markers.** *PLoS One* 2012, **7:**e33470.
13. Bekele WA, Wieckhorst S, Friedt W, Snowdon RJ: **High-throughput genomics in sorghum: from whole-genome resequencing to a SNP screening array.** *Plant Biotechnol J* 2013, **11:**1112–1125.
14. Remington DL, Thornsberry JM, Matsuoka Y, Wilson LM, Whitt SR, Doebley J, Kresovich S, Goodman MM, Buckler ES: **Structure of linkage disequilibrium and phenotypic associations in the maize genome.** *Proc Natl Acad Sci U S A* 2001, **98:**11479–11484.
15. Ching A, Caldwell KS, Jung M, Dolan M, Smith OS, Tingey S, Morgante M, Rafalski AJ: **SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines.** *BMC Genet* 2002, **3:**19.
16. Palaisa KA, Morgante M, Williams M, Rafalski A: **Contrasting effects of selection on sequence diversity and linkage disequilibrium at two phytoene synthase loci.** *Plant Cell* 2003, **15:**1795–1806.

17. Yan JB, Shah T, Warburton ML, Buckler ES, McMullen MD, Crouch J: **Genetic characterization and linkage disequilibrium estimation of a global maize collection using SNP markers.** *PLoS One* 2009, **4:**e8451.
18. Jin L, Lu Y, Xiao P, Sun M, Corke H, Bao J: **Genetic diversity and population structure of a diverse set of rice germplasm for association mapping.** *Theor Appl Genet* 2010, **121:**475–487.
19. Mather KA, Caicedo AL, Polato NR, Olsen KM, McCouch S, Purugganan MD: **The extent of linkage disequilibrium in rice (Oryza sativa L.).** *Genetics* 2007, **177:**2223–2232.
20. Nordborg M, Hu TT, Ishino Y, Jhaveri J, Toomajian C, Zheng H, Bakker E, Calabrese P, Gladstone J, Goyal R, Jakobsson M, Kim S, Morozov Y, Padhukasahasram B, Plagnol V, Rosenberg NA, Shah C, Wall JD, Wang J, Zhao K, Kalbfleisch T, Schulz V, Kreitman M, Bergelson J: **The pattern of polymorphism in *Arabidopsis thaliana*.** *PLoS Biol* 2005, **3:**e196.
21. Hagenblad J, Nordborg M: **Sequence variation and haplotype structure surrounding the flowering time locus FRI in *Arabidopsis thaliana*.** *Genetics* 2002, **161:**289–298.
22. Ecke W, Clemens R, Honsdorf N, Becker HC: **Extent and structure of linkage disequilibrium in canola quality winter rapeseed (*Brassica napus* L.).** *Theor Appl Genet* 2010, **120:**921–931.
23. Zou J, Jiang C, Cao Z, Li R, Long Y, Chen S, Meng J: **Association mapping of seed oil content in *Brassica napus* and comparison with quantitative trait loci identified from linkage mapping.** *Genome* 2010, **53:**908–916.
24. Cruciani F, Santolamazza P, Shen PD, Macaulay V, Moral P, Olckers A, Modiano D, Holmes S, Destro-Bisol G, Coia V, Wallace DC, Oefner PJ, Torroni A, Cavalli-Sforza LL, Scozzari R, Underhill PA: **A back migration from Asia to sub-Saharan Africa is supported by high-resolution analysis of human Y-chromosome haplotypes.** *Am J Hum Genet* 2002, **70:**1197–1214.
25. Gore MA, Chia JM, Elshire RJ, Sun Q, Ersoz ES, Hurwitz BL, Peiffer JA, McMullen MD, Grills GS, Ross-Ibarra J, Ware DH, Buckler ES: **A first-generation haplotype map of maize.** *Science* 2009, **326:**1115–1117.
26. Delourme R, Falentin C, Fomeju BF, Boillot M, Lassalle G, Andre I, Duarte J, Gauthier V, Lucante N, Marty A, Pauchon M, Pichon JP, Ribiere N, Trotoux G, Blanchard P, Riviere N, Martinant JP, Pauquet J: **High-density SNP-based genetic map development and linkage disequilibrium assessment in *Brassica napus* L.** *BMC Genomics* 2013, **14:**120.
27. Allender CJ, King GJ: **Origins of the amphiploid species Brassica napus L. investigated by chloroplast and nuclear molecular markers.** *BMC Plant Biol* 2010, **10:**54.
28. Hasan M, Seyis F, Badani AG, Pons-Kuhnemann J, Lühs W, Friedt W, Snowdon RJ: **Surveying genetic diversity in the *Brassica napus* gene pool using SSR markers.** *Genet Res Crop Evol* 2006, **53:**793–802.
29. Murray MG, Thompsom WF: **Rapid isolation of high molecular weight plant DNA.** *Nucleic Acids Res* 1980, **8:**4321–4325.
30. Chalhoub B, Denoeud F, Liu S, Parkin IAP, Tang H, Wang X, Chiquet J, Belcram H, Tong C, Samans B, Corréa M, Da Silva C, Just J, Falentin C, Koh CS, Le Clainche I, Bernard M, Bento P, Noel B, Labadie K, Alberti A, Charles M, Arnaud D, Guo H, Daviaud C, Alamery S, Jabbari K, Zhao M, Edger PP, Chelaifa H et al.: **Early allopolyploid evolution in the post-neolithic *Brassica napus* oilseed genome.** *Science* 2014. in press.
31. Nei M: **Genetic distance between populations.** *Am Nat* 1972, **106:**283–292.
32. Liu KJ, Muse SV: **PowerMarker: an integrated analysis environment for genetic marker analysis.** *Bioinformatics* 2005, **21:**2128–2129.
33. Pritchard JK, Stephens M, Donnelly P: **Inference of population structure using multilocus genotype data.** *Genetics* 2000, **155:**945–959.
34. Rohlf FJ: *NTSYS-pc. Numerical Taxonomy and Multivariate Analysis System, Version 2.1.* New York: Exeter Software; 2000.
35. Rambaut A: *Fig Tree version 1.3.1. [computer program].* 2009 [http://tree.bio.ed.ac.uk]
36. Hardy OJ, Vekemans X: **spagedi: a versatile computer program to analyze spatial genetic structure at the individual or population levels.** *Mol Ecol Notes* 2002, **2:**618–620.
37. Yu JM, Pressoir G, Briggs WH, Bi IV, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, Kresovich S, Buckler ES: **A unified mixed-model method for association mapping that accounts for multiple levels of relatedness.** *Nat Genet* 2006, **38:**203–208.
38. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES: **TASSEL: software for association mapping of complex traits in diverse samples.** *Bioinformatics* 2007, **23:**2633–2635.
39. Barrett JC, Fry B, Maller J, Daly MJ: **Haploview: analysis and visualization of LD and haplotype maps.** *Bioinformatics* 2005, **21:**263–265.
40. Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D: **The structure of haplotype blocks in the human genome.** *Science* 2002, **296:**2225–2229.
41. Hall TA: *BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT Nucleic Acids Symposium Series*, Volume 41. ; 1999:95–98.
42. Soderlund C, Bomhoff M, Nelson W: **SyMAP: A turnkey synteny system with application to multiple large duplicated plant sequenced genomes.** *Nucleic Acids Res* 2011, **10:**e68.
43. Shin J-H, Blay S, McNeney B, Graham J: **LDheatmap: an R function for graphical display of pairwise linkage disequilibria between single nucleotide polymorphisms.** *J Stat Soft* 2006, **16:**1–10.
44. Roscoe TJ, Lessire R, Puyaubert J, Renard M, Delseny M: **Mutations in the fatty acid elongation 1 gene are associated with a loss of beta-ketoacyl-CoA synthase activity in low erucic acid rapeseed.** *FEBS Lett* 2001, **492:**107–111.
45. Chen S, Nelson MN, Ghamkhar K, Fu T, Cowling WA: **Divergent patterns of allelic diversity from similar origins: thecase of oilseed rape (Brassica napus L.) in China and Australia.** *Genome* 2008, **51:**1–10.
46. Qian W, Meng J, Li M, Frauen M, Sass O, Noack J, Jung C: **Introgression of genomic components from Chinese *Brassica rapa* contributes to widening the genetic diversity in rapeseed (*B. napus* L.), with emphasis on the evolution of Chinese rapeseed.** *Theor Appl Genet* 2006, **113:**49–54.
47. Liu H: *Rapeseed genetics and breeding.* Shanghai: Shanhai Science and Technology Press; 1985:559–566.
48. Shiga T: **Rapa breeding by interspecific crossing between *Brassica napus* and *Brassica campestris* in japan.** *Jpn Agric Res Quart* 1970, **5:**5–10.
49. Wall JD, Pritchard JK: **Assessing the performance of the haplotype block model of linkage disequilibrium.** *Am J Hum Genet* 2003, **73:**502–515.
50. Buckler E, Gore M: **An Arabidopsis haplotype map takes root.** *Nat Genet* 2007, **39:**1056–1057.
51. Mace ES, Tai S, Gilding EK, Li Y, Prentils PJ, Bian L, Campbell BC, Hu W, Innes DJ, Han X, Cruickshank A, Dai C, Frere C, Zhang H, Hunt CH, Wang X, Shatte T, Wang M, Su Z, Li J, Lin X, Godwin ID, Jordan DR, Wang J: **Whole-genome sequencing reveals untapped genetic potential in Africa's indigenous cereal crop sorgum.** *Nat Commun* 2013, **4:**2320.
52. Leflon M, Grandont L, Eber F, Huteau V, Coriton O, Chelysheva L, Jenczewski E, Chèvre AM: **Crossovers get a boost in *Brassica* allotriploid and allotetraploid hybrids.** *Plant Cell* 2010, **22:**2253–2264.
53. Downey RK, Klassen AL, Stringam GR: **Rapeseed and Mustard**. In *Hybridization of crop plants.* Edited by Fehr WR, Hadley H. Madison, WI: ASA, CSA; 1980:495–509.
54. Bennett RA, Thiagarajah MR, King JR, Rahman MH: **Interspecific cross of Brassica oleracea var. alboglabra and B.napus: effects of growth condition and silique age on the efficiency of hybrid production, and inheritance of erucic acid in the self-pollinated backcross generation.** *Euphytica* 2008, **164:**593–601.
55. Gorelick R: **Transposable elements suppress recombination in all meiotic eukaryotes, including automictic ancient asexuals: a reply to Schön and Martens.** *J Nat Hist* 2003, **37:**903–909.
56. Rafalski A, Morgante M: **Corn and humans: recombination and linkage disequilibrium in two genomes of similar size.** *Trends Genet* 2004, **20:**103–111.
57. Olsson G: **Species crosses within the genus *Brassica*.** *Hereditas* 1960, **46:**351.
58. McNaughton IH: **Swedes and Rapes**. In *Evolution of Crop Plants.* Edited by Simmonds NW. London: Longman; 1976:53–56.
59. Guryev V, Smits BMG, van de Belt J, Verheul M, Hubner N, Cuppen E: **Haplotype block structure is conserved across mammals.** *Plos Genetic* 2006, **2:**1111–1118.
60. Wang N, Li F, Chen B, Xu K, Yan G, Qian J, Li J, Gao G, Bancroft I, Meng J, King GJ, Wu X: **Genome-wide investigation of genetic changes during modern breeding of *Brassica napus*.** *Theor Appl Genet* 2014, **127:**1817–1829.
61. Mei J, Li Q, Qian L, Fu Y, Li J, Frauen M, Qian W: **Genetic investigation of the origination of allopolyploid with virtually synthesized lines: application to the C subgenome of *Brassica napus*.** *Heredity* 2011, **106:**955–961.

62. Rygulla W, Friedt W, Seyis F, Lühs W, Eynck C, von Tiedemann A, Snowdon RJ: **Combination of resistance to *Verticillium longisporum* from zero erucic acid *Brassica oleracea* and oilseed *Brassica rapa* genotypes in resynthesized rapeseed (*Brassica napus*) lines.** *Plant Breed* 2007, **126:**596–602.
63. Rygulla W, Snowdon RJ, Eynck C, von Tiedemann A, Lühs W, Friedt W: **Broadening the genetic basis of *Verticillium longisporum* resistance in *Brassica napus* by interspecific hybridisation.** *Phytopathology* 2007, **97:**1391–1396.
64. Ding Y, Mei J, Li Q, Liu Y, Wan H, Wang L, Becker HC, Qian W: **Improvement of *Sclerotinia sclerotiorum* resistance in *Brassica napus* by using *B. oleracea*.** *Genet Resour Crop Evol* 2013, **60:**1615–1619.
65. Jesske T, Olberg B, Schierholt A, Becker HC: **Resynthesized lines from domesticated and wild *Brassica* taxa and their hybrids with *B. napus* L.: genetic diversity and hybrid yield.** *Theor Appl Genet* 2013, **126:**1053–1065.
66. Cornille A, Gladieux P, Smulders MJM, Roldan-Ruiz I, Laurens F, Le Cam B, Nersesyan A, Clavel J, Olonova M, Feugey L, Gabrielyan I, Zhang XG, Tenaillon MI, Giraud T: **New insight into the history of domesticated apple: secondary contribution of the European wild apple to the genome of cultivated varieties.** *Plos Genetics* 2012, **8:**e1002703.
67. Luo MC, Yang ZL, You FM, Kawahara T, Waines JG, Dvorak J: **The structure of wild and domesticated emmer wheat populations, gene flow between them, and the site of emmer domestication.** *Theor Appl Genet* 2007, **114:**947–959.
68. Myles S, Boyko AR, Owens CL, Brown PJ, Grassi F, Aradhya MK, Prins B, Reynolds A, Chia JM, Ware D, Bustamante CD, Buckler ES: **Genetic structure and domestication history of the grape.** *Proc Natl Acad Sci U S A* 2011, **108:**3530–3535.
69. Harper AL, Trick M, Higgins J, Fraser F, Clissold L, Wells R, Hattori C, Werner P, Bancroft I: **Associative transcriptomics of traits in the polyploid crop species *Brassica napus*.** *Nat Biotechnol* 2012, **30:**798–802.

doi:10.1186/1471-2164-15-1170
**Cite this article as:** Qian *et al.*: Sub-genomic selection patterns as a signature of breeding in the allopolyploid *Brassica napus* genome. *BMC Genomics* 2014 **15**:1170.

# 4 Haplotype hitchhiking promotes trait coselection in *Brassica napus*

Lunwen Qian, Wei Qian and Rod J Snowdon

*Plant biotechology journal* 2016

Vol 14, pp.1578-88

Plant Biotechnology Journal

*Plant Biotechnology Journal* (2016), pp. 1–11 doi: 10.1111/pbi.12521

# Haplotype hitchhiking promotes trait coselection in *Brassica napus*

Lunwen Qian[1], Wei Qian[2] and Rod J. Snowdon[1,*]

[1]*Department of Plant Breeding, IFZ Research Centre for Biosystems, Land Use and Nutrition, Justus Liebig University, Giessen, Germany*
[2]*College of Agronomy and Biotechnology, Southwest University, Chongqing, China*

*Received 15 September 2015;*
*revised 16 November 2015;*
*accepted 25 November 2015.*
**Correspondence* (Tel +49 641 9937420; fax +49 641 9937429; email rod.snowdon@agrar.uni-giessen.de)

**Keywords:** linkage disequilibrium, LD, haplogroups, selection, oilseed rape, breeding.

## Summary

Local haplotype patterns surrounding densely spaced DNA markers with significant trait associations can reveal information on selective sweeps and genome diversity associated with important crop traits. Relationships between haplotype and phenotype diversity, coupled with analysis of gene content in conserved haplotype blocks, can provide insight into coselection for nonrelated traits. We performed genome-wide analysis of haplotypes associated with the important physiological and agronomic traits leaf chlorophyll and seed glucosinolate content, respectively, in the major oilseed crop species *Brassica napus*. A locus on chromosome A01 showed opposite effects on leaf chlorophyll content and seed glucosinolate content, attributed to strong linkage disequilibrium (LD) between orthologues of the chlorophyll biosynthesis genes *EARLY LIGHT-INDUCED PROTEIN* and *CHLOROPHYLL SYNTHASE*, and the glucosinolate synthesis gene *ATP SULFURYLASE 1*. Another conserved haplotype block, on chromosome A02, contained a number of chlorophyll-related genes in LD with orthologues of the key glucosinolate biosynthesis genes *METHYLTHIOALKYMALATE SYNTHASE-LIKE 1* and *3*. Multigene haplogroups were found to have a significantly greater contribution to variation for chlorophyll content than haplotypes for any single gene, suggesting positive effects of additive locus accumulation. Detailed reanalysis of population substructure revealed a clade of ten related accessions exhibiting high leaf chlorophyll and low seed glucosinolate content. These accessions each carried one of the above-mentioned haplotypes from A01 or A02, generally in combination with further chlorophyll-associated haplotypes from chromosomes A05 and/or C05. The phenotypic rather than pleiotropic correlations between leaf chlorophyll content index and seed GSL suggest that LD may have led to inadvertent coselection for these two traits.

## Introduction

Chlorophyll is a green photosynthetic pigment with which plant chloroplasts generate energy from light. Leaf chlorophyll content relates to photosynthetic capacity and is thus one of the important physiological traits influencing crop yield (Czyczyło-Mysza *et al.*, 2013; Wang *et al.*, 2008). On the other hand, the presence of chlorophyll in mature seeds can be an undesirable trait that can affect seed maturation, seed oil quality, meal quality and germination (Delmas *et al.*, 2013). Chlorophyll pigments remaining in processed vegetable oils are also associated with increased oxidation causing rancidity (Tautorus and Low, 1993) and difficulties in hydrogenation (Abraham and Deman, 1986). Recent studies suggested a previously unknown relationship between chlorophyll content and metabolism of glucosinolates, a class of sulphurous secondary metabolites expressed in vegetative and generative tissues throughout almost all plants of the *Brassicales*. For example, a total of 11 light-harvesting chlorophyll (LHC) protein complex proteins (including eight LHCB and three LHCA proteins) were identified to be down-regulated in RNAi lines that suppressed molecular networks controlling glucosinolate metabolism in Arabidopsis (Chen *et al.*, 2012). Yang and Zhu (2009) reported a potential negative correlation between chlorophyll and glucosinolate content under abiotic stresses in cabbage plants. These results suggest potential genetic and/or metabolic associations between chlorophyll content and glucosinolate metabolism in *Brassica* crops. A molecular basis for such a pleiotropic relationship might be associated with the role of plastids in sulphate reduction and cysteine/methionine synthesis in the chloroplasts, providing sulphuric amino acids that are exported across the chloroplast membrane and play an important role as precursors for glucosinolate synthesis (Takahashi *et al.*, 2011). Such a relationship is potentially relevant both evolutionarily and agronomically. Leaf glucosinolates have demonstrated positive nutritional value in vegetable *Brassica* crops (Murillo and Mehta, 2001) and play an important role, either antagonistically or mutualistically, in plant interactions with insect pests. On the other hand, glucosinolates are highly undesirable in *Brassica* oilseed meals fed to livestock (Friedt and Snowdon, 2010). The huge global importance of oilseed rape and canola (*B. napus*), the world's second most important oilseed crop, imparts enormous agroeconomic relevance on these compounds.

Genome-wide association studies (GWAS) examine cotransmission of phenotypes with genetic markers, normally based on

Please cite this article as: Qian, L., Qian, W. and Snowdon, R. (2016) Haplotype hitchhiking promotes trait coselection in *Brassica napus*. *Plant Biotechnol. J.*, doi: 10.1111/pbi.12521

linkage disequilibrium (LD) analysis in genetically diverse populations using panels of markers spanning the entire genome at high density. Besides providing high mapping resolution by incorporating historical recombination events, LD analyses can also provide important insight into the history of both natural and artificial selection (breeding) and give valuable guidance to breeders seeking to diversify crop gene pools.

Natural and artificial selection can cause conservation of haplotype blocks, comprising specific combinations of nucleotides on the same chromosome, in genome regions carrying genes under positive or negative selection. Haplotypes can therefore provide more information than any single SNP regarding the complex relationship between DNA variation and quantitative phenotypes (Stephens *et al.*, 2001). Elucidating the evolutionary relationships among local haplotypes can further improve the detection power of GWAS scans (Buntjer *et al.*, 2005). For example, detailed analysis of LD surrounding major QTL revealed strong signatures of artificial selection associated with important traits in different breeding pools of rapeseed and bread wheat (Qian *et al.*, 2014; Voss-Fels and Snowdon, 2015; Voss-Fels *et al.*, 2015). In such cases, haplotypes reveal the extent to which genetic variation in a given chromosome region is described by clustering markers. Comparing haplotype diversity can help to understand the effects of natural and artificial selection on genome-scale and single-gene variation, as shown recently in grapevine (Fernandez *et al.*, 2014), maize (Yang *et al.*, 2013) and Arabidposis (Li *et al.*, 2014). New high-density genome screening tools provide an unprecedented level of insight into local LD patterns in even complex crop genomes (Edwards *et al.*, 2013; Voss-Fels and Snowdon, 2015). For example, strongly selected haplotype patterns detected in high-density population genomic studies have been associated with domestication, adaptation and breeding in sorghum (Mace *et al.*, 2013), rapeseed (Qian *et al.*, 2014) and bread wheat (Voss-Fels *et al.*, 2015), respectively.

Evolutionary selection pressures frequently act on entire pathways or their functional subnetworks. Multiple interacting genes may change in the same fitness direction, at a similar evolutionary rate and across the same timescale, to achieve a common phenotypic outcome. Associations in evolutionary patterns may therefore simply reflect parallel selection of different genes in the same pathway with shared functionality. On the other hand, artificial selection in plant breeding targets recombinations surrounding chromosome regions that carry variants conferring traits of agricultural or economic interest. Different traits that interact via molecular networks may be unintentionally coselected due to pleiotropy, whereas traits controlled by distinct networks may be coselected via LD between functionally independent genes. Haplotype blocks can provide powerful insight into the causes of correlations between different, quantitative trait phenotypes and their associations with responsible genes within haplotype regions.

In this study, we used a high-density SNP genotyping array to identify haplotype blocks associated with leaf chlorophyll content index (CCI) and seed glucosinolate content in a diverse *B. napus* population. Gene content within haplotype blocks for these two traits suggests they have been coselected during breeding of high-yielding, high-quality, modern oilseed rape cultivars. Our results suggest their interrelationship in *B. napus* can be explained by hitchhiking selection due to LD between functionally nonrelated genes.

## Results

### Variation and correlations for leaf chlorophyll and seed glucosinolate content

Broad variation was observed in leaf CCI across the diversity panel, in different field and glasshouse environments, and across different plant developmental stages. Highly significant ($P < 0.001$) positive correlations were seen between different developmental stages in the same location (Figure 1). Weaker but nevertheless significant correlations were measured between field and glasshouse, while leaf CCI at bolting stage in the glasshouse was not significantly correlated with bolting stage, flowering stage and mature period in the field. A low heritability of $H^2 = 0.24$ was calculated for leaf CCI, reflecting the strong G*E interaction (Table S1). As expected from previous findings, seed GSL was highly heritable ($H^2 = 0.86$) (Table S1). Correspondingly, highly significant positive correlations ($P < 0.001$) were found among the GSL values from the diversity panel among the three different environments (Figure 1).

### Homoeologous haplotype blocks containing chlorophyll-related genes associated with leaf CCI

Manhattan plots and quantile–quantile plots describing significant SNP associations for leaf CCI in glasshouse and field experiments are shown in Figure S1. A total of 35 and 32 SNPs distributed throughout the genome were detected with the significance threshold of $-\log10^{(P)} = 4$ using the glasshouse and field data, respectively. Associations to CCI that could be corroborated in multiple environments and tissues were selected as candidate loci for further investigation. Candidate regions containing SNPs associated with leaf CCI were investigated at high resolution by assaying haplotype blocks ($r^2 > 0.65$) in flanking chromosome segments. Details of SNPs and candidate genes in haplotype blocks with significant associations to leaf CCI are provided in Figure S2 and Table S2.

Two SNPs (Bn-A05-p19777231 and Bn-A05-p19777547 with $P = 8.61 \times 10^{-7}$ and $8.59 \times 10^{-5}$, respectively) with significant associations to CCI were located in a 113 kb haplotype region on chromosome A05 (position 17 873 133–17 986 390 bp; $r^2 = 0.74$) (Figure 2; Table S3). As shown in Figure 2, comparative analysis via synteny alignments revealed that this region is homologous with a 692 kb haplotype block on chromosome C05 (position 36 453 492–36 605 598 bp; $r^2 = 0.68$) that also showed significant associations to leaf CCI (Table S3). Both homologous regions contain multiple *B. napus* orthologues of the *Arabidopsis* chloroplast membrane gene *PALMITOYL-MONOGALACTOSYLDIACYLGLYCEROL DELTA-7 DESATURASE* (*FAD5*; BnaA05g23670D, BnaA05g23680D, BnaA05g23690D, BnaC05g37420D, BnaC05g37450D and BnaC05g37460D), along with two copies of the photosynthesis gene *POST-ILLUMINATION CHLOROPHYLL FLUORESCENCE INCREASE PROTEIN* (*PIFI*; BnaA05g23700D and BnaC05g37470D) (Table S3). Three and nine haplogroups were observed for the A05 and C05 haplotype regions, respectively. Two haplogroups, A05_Hap2 and C05_Hap5, were found to exhibit higher CCI than the remaining two and eight haplogroups on chromosomes A05 and C05, respectively (*t*-test and mean values; Figure S3; Table S3).

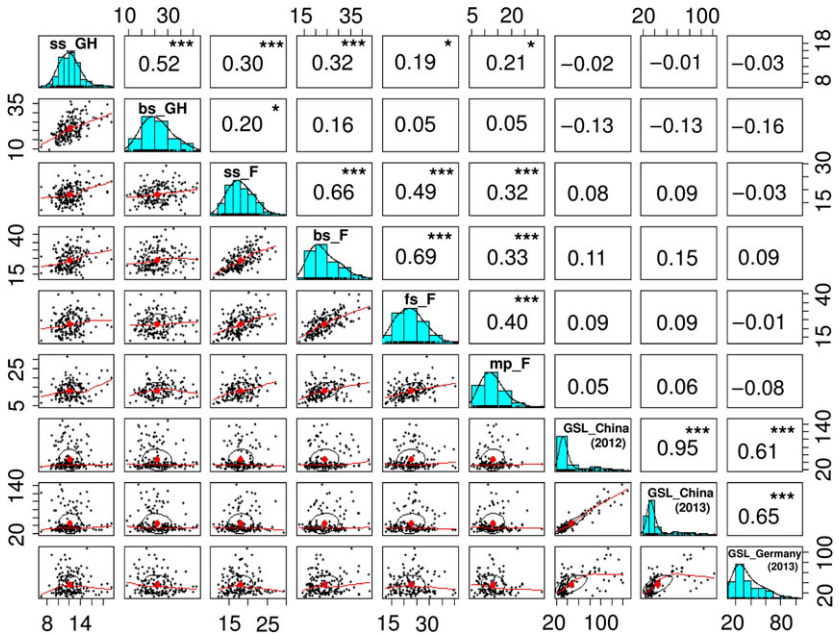



**Figure 1** Correlation coefficients and frequency distributions for chlorophyll content index and GSL in 203 Chinese semi-winter rapeseed accessions. ss_GH: chlorophyll content index in seedling stage (glasshouse experiments, 2012); bs_GH: chlorophyll content index in bolting stage (glasshouse experiments, 2012); ss_F: chlorophyll content index in seedling stage (Field, 2013); bs_F: chlorophyll content index in bolting stage (Field, 2013); fs_F: chlorophyll content index in flowering stage (field experiments, 2013); mp_F: chlorophyll content index in mature period (Field, 2013); GSL: glucosinolate content (*$P \leq 0.05$, **$P \leq 0.01$, ***$P \leq 0.001$).

## Chlorophyll-associated haplotypes carry chloroplast membrane protein genes

Three SNPs (Bn-A05-p9118240, Bn-A05-p9079762 and Bn-A05-p9078313) in a 147 kb haplotype block ($r^2 = 0.66$), from 12 869 710 to 13 017 024 bp on chromosome A02, showed significant associations ($P = 2.088 \times 10^{-5}$, $3.63 \times 10^{-5}$ and $5.17 \times 10^{-5}$, respectively) to leaf CCI (Fig 3a; Table S4). The haplotype block containing this gene also contains *B. napus* orthologues of the Arabidopsis genes *TRANSLOCON AT THE OUTER ENVELOPE MEMBRANE OF CHLOROPLASTS 159* (*TOC159*; BnaA02g20610D) and AT4G02530 (BnaA02g20650D). *TOC159* encodes an essential component of the TOC-protein complex, responsible for recognition and translocation of photosynthetically active proteins through the chloroplast envelope membrane (Bauer *et al.*, 2000), while BnaA02g20650D encodes a chloroplast thylakoid lumen protein involved in photosynthesis and chlorophyll biosynthesis (Ferro *et al.*, 2010; Friso *et al.*, 2004). By comparing leaf CCI phenotypes of the haplogroups for this haplotype block, we found that the haplogroup A02_Hap1 was associated with higher chlorophyll levels than the other five haplogroups (*t*-test and mean value; Figure 3b; Table S4). A02_Hap1 differs strongly differs from all other haplogroups except for A02-Hap4, which has differences at only the first two bases. Genotypes carrying A02-Hap4 also have higher median CCI in later developmental stages (Figure 3, Table S4).

In addition, six SNPs in this haplotype region were located within the gene BnaA02g20650D. Two of these SNPs (Bn-A05-p9079762 and Bn-A05-p9078313), located within a 2 kb haplotype block inside intron 3 of BnaA02g20650D, exhibited significant associations with leaf CCI (Figure 4a; Table S5). We used t-test and mean values to compare phenotype values for three haplogroups identified in the 2 kb haplotype block. Haplogroup A02_Hap1_1 was found to have higher leaf CCI than other two haplogroups (Figure 4b; Table S5). Comparison of A02_Hap1_1 with A02_Hap1 showed that A02_Hap1 (including both BnaA02g20650D and BnaA02g20610D) has higher leaf CCI than A02_Hap1_1 (comprising only BnaA02g20650D) (Figure 5). This result suggests that the multigene haplogroup has a significantly greater contribution to higher leaf chlorophyll content than the single gene haplogroup.

## Haplotype conservation associates with coselection for leaf CCI and seed GSL

In a number of CCI-associated haplotype blocks, we observed strong LD between genes involved in chlorophyll biosynthesis, photosynthesis or chloroplast membrane fatty acid synthesis, and further genes implicated in GSL synthesis. We therefore mapped SNP-trait associations for seed GSL, using phenotype data from three different environments, to test whether the same haplotype block regions are also associated with GSL. A genome-wide significance threshold of $-\log^{10(P)} = 4.4$ was applied to determine SNPs with significant associations to GSL.

Conserved LD blocks containing significant GWAS hits for both seed GSL content and leaf chlorophyll content in chromosome A01 are shown in Figure 6 and Table S6. The SNP markers Bn-A01-p12454306 ($P = 6.00 \times 10^{-5}$) and Bn-A01-p12314813 ($P = 4.57 \times 10^{-6}$) were significantly correlated to CCI and GSL, respectively (Figure 6a; Table S6). These two markers share conserved LD within a 338 kb haplotype block ($r^2 = 0.67$) containing *B. napus* orthologues of three chloroplast-associated *Arabidopsis* genes with functional annotations associated to chlorophyll and glucosinolates, respectively (Figure 6a; Table S6). The genes *EARLY LIGHT-INDUCED PROTEIN* (*ELIP2*; BnaA01g19110D) and *CHLOROPHYLL SYNTHASE* (*CHLG*: BnaA01g19280D) are involved in photosynthesis via regulation

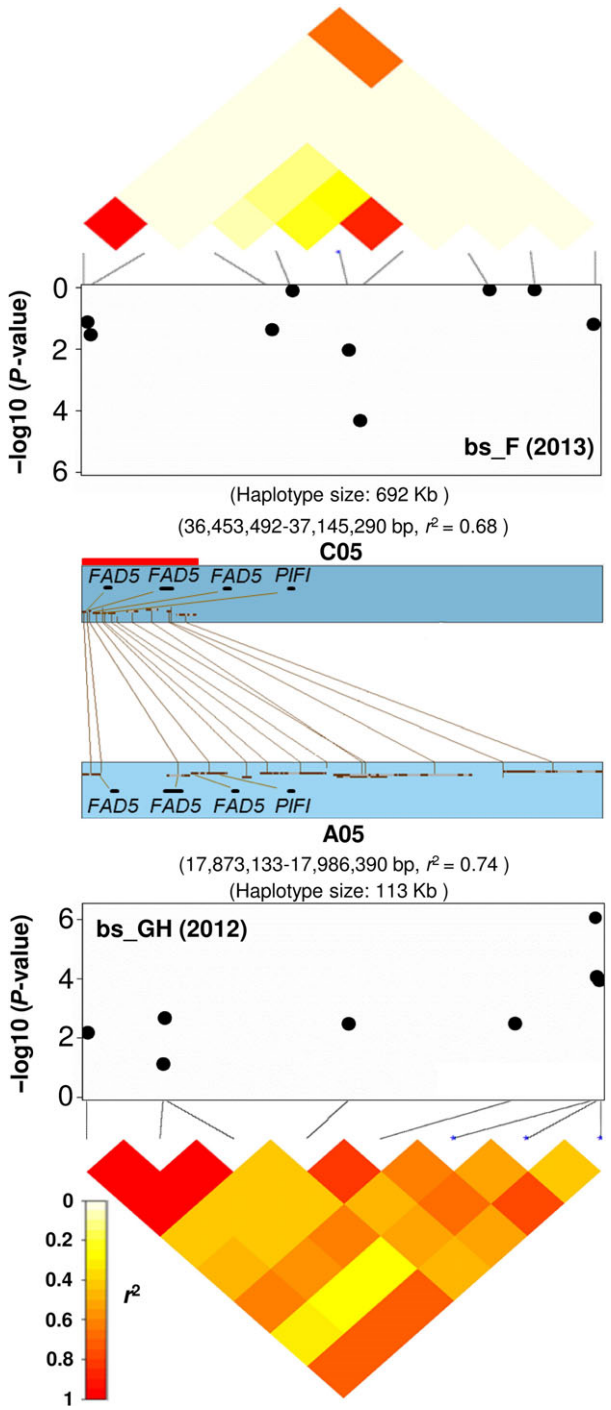



**Figure 2** Association mapping for leaf chlorophyll content index at bolting stage in glasshouse (bs_GH) and field (bs_F) to homoeologous haplotype regions on of *Brassica napus* chromosomes A05 (17 873 133–17 986 390 bp) and C05 (36 453 492–37 145 290 bp). The heatmaps span the SNP markers that show linkage disequilibrium with the most strongly associated SNPs. Three *B. napus* orthologues of the gene *FAD5* and one of *PIFI* are located in each of the homoeologous segments (indicated by the red bar).

of chlorophyll biosynthesis, whereas *ATP SULFURYLASE 1* (*APS1*; BnaA01g19120D) encodes the first enzyme in the sulphate assimilation pathway and therefore has a potential direct impact on synthesis of sulphuric glucosinolates (Table S6). For this haplotype block, the haplogroup A01_Hap5 showed higher leaf CCI, especially at seedling and bolting stages in the glasshouse experiments ($P < 0.05/0.01$), whereas A01_Hap1 showed lower CCI than the other four haplogroups (*t*-test and mean value; Figure 6b; Table S6). Conversely, in all three test environments, haplogroup A01_Hap5 showed significantly lower GSL and haplogroup A01_Hap1 significantly higher GSL than the other four haplogroups (*t*-test and mean value; Figure 6b; Table S6). In almost all cases, the CCI in the glasshouse experiments showed a negative correlation between leaf CCI and seed GSL between haplogroups A01_Hap1 and A01_Hap5, suggesting that this haplotype may be associated with inadvertent coselection for high leaf chlorophyll in breeding materials with low seed GSL.

## Introgressions from winter oilseed rape contribute additively to chlorophyll content

On chromosomes A01, A02, A05 and C05, the high-chlorophyll haplogroups A01_Hap5, A02_Hap1, A05_Hap2 and C05_Hap5 were found in 22 of the 203 accessions (Figures 6b, 3b and S3). Haplotype network analysis across these four haplotype regions showed that A01_Hap5, A02_Hap1 and C05_Hap5 correspond to accessions distributed throughout the subpopulations Q1 and 'mixed', respectively (Figure 7a, b and d), while accessions carrying A05_Hap2 were found only in Q1 (Figure 7c). Seven of the 22 accessions carried more than one of the high-chlorophyll haplogroups, and a further 7 also exhibited elevated leaf CCI (Table S7). 11 of these 14 high-chlorophyll accessions were assigned to subpopulation Q1, which has been shown to be associated with a strong winter rapeseed genetic background Qian *et al.* (2014, Table S7). Collectively, these results suggest that multiple introgressions from winter rapeseed may have elevated chlorophyll content in this group of accessions.

## Haplogroups have additive effects on chlorophyll content

Among the 14 accessions with higher leaf CCI related to haplogroups A01_Hap5, A02_Hap1, A05_Hap2 and C05_Hap5, seven carried combinations of two or three of these haplogroups (designated group A in Table S8). The remaining seven accessions were (designated group B in Table S8) each carry only one of the haplogroups. Comparative phenotype analyses showed the accessions with multiple haplogroups (group A) have higher leaf CCI than those with only one haplogroup (B group; Figure S4; Table S8). This indicates putative additive effects of the selected haplogroups on leaf chlorophyll content.

## Higher chlorophyll content accessions relate to lower GSL

The reanalysis of detailed substructure in subpopulations Q1 and 'mixed' by PCA and UPGMA revealed three clear subgroups (clades Q1_1, Q1_2 and Q1_3), comprising 69, 16 and 59 accessions, respectively (Figure 8a,b; Table S7). The results of the UPGMA tree analysis corresponded with around 84% similarity to the PCA. Haplotype A01_Hap5 (associated with low seed GSL and high leaf CCI) was unique to four accessions distributed across the clade Q1_2. This clade (Figure 8b) also included six accessions with higher leaf CCI associated with the haplotypes A02_Hap1, A05_Hap2 and C05_Hap5, respectively.

Extending the haplotype block A02_Hap1 by only one additional upstream SNP, corresponding to a slightly relaxed LD threshold ($r^2 = 0.62$; extended haplotype block from 12 869 710 to

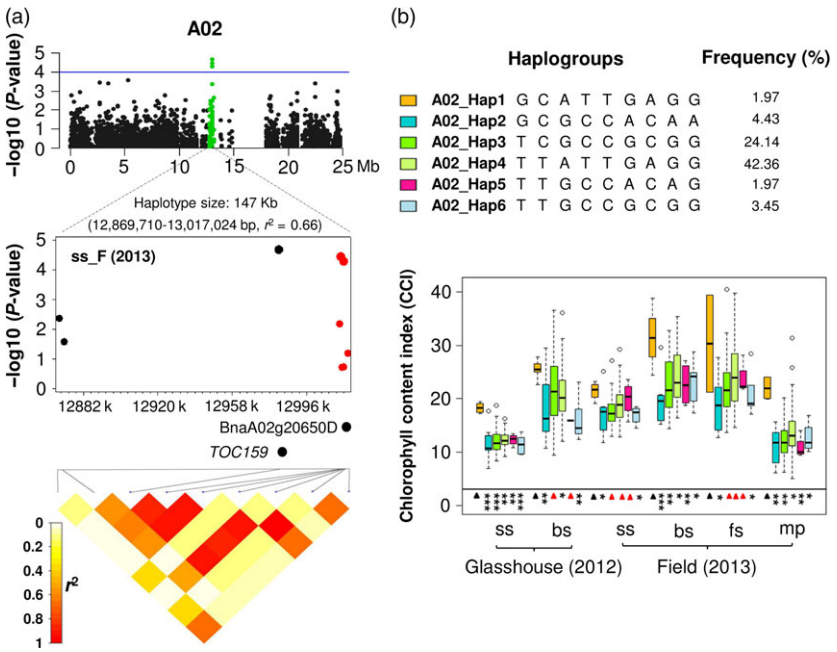



**Figure 3** Association mapping for leaf chlorophyll content index (CCI) on chromosome A02 in 203 Chinese semi-winter rapeseed accessions. (a) Green plots show a 12 869 710–13 017 024 bp haplotype region significantly associated with leaf CCI. The blue horizontal line indicates a threshold of genome-wide significance at a *P* value of $1.0 \times 10^{-4}$. The heatmaps span the SNP markers that show linkage disequilibrium (LD) with the most strongly associated SNPs. Positions of *Brassica napus* orthologues of the *Arabidopsis* genes *TOC159* (BnaA02g20610D) and a chloroplast thylakoid lumen protein (BnaA02g20650D) are marked. The six SNPs labelled in red are located within the gene BnaA02g20650D. (b) Boxplots showing leaf CCI values for six haplogroups with frequency >0.01. Symbols show significant differences of haplogroups compared with A02_Hap1: $*P \leq 0.05$, $**P \leq 0.01$, $***P \leq 0.001$; Red triangles: not significant ($P > 0.05$).

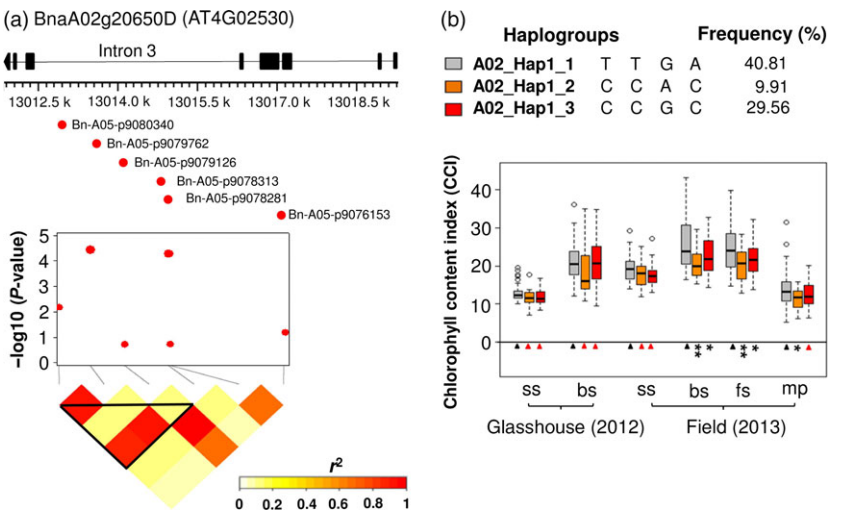



**Figure 4** (a) The SNPs Bn-A05-p9079762 and Bn-A05-p9078313, located in intron 3 of BnaA02g20650D within a 2 kb haplotype block, exhibited significant associations with leaf chlorophyll content index (CCI). (b) Boxplots showing leaf CCI values for three haplogroups with frequency >0.01 within the BnaA02g20650D gene-haplotype region. Symbols show significant differences of haplogroups compared with A02_Hap1_1: $*P \leq 0.05$, $**P \leq 0.01$, $***P \leq 0.001$; Red triangles: not significant ($P > 0.05$).

13 1667 084 bp), resulted in detection of significant associations with seed GSL (Figure S6; Table S9). As was the case for A01_Hap5, haplotypes associated with lowest GSL overlapped with the haplotypes for high CCI. Within the extended A02_GSL_Hap1 haplotype, we located *B. napus* orthologues of the Arabidopsis glucosinolate biosynthesis genes *METHYLTHIOALKYMALATE SYNTHASE-LIKE 3* (*MAM3*; BnaA02g20830D) and *MAM 1* (BnaA02g20840D). In Arabidopsis, *MAM1* and *MAM3* play important roles in the biosynthesis of aliphatic glucosinolates (Kroymann *et al.*, 2001; Field *et al.*, 2004; Textor *et al.*, 2007). The very close proximity of these genes (approximately 100 kb) to the CCI-associated haplotype block confirms the presence of strong LD

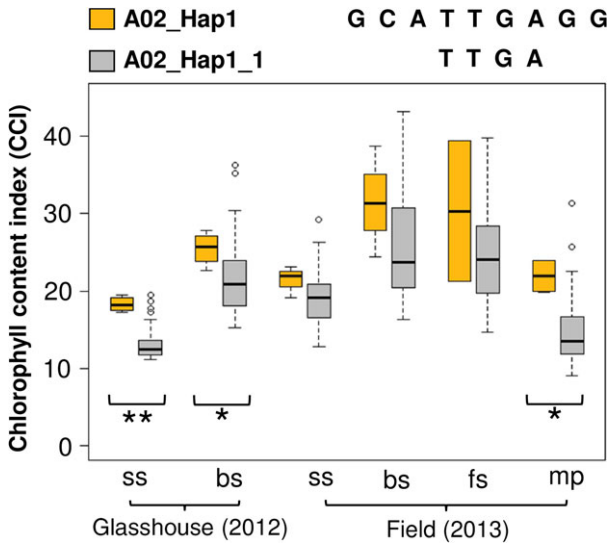



**Figure 5** Boxplots showing phenotypic values for leaf chlorophyll content index (CCI) related to two haplogroups. Comparison between the haplogroups show higher leaf CCI in A02_Hap1 than A02_Hap1_1 (*$P \leq 0.05$, **$P \leq 0.01$).

between putative causal loci for GSL variation and loci with strong effects on CCI. Four accessions belonging to clades Q1_2 and Q1_3 carried the low-GSL haplotype associated with A02_GSL_Hap1, but not A01_Hap5.

A similar situation to that described for A02_Hap1 was also observed for the haplotype A05_Hap2. Within 200 kb from the 113 kb CCI-associated haplotype block ($r^2 = 0.74$), we found additional genes with a putative function in GSL biosynthesis. This extended haplotype block ($r^2 = 0.41$) also showed associations to seed GSL content.

Interestingly, phenotypic comparison among the three clades Q1_1, Q1_2 and Q1_3 revealed higher leaf CCI and lower seed GSL in Q1_2 than in the other two clades (t-test and mean value; Figure S5; Table S7). This suggests that coselection for these two traits have occurred particularly within clade Q1_2. All individuals belonging to clade Q1_2 carried either A01_Hap5 or A02_Hap1, either alone or in combination with A05_Hap2.

## Discussion

Selection for specific agronomic traits during plant domestication and breeding has strong influences on the genetic diversity and population structure within available gene pools for further crop improvement. High-density genotyping tools today provide a means for unprecedented insight into patterns of diversity associated with breeding in major crop species (Qian *et al.*, 2014; Snowdon *et al.*, 2015; Voss-Fels *et al.*, 2015), enabling identification of loci under strong selection and potentially allowing islands of depleted diversity to be addressed by targeted, marker-assisted introgressions without compromising desirable adaptation, yield or quality traits (Voss-Fels and Snowdon, 2015).

Besides their negative influence on genetic diversity, strong signatures of selection associated with key traits can also cause coselection of loci with undesirable effects, resulting in what is commonly known as linkage drag. In this study, we present an interesting example in oilseed rape for coselection of increased leaf chlorophyll content along with reduced seed glucosinolate content, two traits with no immediately obvious biological relationship. We demonstrate that introgressions between ecogeographically distinct gene pools resulted in indirect selection of plants with elevated levels of leaf chlorophyll, due to coselection of beneficial haplotypes at four independent haplotype blocks. Within one of these haplotype blocks, a gene involved in chlorophyll synthesis showed almost complete LD with a gene associated with seed glucosinolate content. The relationship between the observed haplotypes and the respective phenotypic behaviour suggests that introgression of this locus during breeding, to reduce seed glucosinolate levels in nutritionally valuable varieties, causes hitchhiking selection resulting in increased chlorophyll content. Elevated chlorophyll may result in improved photosynthetic performance. While this does not necessarily relate *per se* to improved agronomic performance, stay-green traits associated with increased chlorophyll production, or suppression of chlorophyll degradation, are implicated in improved adaptation of crops to abiotic stresses like water or nutrient deficiency (Thomas and Ougham, 2014).

In this study, a total of nine haplotype blocks were found to harbour significant associations ($r^2 \geq 0.65$) with leaf CCI. Within these regions, we found eighteen genes implicated in chlorophyll synthesis or catabolism, respectively. These included six orthologues of *FAD5*, which was found in Arabidopsis to influence chlorophyll biosynthesis (Heyndrickx and Vandepoele, 2012) and restore leaf chlorophyll content (Heilmann *et al.*, 2004), along with two orthologues of the gene *PIFI*, which has an indirect effect on photosynthesis (Gotoh *et al.*, 2010). The observations of homoeologous, trait-associated haplotypes, carrying multiple duplicated genes, provide a further example for gene amplification by polyploidization. In the complex allopolyploid crop *B. napus*, this phenomenon has been shown to influence numerous agronomically important traits (Chalhoub *et al.*, 2014).

Additional chlorophyll-associated haplotypes carried *B. napus* orthologues of the genes *ELIP2* and *CHLG*, which in Arabidopsis are directly involved in chlorophyll biosynthesis (Oster and Rüdiger, 1998; Tzvetkova-Chevolleau *et al.*, 2007), and the chloroplast preprotein transporter protein gene *TOC159*. The latter encodes an essential component of the chloroplast assembly mechanism, which acts as a selective import receptor for preproteins required in chloroplast development (Smith *et al.*, 2004). Collectively, the genetic associations we found, involving multiple loci carrying these very different contributors to the photosynthesis apparatus, suggest a broad genetic variation for factors involved in photosynthesis in crop plants. Given the key role of photosynthetic activity in ecogeographical and stress adaptation, it is interesting that multiple loci with positive additive influences on chlorophyll content are coselected in introgressions between different breeding pools in this major crop.

The close genetic relationship among the clade of accessions with simultaneously elevated leaf CCI and low seed GSL suggests that this clade may be the product of artificial selection from a common genetic background. Of the four CCI-associated haplotypes, two (on chromosomes A01 and A02) carry well-known glucosinolate biosynthesis genes either within or directly adjacent to the LD block responsible for increase chlorophyll concentration. The third haplotype block, on chromosome A05, is closely linked to genes that have a putative function in GSL synthesis, while its homoeologous locus on chromosome C05 can be

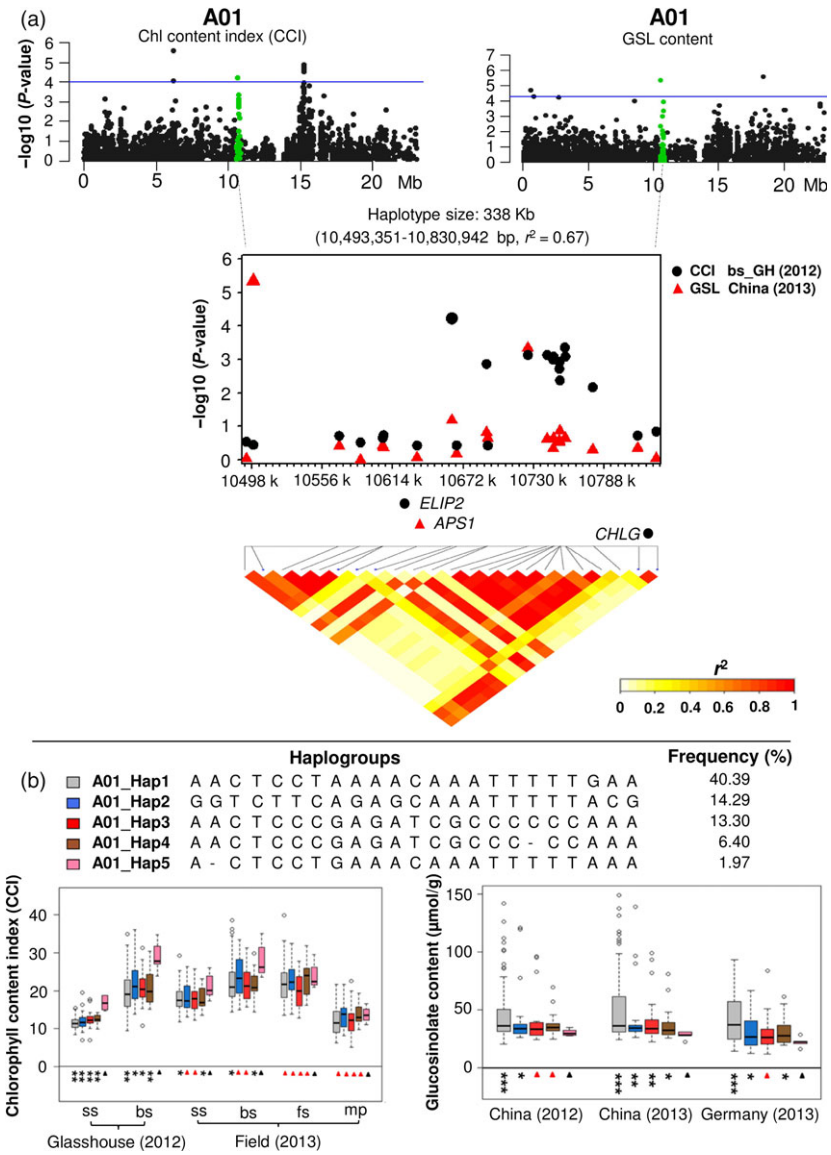



**Figure 6** Association mapping for leaf chlorophyll content index (CCI) and seed GSL on chromosome A01. Green plots represent a haplotype region (36 453 492–37 145 290 bp) significantly associated with leaf CCI and seed GSL, respectively. The blue horizontal line indicates a threshold of genome-wide significance at a *P* value of $1.0 \times 10^{-4}$ and $4.5 \times 10^{-5}$ for CCI and GSL, respectively. The SNPs with highest p values for each trait were used to define a haplotype region with strong linkage disequilibrium (LD) to the traits, containing *Brassica napus* orthologues of two Arabidopsis genes related to chlorophyll (*ELIP2* and *CHLG*) and one related to GSL (*APS1*). (b) Five haplogroups with frequency >0.01 were found in this haplotype region. Boxplots show that A01_Hap5 has higher leaf CCI and lower seed GSL than the other four haplogroups. Symbols show significant differences of haplogroups compared with A01_Hap5: $*P \leq 0.05$, $**P \leq 0.01$, $***P \leq 0.001$; Red triangles: not significant ($P > 0.05$).

expected to carry a corresponding repertoire of homologous genes. Collectively, these examples represent an accumulation of experimental evidence (A02_Hap5 and A02_Hap1) and putative support (A05_Hap2 and C05_Hap5) for a general genetic linkage between loci involved in expression of seed GSL and leaf chlorophyll content in *B. napus*. Leaf CCI and seed GSL exhibit significant associations to different SNP loci that in turn do not appear to show pleiotropic associations to the different traits. The presence of highly plausible positional and functional candidate genes for both traits, and their strong LD to most of the trait-associated SNPs, provides additional support for the hypothesis that the phenotypic associations we observed are caused by linkage rather than pleiotropy.

The genetic determinants of these two traits appear to differ among the three subclades we identified, however, with overlapping associations among some loci but not others. This suggests that the apparent coselection of chlorophyll content, which in contrast to seed GSL, is not a breeding target in rapeseed, most likely occurred inadvertently (through linkage to loci conferring low seed GSL) rather than intentionally. As such, this represents an interesting example for local enrichment of a trait with specific adaptation potential as an indirect consequence of intensive breeding for seed quality characters.

It is likely that the expression of many genes is jointly controlled by the actions of multiple regulatory alleles. Despite this, GWAS studies rarely consider the possibility that, at a given locus,

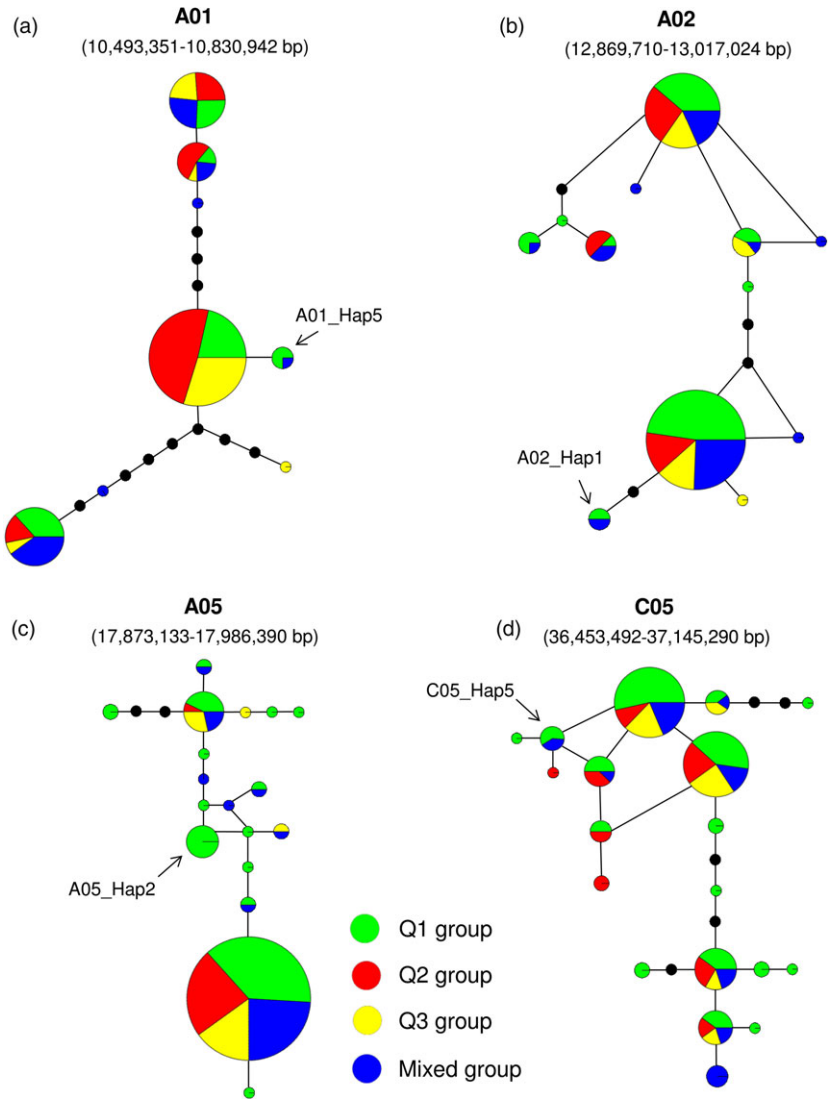



**Figure 7** Haplotype networks in the four selected haplotype regions on chromosomes (a) A01 (10 493 351–10 830 942 bp), (b) A02 (12 869 710–13 017 024 bp), (c) A05 (17 873 133–17 986 390 bp) and (d) C05 (36 453 492–37 145 290 bp). Each circle represents a haplogroup, and the size of the circle is proportional to the number of lines within the haplogroup. Colours represent four different subgroups. The arrows indicate the four haplogroups from the Q1 and mixed subpopulations, respectively, which consistently exhibit higher leaf chlorophyll content index.

multiple genes may impact a phenotype by interactions between more than one regulatory allele in across a gene-haplotype region (Corradin *et al.*, 2014). Particularly for complex, multigenic traits such as chlorophyll content, a spatial proximity of interacting genes can potentially facilitate coordinated expression in certain tissues, developmental timepoints or in response to environmental stimuli. This may help explain why multigenic haplogroups associate more strongly with CCI phenotypes than single-gene haplotypes. The results of the present study underline the benefit of combining haplotype diversity analysis with GWAS studies to dissect additive effects of quantitative trait loci in crops and understand their underlying biology (Buntjer *et al.*, 2005). Detailed investigations of trait relationships at the local haplotype level, using high-density SNP markers in large populations, also provide plant breeders with a means to distinguish between genetic and pleiotropic trait correlations.

## Materials and methods

### Plant materials, genotype and phenotype data

A diversity panel of 203 homozygous *B. napus* inbred lines was constructed to broadly represent variability in Chinese semi-winter rapeseed, an intermediate form of oilseed *B. napus* that is broadly grown in China and also commonly used to enrich gene pools of European winter oilseed rape and Australian or North American spring canola. The plant population (Table S1) and high-density SNP data, generated by genotyping with the *Brassica* 60k SNP Illumina consortium genotyping array (Illumina, San Diego, CA), were described in detail by Qian *et al.* (2014).

The association panel was evaluated in separate glasshouse and field experiments. A CCI was calculated based on absorbance measurements at 653 and 931 nm with a CCM-200 chlorophyll content metre (Opti-Sciences, Inc., Hudson, NH). Measurements of leaf CCI were performed at seedling and bolting stage in the glasshouse experiment in 2012. At each developmental stage, two independent measurements were taken from each side of a single young leaf on three individual plants per accession. In the field, the accessions were sown in single rows with two replications. Chlorophyll content index measurements were performed on five plants per accession per replication at seedling stage, bolting stage, flowering stage and maturity, using the same measurement procedure as in the glasshouse experiment. Additional field trials for seed quality analysis were performed at the experimental farm of Southwest University in Beibei, Chongqing, China, in 2012 and 2013. Glucosinolate content in

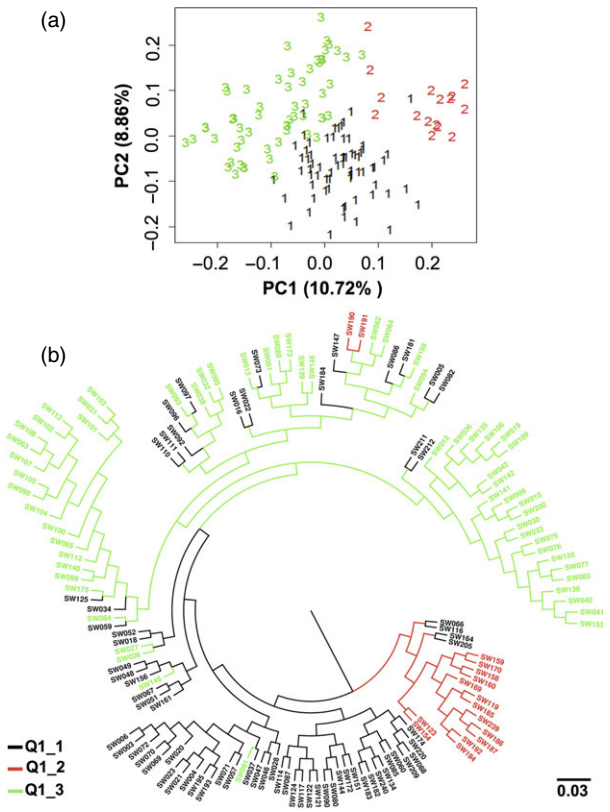



**Figure 8** One hundred and forty-four accessions belonging to the Q1 and mixed subpopulations were reanalysed for detailed population structure by (a) PCA and (b) UPGMA. The 144 accessions clustered into the three clades Q1_1, Q1_2 and Q1_3. The blue dots represent accessions carrying haplogroups, A02_Hap1, A05_Hap2 and C05_Hap5, while the red dots represent accessions carrying haplogroup A01_Hap5.

harvested, fully mature seeds from all 203 accessions was measured by near-infrared spectroscopy on seeds grown from all three field environments, recording mean values in glucosinolate (μmol) per seed dry weight (g) from at least two technical and two biological replicates per accession and environment.

### Statistical analysis

Heritability ($H^2$) for the two traits was calculated using the statistical software package SPSS Statistics for Windows Version 22.0 (IBM Corp., Armonk, NY). Distributions and correlations among the traits and environments were analysed using the R package *psych* (Revelle, 2014) and HMISC (Harrell and Dupont, 2012).

### Genome-wide association analysis

A total of 24 338 high-quality, single-locus single-nucleotide polymorphism (SNP) markers with minor allele frequency (MAF) ≥0.05 were used for the GWAS and LD analyses. The mixed linear model was as follows:

$$y = X\alpha + P\beta + K\mu + e$$

It was used to test associations between the SNPs and phenotypes, where $y$ is the vector of phenotypic observations, $\alpha$ is the vector of SNP effects, $\beta$ is the vector of population structure effects, $\mu$ is the vector of kinship background effects, $e$ is the vector of residual effects, $P$ is the PCA matrix relating $y$ to $\beta$, and $X$ and $K$ are incidence matrices of 1s and 0s relating $y$ to $\alpha$ and $\mu$, respectively (Yu *et al.*, 2006). The observed $P$ values from marker-trait associations were used to display Q–Q plots and Manhattan plots, using R. Kinship analysis was performed using the software TASSEL 5.0 (Bradbury *et al.*, 2007), while detailed information on population structure was imported from the previous analysis described by Qian *et al.* (2014). The critical $P$-value for assessing the significance of SNP-trait associations was calculated separately for CCI based on a false discovery rate (FDR; Benjamini and Hochberg, 1995). An FDR <0.05 was used to identify significant associations for CCI at cut-off values of $-\log_{10}{}^{(P)} = 4$. To simplify the procedure, we used the uniform Bonferroni-corrected thresholds at $\alpha = 1$ as the cut-offs, so, the Bonferroni threshold ($-\log^{(1/24338)} = 4.4$) was used to identify significant associations for GSL.

### Phenotypic correlations to haplotype diversity groups

Significant haplotype blocks were identified using the R package *LDheatmap* (Shin *et al.*, 2006), with haplotypes being defined across regions of homozygous markers LD ($r^2$) > 0.65 between the first and last markers in the block. We use the term haplogroup to refer to groups of individuals carrying a common haplotype across a specific haplotype block. Haplogroups with frequency >0.01 were used for comparative phenotype analysis. A two-sample *t*-test (assuming unequal variances) was used to test for significant phenotypic differences between haplogroups with regard to leaf CCI and seed GSL. Haplotype networks were constructed based on the SNPs number of haplotype region using the program TCS1.21 (Clement *et al.*, 2000).

### Gene content in homoeologous haplotype blocks

A chromosome-scale alignment of syntenic haplotype block regions on homoeologous chromosomes A05 (position 17 873 133–17 986 390 bp, $r^2 = 0.74$) and C05 (36 453 492–37 145 290 bp, $r^2 = 0.68$) was performed using the large-scale genome synteny tool *SYMAP* version 4.2 (Soderlund *et al.*, 2011). All annotated genes within the corresponding haplotype regions were extracted from the *B. napus* Darmor-bzh reference genome v. 4.2 (Chalhoub *et al.*, 2014; accessed from https://genomevolution.org/CoGe/). For verification of the most likely gene functions, we accessed annotations of the closest orthologous *Arabidopsis thaliana* gene by blasting to the Arabidopsis genome database http://www.arabidopsis.org/.

### Population structure analysis

The general population structure of the diversity panel was described previously by Qian *et al.* (2014). Here, we further elucidated the detailed substructure within the largest subpopulations, Q1 and 'mixed', by reanalysing with a random selection of 11 910 polymorphic, single-copy SNPs that have MAF ≥0.05 across these subpopulations. This enabled us to accurately distinguish a small clade of 10 related individuals, with high leaf chlorophyll and low seed glucosinolate content, that form a subcluster (clade) within subpopulation Q1. The R package *SelectionTools* (http://www.uni-giessen.de/population-genetics/downloads) was used to perform a principal component analysis, while an unweighted pair group matrix algorithm tree was calculated by POWERMARKER version 3.25 (Liu and Muse, 2005) and drawn using the software FIGTREE version 1.3.1 (Rambaut, 2009).

## Competing interests

The authors declare no competing interests.

## Authors' contributions

LQ and RS conceived the study, generated the genome-wide SNP data and drafted the manuscript. WQ generated the plant population. LQ and WQ performed field phenotyping, while LQ performed glasshouse trials and was responsible for the data analysis and interpretation. All authors read and approved the final manuscript.

## Acknowledgements

The authors thank Benjamin Wittkop, Petra Degen, Malte Luh, Mechthild Schwarte and Stavros Tzigos for excellent technical assistance. Funding was provided by DFG grants 14/16-2 and 14/17-1 to RS and a doctoral fellowship to LQ from the Chinese Scholarship Council.

## References

Abraham, V. and Deman, J.M. (1986) Hydrogenation of Canola Oil as effected by chlorophyll. *J. Am. Oil Chem. Soc.* **63**, 1185–1188.

Bauer, J., Chen, K., Hiltbunner, A., Wehrli, E., Eugster, M., Schnell, D. and Kessler, F. (2000) The major protein import receptor of plastids is essential for chloroplast biogenesis. *Nature*, **403**, 203–207.

Benjamini, Y. and Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B*, **57**, 289–300.

Bradbury, P.J., Zhang, Z., Kroon, D.E., Casstevens, T.M., Ramdoss, Y. and Buckler, E.S. (2007) TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*, **23**, 2633–2635.

Buntjer, J.B., Sørensen, A.P. and Peleman, J.D. (2005) Haplotype diversity: the link between statistical and biological association. *Trends Plant Sci.* **10**, 466–471.

Chalhoub, B., Denoeud, F., Liu, S., Parkin, I.A.P., Tang, H., Wang, X., Chiquet, J. *et al.* (2014) Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *Science*, **345**, 950–953.

Chen, Y.Z., Pang, Q.Y., He, Y., Zhu, N., Branstrom, I., Yan, X.F. and Chen, S. (2012) Proteomics and metabolomics of *Arabidopsis* responses to perturbation of glucosinolate biosynthesis. *Mol. Plant*, **5**, 1138–1150.

Clement, M., Posada, D. and Crandall, K.A. (2000) TCS: a computer program to estimate gene genealogies. *Mol. Ecol.* **9**, 1657–1660.

Corradin, O., Saiakhova, A., Akhtar-Zaidi, B., Myeroff, L., Willis, J., Cowper-Sallari, R., Lupien, M. *et al.* (2014) Combinatorial effects of multiple enhancer variants in linkage disequilibrium dictate levels of gene expression to confer susceptibility to common traits. *Genome Res.* **24**, 1–13.

Czyczyło-Mysza, I., Tyrka, M., Marcińska, I., Skrzypek, E., Karbarz, M., Dziurka, M., Hura, T. *et al.* (2013) Quantitative trait loci for leaf chlorophyll fluorescence parameters, chlorophyll and carotenoid contents in relation to biomass and yield in bread wheat and their chromosome deletion bin assignments. *Mol. Breed.* **32**, 189–210.

Delmas, F., Sankaranarayanan, S., Deb, S., Widdup, E., Bournonville, C., Bollier, N., Northey, J.G. *et al.* (2013) *ABI3* controls embryo degreening through Mendel's I locus. *Proc. Natl Acad. Sci. USA*, **110**, 3888–3894.

Edwards, D., Batley, J. and Snowdon, R.J. (2013) Accessing complex crop genomes with next-generation sequencing. *Theor. Appl. Genet.* **126**, 1–11.

Fernandez, L., Le Cunff, L., Tello, J., Lacombe, T., Boursiquot, J.M., Fournier-Level, A., Bravo, G. *et al.* (2014) Haplotype diversity of *VvTFL1A* gene and association with cluster traits in grapevine (*V. vinifera*). *BMC Plant Biol.* **14**, 209.

Ferro, M., Brugière, S., Salvi, D., Seigneurin-Berny, D., Court, M., Moyet, L., Ramus, C. *et al.* (2010) AT_CHLORO, a comprehensive chloroplast proteome database with subplastidial localization and curated information on envelope proteins. *Mol. Cell Proteomics*, **9**, 1063–1084.

Field, B., Cardon, G., Traka, M., Botterman, J., Vancanneyt, G. and Mithen, R. (2004) Glucosinolate and Amino Acid Biosynthesis in Arabidopsis. *Plant Physiol.* **135**, 828–839.

Friedt, W. and Snowdon, R.J. (2010) Oilseed rape. In *Handbook of Plant Breeding*, Vol. **4** (Vollmann, J. and Istvan, R., eds), pp. 91–126., Oil Crops Breeding. Berlin/Heidelberg: Springer Verlag.

Friso, G., Giacomelli, L., Ytterberg, A.J., Peltier, J.B., Rudella, A., Sun, Q. and Wijk, K.J. (2004) In-depth analysis of the thylakoid membrane proteome of *Arabidopsis thaliana* chloroplasts: new proteins, new functions, and a plastid proteome database. *Plant Cell*, **16**, 478–499.

Gotoh, E., Kobayashi, Y. and Tsuyama, M. (2010) The post-illumination chlorophyll fluorescence transient indicates the RuBP regeneration limitation of photosynthesis in low light in Arabidopsis. *FEBS Lett.* **14**, 3061–3064.

Harrell, F.E. and Dupont, M.C. (2012) Package 'Hmisc'. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.

Heilmann, I., Mekhedov, S., King, B., Browse, J. and Shanklin, J. (2004) Identification of the Arabidopsis palmitoyl-monogalactosyldiacylglycerol delta7-desaturase gene *FAD5*, and effects of plastidial retargeting of *Arabidopsis* desaturases on the *FAD5* mutant phenotype. *Plant Physiol.* **136**, 4237–4245.

Heyndrickx, K.S. and Vandepoele, K. (2012) Systematic identification of functional plant modules through the integration of complementary data sources. *Plant Physiol.* **159**, 884–901.

Kroymann, J., Textor, S., Tokuhisa, J.G., Falk, K.L., Bartram, S., Gershenzon, J. and Mitchell-Olds, T. (2001) A gene controlling variation in Arabidopsis glucosinolate composition is part of the methionine chain elongation pathway. *Plant Physiol.* **127**, 1077–88.

Li, P., Filiault, D., Box, M.S., Kerdaffrec, E., van Oosterhout, C., Wilczek, A.M., Schmitt, J. *et al.* (2014) Multiple *FLC* haplotypes defined by independent cis-regulatory variation underpin life history diversity in *Arabidopsis thaliana*. *Genes Dev.* **15**, 1635–1640.

Liu, K.J. and Muse, S.V. (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics*, **21**, 2128–2129.

Mace, E.S., Tai, S., Gilding, E.K., Li, Y., Prentils, P.J., Bian, L., Campbell, B.C. *et al.* (2013) Whole-genome sequencing reveals untapped genetic potential in Africa's indigenous cereal crop sorghum. *Nat. Commun.* **4**, 2320.

Murillo, G. and Mehta, R.G. (2001) Cruciferous vegetables and cancer prevention. *Nutr. Cancer*, **41**, 17–28.

Oster, U. and Rüdiger, W. (1998) The G4 gene of *Arabidopsis thaliana* encodes a chlorophyll synthase of etiolated plants. *Bot. Acta*, **110**, 420–423.

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

Rambaut, A. (2009) *Fig Tree version 1.3.1*. [computer program]. [Hattp://tree.bio.ed.ac.uk].

Revelle, W. (2014) *Psych: Procedures for Personality and Psychological Research*. Evanston, Illinois, USA: Northwestern University. http://CRAN.R-project.org/package=p-sych Version=1.4.8.

Shin, J.-H., Blay, S., McNeney, B. and Graham, J. (2006) LDheatmap: an R function for graphical display of pairwise linkage disequilibria between single nucleotide polymorphisms. *J. Stat. Soft.* **16**, 1–9.

Smith, M.D., Rounds, C.M., Wang, F., Chen, K., Afitlhile, M. and Schnell, D.J. (2004) *atToc159* is a selective transit peptide receptor for the import of nucleus-encoded chloroplast proteins. *J. Cell Biol.* **165**, 323–334.

Snowdon, R.J., Abbadi, A., Kox, T., Schmutzer, T. and Leckband, G. (2015) Heterotic haplotype capture: precision breeding for hybrid performance. *Trends Plant Sci.* **20**, 410–413.

Soderlund, C., Bomhoff, M. and Nelson, W. (2011) SyMAP: a turnkey synteny system with application to multiple large duplicated plant sequenced genomes. *Nucleic Acids Res.* **39**, e68.

Stephens, J.C., Schneider, J.A., Tanguay, D.A., Choi, J., Acharya, T., Stanley, S.E., Jiang, R. *et al.* (2001) Haplotype variation and linkage disequilibrium in 313 human genes. *Science*, **293**, 489–493.

Takahashi, H., Kopriva, S., Giordano, M., Saito, K. and Hell, R. (2011) Sulfur assimilation in photosynthetic organisms: molecular functions and

regulations of transporters and assimilatory enzymes. *Annu. Rev. Plant Biol.* **62**, 157–184.

Textor, S., de Kraker, J.W., Hause, B., Gershenzon, J. and Tokuhisa, J.G. (2007) MAM3 catalyzes the formation of all aliphatic glucosinolate chain lengths in Arabidopsis. *Plant Physiol.* **144**, 60–71.

Tautorus, C.L. and Low, N.H. (1993) Chemical aspects of chlorophyll breakdown products and their relevance to canola oil stability. *J. Am. Oil Chem. Soc.* **70**, 843–847.

Thomas, H. and Ougham, H. (2014) The stay-green trait. *J. Exp. Bot.* **65**, 3889–3900.

Tzvetkova-Chevolleau, T., Franck, F., Alawady, A.E., Dall'Osto, L., Carrière, F., Bassi, R., Grimm, B. *et al.* (2007) The light stress-induced protein *ELIP2* is a regulator of chlorophyll synthesis in *Arabidopsis thaliana*. *Plant J.*, **50**, 795–809.

Voss-Fels, K. and Snowdon, R.J. (2015) Understanding and utilizing crop genome diversity via high-resolution genotyping. *Plant Biotechnol. J.* doi: 10.1111/pbi.12456.

Voss-Fels, K., Frisch, M., Qian, L., Kontowski, S., Friedt, W., Gottwald, S. and Snowdon, R.J. (2015) Subgenomic diversity patterns caused by directional selection in bread wheat gene pools. *Plant Genome*, **8**. doi: 10.3835/plantgenome2015.03.0013.

Wang, F., Wang, G., Li, X., Huang, J. and Zheng, J. (2008) Heredity, physiology and mapping of a chlorophyll content gene of rice (*Oryza sativa* L.). *J. Plant Physiol.* **165**, 324–330.

Yang, J. and Zhu, Z.J. (2009) *Glucosinolates in Pakchoi (Brassica Campestris* L. ssp. Chinensis) as Affected by Nutrient Status and Postharvest. PhD thesis, Zhejiang University, China. http://www.dissertationtopic.net/down/1524719.

Yang, Q., Li, Z., Li, W., Ku, L., Wang, C., Ye, J., Li, K *et al.* (2013) CACTA-like transposable element in ZmCCT attenuated photoperiod sensitivity and accelerated the postdomestication spread of maize. *Proc. Natl Acad. Sci. USA*, **42**, 16969–16974.

Yu, J., Pressoir, G., Briggs, W.H., Vroh Bi, I., Yamasaki, M., Doebley, J.F., McMullen, M.D. *et al.* (2006) Aunified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.* **38**, 203–208.

## Supporting information

Additional Supporting information may be found in the online version of this article:

**Figure S1** Manhattan and quantile–quantile plots of MLM showing genome-wide associations for leaf chlorophyll content index in two different environments (glasshouse and field) in 203 Chinese semi-winter rapeseed accessions.

**Figure S2** Genome-wide associations for leaf chlorophyll content index on chromosomes A07, A08, C03 and C08, respectively.

**Figure S3** Boxplots showing phenotypic values for leaf chlorophyll content index in nine and three haplogroups, with frequency >0.01, found in haplotype regions on chromosomes A05 and C05, respectively.

**Figure S4** Comparative analysis of leaf chlorophyll content index (CCI) between groups of accessions carrying combinations of two or three of CCI-associated haplogroups (group A) and accessions carrying only one CCI-associated haplogroup (group B).

**Figure S5** Boxplots showing leaf chlorophyll content index and seed GSL content in the three different subgroups.

**Figure S6** Association mapping for seed GSL on chromosome A02 in 203 Chinese semi-winter rapeseed accessions.

**Table S1** Source, population structure and heritability (leaf chlorophyll content index and seed GSL) in 203 Chinese semi-winter rapeseed accessions

**Table S2** Detailed description of five haplotype regions significantly associated with leaf chlorophyll content index

**Table S3** Comparative analysis of haplogroups related to leaf chlorophyll content index, along with gene content in homologous haplotype regions on chromosomes A05 and C05

**Table S4** Comparative analysis of six haplogroups related to leaf chlorophyll content index, along with gene information in the A02 haplotype region

**Table S5** Comparative analysis of three haplogroups corresponding to leaf chlorophyll content index in BnaA02g20650D gene region

**Table S6** Comparative analysis of five haplogroups corresponding to leaf chlorophyll content index and seed GSL and gene information in chromosome A01 coselection haplotype region

**Table S7** The distribution of haplogroups related to higher leaf chlorophyll content index accessions and comparative analyses of three clades corresponding to leaf CCI and seed GSL in subpopulations 'Q1' and 'mixed', respectively

**Table S8** Comparative analysis of leaf chlorophyll content index between A and B groups

**Table S9** Gene information in the A02 haplotype region

# 5 Discussion

This chapter containing sections that have been previously published by the author in the following publications:

**Lunwen Qian**, Wei Qian and Rod J Snowdon (2014) Sub-genomic selection patterns as a signature of breeding in the allopolyploid *Brassica napus* genome. *BMC Genomics* 15, 1170.

**Lunwen Qian**, Wei Qian and Rod J Snowdon (2016) Haplotype hitchhiking promotes trait coselection in *Brassica napus*. *Plant Biotechnology Journal* 14, 1578-88.

Ying Fu, Kun Lu, **Lunwen Qian**, Jiaqin Mei, Dayong Wei, Xuhui Peng, Xinfu Xu, Jiana Li, Martin Frauen, Felix Dreyer, Rod J Snowdon and Wei Qian (2015) Development of genic cleavage markers in association with seed glucosinolate content in canola. *Theoretical and Applied Genetics* 128, 1029-37.

## 5.1 Development of cleavage markers for breeding of ultra-low GSL content

Cleaved Amplified Polymorphic Sequences (CAPS) are molecular markers with restriction fragment lengths caused by genetic differences between accessions that create or abolish restriction endonuclease recognition sites. These DNA fragment lengths are easily detected in different accessions by agarose gel electrophoresis. CAPS markers have proved to be powerful tools for genotyping in positional or map-based cloning projects (Konieczny and Ausubel 1993; Bello et al. 2014). Chapter 2 of this thesis work identified candidate genes from the GSL biosynthesis pathway within the confidence intervals of QTL that segregate transgressively in low-GSL (doublelow, canola quality) oilseed rape. This represents a first step towards potential cloning of genes having minor but positive effects on this valuable seed quality trait and provides useful CAPS markers for breeding of ultra-low GSL content in oilseed rape and canola meals. Although confirmation of the candidate genes will require additional functional analyses, for breeding purposes, the availability of effective markers tightly linked to interesting QTL can be of considerable assistance both for identification of new useful diversity and for its introgression into elite materials by marker-assisted backcrossing. This is particularly true for traits like seed GSL content, where small-effect loci are often masked by a few large-effect QTL and field testing in multiple environments is necessary for accurate selection of ultra-low phenotypes.

Defatted rapeseed meal is enriched with a high-quality protein with a desirable amino acid composition similar to soybean protein. Reducing GSL content will further improve the nutritive value of meal in oilseed rape and canola meals. Although the low GSL trait in all current *B. napus* cultivars derives from the same founder accession, Bronowski (Krzymanski 1970), with wide variances for the composition of GSL were detected in canola (Li et al. 2005). This suggests the presence of additional genetic factors besides the well described major QTL. Hutcheson et al. (2000) developed an ultra-low GSL spring-type turnip rape (*B. rapa*) with a seed GSL content of only 4.2 µmol/g meal. This material was derived from a cross between members of a low aliphatic GSL *B. rapa* breeding population (BC86-18) and the low indole GSL *B. rapa* breeding population DLY (Hutcheson et al. 2000). We also

detected numerous minor-effects QTL for GSL content in a DH population derived from two lines of oilseed rape, each with low seed GSL. Those findings indicated variant alleles for seed GSL content in canola.

Studies reporting QTL in rapeseed by linkage mapping studies or by association approaches (Uzunova et al. 1995; Howell et al. 2003; Zhao and Meng 2003; Li et al. 2014) were compared with QTL detected in the present study. Four major QTL on chromosomes A09, C02, C07 and C09, which were detected independently in different studies (Uzunova et al. 1995; Howell et al. 2003; Zhao and Meng 2003), were positioned at four common regions at 3.2, 50.0, 39.9 and 2.8 Mb on chromosomes A09, C02, C07 and C09, respectively. These QTL proved to be associated with homologues of the key gene HAG1 (At5g61420) controlling aliphatic glucosinolate biosynthesis (Li et al. 2014). Interestingly, none of the 43 QTL detected in the present study overlapped or were nearby the previously reported major-effect loci. Routine phenotypic selection for low GSL content is generally performed in advanced, homogeneous breeding generations by use of high-throughput near-infrared spectroscopy (NIRS; Wittkop et al. 2009). Although this approach has been highly successful in the breeding of canola-quality *B. napus*, environmentally induced variation among low-GSL materials makes it challenging to implement new loci with additional small effects using conventional selection methods. The markers identified in this work and the plant materials that carry them, therefore, represent important resources to develop ultra-low GSL rapeseed by pyramiding low GSL alleles with both major and minor effects. A previous study by Hasan et al. (2008) confirmed marker–trait associations of gene-linked SSR markers to seed glucosinolate content in genetically diverse *B. napus* germplasm. These included loci associated with homologues of the GSL biosynthesis genes CYP79A2 and MAM1/MAM3, which we also identified within QTL intervals in the present study. Due to the narrow genetic background of the low GSL trait in current canola and oilseed rape cultivars, most of which carry common major-effect QTL from the same origin, we expect the novel minor QTL we describe in this work to be generally effective across a broad range Asian, North American and European breeding materials.

## 5.2 Population structure and genetic diversity

Rapeseed breeding materials in Australia and China have similar origins, with introductions from Europe, Canada and Japan in the mid-20th century and subsequent interchanges of germplasm since then (Chen et al. 2008). Recently, hybrid breeding has received considerable attention, with development of genetically diverse gene pools through recurrent, reciprocal selection of genetic diversity from different *B. napus* ecotypes [9]. Chapter 3 of this thesis showed that the three main population subgroups we observed in our diversity panel may reflect breeding efforts to diversify Chinese semi-winter rapeseed by introgressing genetically distant winter rapeseed (in the case of Q1) and spring canola (in the case of Q3) into different hybrid breeding pools.

Genetic diversity in Chinese *B. napus* has been further improved by introgressions from Asian *B. rapa* (Chen et al. 2008; Qian et al. 2006), the diploid donor of the *B. napus* A-subgenome. According to Liu (1985) and Shiga (1970), more than 50% of *B. napus* cultivars in China and Japan are derived from *B. napus* × *B. rapa* crossings. In the present study, considerably more genetic diversity was observed in the A-subgenome of Chinese rapeseed than the C-subgenome. Interestingly, however, 67 accessions were found with a stronger diversity in the C-subgenome than the A-subgenome. These may have derived from programmes to introgress additional diversity and resistance alleles from C-genome donors.

## 5.3 Comparative analysis of LD and haplotype block between A and C genome

Around 15.17% of the assembled *B. napus* genome could be assigned to haplotype blocks, with large gaps between blocks (data not shown). With an average SNP density of only one SNP per 48.01 Kb, it is difficult to detect very small haplotype blocks (Wall et al. 2003). Studies in maize (Gore et al. 2009) and *Arabidopsis*, (Buckler and Gore 2007) have demonstrated the power of comprehensive genome-wide SNP genotyping arrays for generation of detailed haplotype maps and high-resolution LD analysis. Whole-genome resequencing data provides the ultimate dimension to uncover LD in association with signatures of natural and artificial selection, but so far has been limited to species with

relatively small diploid genomes, like sorghum (Mace et al. 2013). Many of the problems with duplicated SNP loci on the *Brassica* 60k SNP array result from extensive recent genome duplications, which make it quite challenging to design locus-specific SNP assays in many strongly homoeologous regions of the genome. These technical difficulties can cause a reduction in resolution in some genome regions. Nevertheless, use of a high-density SNP array to analyse and compare LD and selection in homoeologous QTL is a unique feature of this study in comparison to previous work in simple diploid species. As a recent allopolyploid *B. napus* thus provides interesting insight into the evolutionary processes of selection in an important crop (Chalhoub et al. 2014).

In Chapter 3 of this thesis, 24,994 unique, polymorphic SNPs were used to perform a preliminary whole-genome analysis of haplotype block structure in *B. napus*. In particular we were able to demonstrate that some *B. napus* chromosomes carry extremely large segments of highly conserved LD, and that this phenomenon is a particular feature of C-subgenome chromosomes. This may indicate increased recombination rates of A-subgenome chromosomes after interspecific hybridisations with *B. rapa*. Boosts of homoeologous recombination among diploid chromosome pairs after interspecific hybridisation were documented in *Brassica* crosses by Leflon et al. (Leflon et al. 2010); this might have caused more rapid LD decay and subsequently shorter-range haplotype blocks in A-subgenome chromosomes in the present materials after hybridisations with *B. rapa*. Although breeders have used interspecific crosses to improve agronomic traits and increase C-subgenome genetic diversity in *B. napus*, it is extremely difficult to obtain viable hybrid seeds from *B. napus* × *B. oleracea* crosses (Downey et al. 1980; Bennett et al. 2008), causing a constraint in the ability to diversify the C-subgenome genetic component. It is thought that *B. napus* arose only in post-neolithic times and from only a small number of independent hybridisation events (Allender et al. 2010), and that the Chinese rapeseed genepool may predominantly represent only one or a few of these events. Hence it is perhaps not surprising that recombination and diversity appear to be considerably lower in the C-subgenome of Chinese oilseed rape. An alternative explanation, which may also partly explain the great overall

difference in LD between the A and C subgenomes, is the considerably greater expansion of transposable elements in the *B. napus* C-subgenome compared to the A-subgenome (Chalhoub et al. 2014), since transposon-rich regions are often observed to be recombination-poor (Gorelick 2003). On the other hand, this fails to explain the great variation in the size of LD and long-range haplotype blocks we observed among different C-subgenome chromosomes. A more simple contributing factor is likely to be strong natural and artificial selection for key adaptation and seed quality traits, where specific variants seem to have been selected during the face of ecogeographical adaptation and human selection, for example for flowering time or quality traits. Strong selection at a locus is expected to reduce diversity and increase LD and haplotype block size in the surrounding region (Rafalski et al. 2004).

In particular, stronger LD and longer-range LD blocks on chromosomes C01, C02, C04 and C09 suggest particularly strong selection of the corresponding region of these chromosomes. According to Liu (1985) rapeseed was introduced into China from Europe in the 1930-1940s, although a later origin within the past few hundred years in Japan is also postulated (Olsson 1960; McNaughton 1976). Guryev et al. (2006) showed that the evolutionary selection process drives conservation of long-range allele combinations, causing chromosome regions to retain a long-range haplotype block structure. Artificial selection can also have a profound effect on LD in crop plants, with selection for key agronomic traits like flowering behaviour, resistances or essential quality parameters causing genetic bottlenecks that lead to extensive conserved haplotype blocks in chromosome regions carrying the responsible gene loci or major QTL for selected traits. Modern double-low quality oilseed rape has undergone selective sweeps for reduction of seed erucic acid and glucosinolate contents, along with flowering time, winter hardiness and vernalisation-related traits. Such selection tends to reduce allele diversity and increase haplotype block structure around the major responsible loci, however detailed studies of LD conservation in oilseed rape breeding pools on a DNA sequence level has only recently become possible since the availability of high-density genome-wide SNP markers (Edwards et al. 2013) in combination with annotated *Brassica*

genome sequences. Here we identified chromosome-scale LD patterns in *B. napus* genome regions carrying important QTL for both a simple, bigenically inherited trait (erucic acid content) and for a complex quantitative trait (seed glucosinolate content).

The observed distance of LD decay in Chinese oilseed rape was 0.85-0.90 Mb, which is higher than maize with 0.5-10 Kb (Yan et al. 2009) and *Arabidopsis* with 50–250 Kb (Hagenblad and Nordborg 2002). This reflects the very recent domestication of *B. napus*, its exclusive use in cultivation, with no known wild forms, and the strong selection bottlenecks associated with cultivation and breeding. Previous studies (e.g. Wang et al. 2014) have shown that the A-subgenome has been successfully improved by closely related species, leading to more rapid decay of LD in the A-subgenome than the C-subgenome. Our results showed that the relatively low overall LD conservation in Chinese rapeseed is caused mainly by a lack of genetic diversity in the C-subgenome. According to Mei et al. (2011), natural *B. napus* has very low genetic diversity compared with its diploid progenitors, therefore intercrossing with the parental species can be an effective way to broaden genetic diversity in rapeseed. To achieve this, it may be necessary to overcome sexual compatibility barriers by using embryo rescue techniques, for example. In recent years considerable progress has been made in introducing novel C-genome donors to European winter oilseed rape, in order to improve genetic diversity particularly for disease resistance (Rygulla et al. 2007a; Rygulla et al. 2007b; Ding et al. 2013 ) or heterosis (Jesske et al. 2013).

## 5.4 Haplotype blocks and extent of LD between homologous QTL region in the A and C subgenomes of *B. napus*

Chapter 3 of this thesis showed that conserved haplotype blocks with strong LD spanning major homoeologous QTL for seed GLS (chromosomes A02/C02 and A09/C09) and erucic acid (A08/C03) reflect the strong selection bottlenecks for these traits. On the other hand, the introgression of exotic A-subgenome diversity from *B. rapa* has apparently led to shorter-range haplotype blocks and lower LD in A-subgenome than C-subgenome QTL. Chinese *B. napus* originated from Europe (Liu 1985), being introduced to China in the 1930-1940s

replacing the traditional oilseed crop *B. rapa*. Local adaptation to the new ecogeographical environment and diversification of breeding pools was achieved by introgressing local populations of the wild progenitors and closely related species, particularly *B. rapa* (Cornille et al. 2012; Luo et al. 2007; Myles et al. 2011). Our results show that this process resulted in substantial decay of LD surrounding important A-subgenome QTL, whereas longer-range haplotype blocks and higher LD are retained around C-subgenome QTL regions. Importantly, conserved haplotype blocks in C-subgenome QTL tend to retain multiple genes related to relevant biosynthetic processes, which can potentially cause linkage drag and ultimately slows down breeding progress for the trait of interest.

Various forces have potentially contributed to haplotype conservation in C-subgenome QTL in *B. napus*, including genetic bottlenecks from artificial or natural selection or a simple lack of recombination and sequence diversity. We found rates of sequence polymorphism to be generally lower in C-subgeneome QTL regions than their corresponding A-subgenome homoeologues, suggesting that the former may be the dominant mechanism. On the other hand a suppression of recombination, due to the increased density of transposable elements in the C-subgenome (Chalhoub et al. 2014), cannot be ruled out. Detailed haplotype block analysis of important QTL can help in the precise mapping of important genomic regions and location of favorable alleles. In association with genomic sequence data it can also help to more precisely predict quantitative trait-related genes (QTG) in QTL regions using targeted association mapping with high-density markers.

Strongly conserved LD observed across the QTL on chromosomes C02 and C09 in this study was found to be associated with a large number of functionally related genes in close genetic linkage. The corresponding homoeologous QTL on chromosomes A02 and A09 each contained fewer genes annotated to the QTL function. This result demonstrates the important role of gene loss during or after allopolyploidisation in natural and/or artificial selection of key traits like GSL content (Chalhoub et al. 2014). Natural evolution results in a positive and balancing selection within the genome, whereas artificial selection can lead to partial

separation of phenotypic traits. According to Harper et al. (2012), deletions affecting homologues of the GSL biosynthesis gene Bna.HAG1/myb28 resulted in selective sweeps affecting the QTL for GSL on A09 and C02. Both in this case, and in the case of homoeologous QTL for erucic acid content on chromosomes A08 and C03, it was demonstrated that selective sweeps can also incorporate additional, functionally-related genes for which alleles in strong LD may have either a positive or negative influence (linkage drag) on target traits (e.g. GSL content, fatty acid composition or oil content). Detailed analysis of LD structure and signatures of selection in important QTLs can guide breeders towards a knowledge-based crop improvement through genome-based introgression of useful genetic diversity.

## 5.5 Association mapping for leaf CCI and seed GSL in 203 Chinese semi-winter rapeseed accessions

Chlorophyll is a green photosynthetic pigment with which plants generate energy in the presence of light. Leaf chlorophyll content relates to photosynthetic capacity and is thus one of the important physiological traits influencing crop yield (Czyczyło-Mysza et al. 2013; Wang et al. 2008). On the other hand, the presence of chlorophyll in mature seeds can be an undesirable trait that can affect seed maturation, seed oil quality, meal quality and germination (Delmas et al., 2013). Chlorophyll pigments remaining in processed vegetable oils are also associated with increased oxidation causing rancidity (Tautorus and Low 1993) and difficulties in hydrogenation (Abraham and Deman 1986). Recent studies suggested a previously unknown relationship between chlorophyll content and metabolism of glucosinolates, a class of sulphur-containing secondary metabolites expressed in vegetative and generative tissues throughout almost all plants of the *Brassicales*. For example, a total of 11 light-harvesting chlorophyll (LHC) complex proteins (including eight LHCB and three LHCA proteins) were identified to be down-regulated in RNAi lines that suppressed molecular networks controlling glucosinolate metabolism in *Arabidopsis* (Chen et al. 2012). Yang et al. (2009) reported an upward trend in total glucosinolate content in cabbage plants grown under low light intensity. These results suggest potential genetic and/or metabolic

associations between chlorophyll content and glucosinolate metabolism in *Brassica* crops. Although a molecular basis for such a pleiotropic relationship might be associated with the role of plastids in sulphate reduction and cysteine/methionine synthesis, providing sulphuric amino acids that serve as precursors for glucosinolate synthesis (Takahashi et al. 2011). Such a relationship is potentially relevant both evolutionarily and agronomically. Leaf glucosinolates have demonstrated positive nutritional value in vegetable *Brassica* crops (Murillo and Mehta 2001) and play an important role, either antagonistically or mutualistically, in plant interactions with insect pests. On the other hand, glucosinolates are highly undesirable in *Brassica* oilseed meals fed to livestock (Friedt and Snowdon 2010). The huge global importance of oilseed rape and canola (*Brassica napus*), the world's second most important oilseed crop, imparts enormous agroecomonic relevance on these compounds.

GWAS examines co-transmission of phenotypes with genetic markers, normally based on linkage disequilibrium (LD) analysis in genetically diverse populations using panels of markers spanning the entire genome at high density. Besides providing high mapping resolution by incorporating historical recombination events, LD analyses can also provide important insight into the history of both natural and artificial selection (breeding) and give valuable guidance to breeders seeking to diversify crop gene pools.

Selection for specific agronomic traits during plant domestication and breeding has strong influences on the genetic diversity and population structure within available gene pools for further crop improvement. High-density genotyping tools today provide a means for unprecedented insight into patterns of diversity associated with breeding in major crop species (Qian et al., 2014, Snowdon et al., 2015, Voss-Fels et al., 2015), enabling identification of loci under strong selection and potentially allowing islands of depleted diversity to be addressed by targeted, marker-assisted introgressions without compromising desirable adaptation, yield or quality traits (Voss-Fels and Snowdon 2015).

Besides their negative influence on genetic diversity, strong signatures of selection associated with key traits can also cause co-selection of loci with undesirable effects,

resulting in what is commonly known as linkage drag. This thesis presents an interesting example in oilseed rape for co-selection of increased leaf chlorophyll content along with reduced seed glucosinolate content, two traits with no immediately obvious biological relationship. It could be demonstrated that introgressions between ecogeographically distinct gene pools resulted in indirect selection of plants with elevated levels of leaf chlorophyll, due to co-selection of beneficial haplotypes at four independent haplotype blocks. Within one of these haplotype blocks, a gene involved in chlorophyll synthesis showed almost complete LD with a gene associated to reduce seed glucosinolate content. The relationship of the observed haplotypes to the respective phenotypic behaviour suggests that introgression of this locus during breeding, to reduce seed glucosinolate levels in nutritionally valuable varieties, causes hitchhiking selection resulting in increased chlorophyll content. Elevated chlorophyll may result in improved photosynthetic performance. While this does not necessarily relate *per se* to improved agronomic performance, stay-green traits associated with increased chlorophyll production, or suppression of chlorophyll degradation, are implicated in improved adaptation of crops to abiotic stresses like water or nutrient deficiency (Thomas and Ougham 2011).

A total of nine haplotype blocks were found to harbour significant associations ($r2 \geqslant 0.65$) with leaf CCI. Within these regions eighteen genes weew found that are implicated in chlopophyll synthesis or catabolism, respectively. These included six ortholgoues of *FAD5*, which was found in *Arabidopsis* to influence chlorophyll biosynthesis (Heyndrickx et al. 2012) and restore leaf chlorophyll content (Heilmann et al., 2004), along with two orthologues of the gene PIFI, which has an indirect effect on photosynthesis (Gotoh et al. 2010). The observation of homoeologous, trait-associated haplotypes, carrying multiple duplicated genes, provides a further example for gene amplification by polyploidisation. In the complex allopolyploid crop *B. napus* this phenomenon has been shown to influence numerous agronomically important traits (Chalhoub et al. 2014).

Additional chlorophyll-associated haplotypes carried *B. napus* orthologues of the genes *ELIP2* and *CHLG*, which in *Arabidopsis* are directly involved in chlorophyll biosynthesis (Tzvetkova-Chevolleau et al., 2007; Oster and Rudiger 1998), and the chloroplast preprotein transporter protein gene *TOC159*. The latter encodes an essential component of the chloroplast assembly mechanism, which act as a selective import receptor for pre-proteins required in chloroplast development (Smith et al. 2004). Collectively, the genetic associations that were found, involving multiple loci carrying these very different contributors to the photosynthesis apparatus, suggest a broad genetic variation for factors involved in photosynthesis in crop plants. Given the key role of photosynthetic activity in ecogeographical and stress adaptation it is interesting that multiple loci with positive additive influences on chlorophyll content are co-selected in introgressions between different breeding pools in this major crop.

It is likely that the expression of many genes is jointly controlled by the actions of multiple regulatory alleles. Despite this, GWAS studies rarely consider the possibility that, at a given locus, multiple genes may impact a phenotype by interactions between more than one regulatory allele in across a gene-haplotype region (Corradin et al. 2014). Particularly for complex, multigenic traits like chlorophyll content, a spatial proximity of interacting genes can potentially facilitate coordinated expression in certain tissues, developmental timepoints or in response to environmental stimuli. This may help explain why multi-genic haplogroups associate more strongly with CCI phenotypes than single-gene haplotypes. The results of the present study underline the benefit of combining haplotype diversity analysis with GWAS studies to dissect additive effects of quantitative trait loci in crops and understand their underlying biology (Buntjer et al., 2005). Detailed investigations of trait relationships at the local haplotype level, using high-density SNP markers in large populations, also provide plant breeders with a mean to distinguish between genetic and pleiotropic trait correlations.

## 5.6 Conclusions

In this thesis, sequence polymorphisms between the corresponding coding regions of parental rapeseed lines were used to develop cleaved amplified polymorphic site markers for two QTL linked genes associated to two genes contributing to ultra-low GSL content, IPMDH1 and APR3. Furthermore, high-resolution genome data obtained using a 60k SNP array was implemented for diversity analysis in major homoloeogous QTL of important quality traits (GSL and erucic acid content) and for fine mapping for co-selection analysis of leaf CCI and seed GSL content in 203 Chinese semi-winter rapeseed inbred lines.

The following general conclusions arose from the work:

- Orthologues of *Arabidopsis* genes involved in seed GSL metabolism were identified within QTL confidence intervals, and two CAPS markers were developed to facilitate breeding for ultra-low glucosinolates in rapeseed.

- Three main population subgroups observed in the diversity panel may reflect breeding efforts to diversify Chinese semi-winter rapeseed by introgressing genetically distant winter rapeseed (in the case of Q1) and spring canola (in the case of Q3) into different hybrid breeding pools.

- 67 accessions were found with a stronger diversity in the C-subgenome than the A-subgenome. These may derive from programmes to introgress additional diversity and resistance alleles from C-genome donors.

- A total of 24,994 unique, polymorphic SNPs, used to perform a whole-genome analysis of haplotype block structure, showed that some *B. napus* chromosomes carry extremely large segments of highly conserved LD, and that this phenomenon is a particular feature of C-subgenome chromosomes.

- Comparative analysis of conserved haplotype blocks with strong LD spanning major homoeologous QTL for seed GSL (chromosomes A02/C02 and A09/C09) and erucic acid (A08/C03) showed that longer-range haplotype blocks and higher LD are

retained around C-subgenome QTL regions. Importantly, conserved haplotype blocks in C-subgenome QTL tend to retain multiple genes related to relevant biosynthetic processes, which can potentially cause linkage drag that slows breeding progress for the trait of interest.

- Analysis of homologous haplotypes on chromosomes A05 and C05 revealed multiple orthologous copies of the chloroplast membrane protein gene *PALMITOYL-MONOGALACTOSYLDIACYLGLYCEROL DELTA-7 DESATURASE* (*FAD5*) associated with leaf chlorophyll content. Another conserved haplotype block, on chromosome A02, contained a number of genes related to chlorophyll synthesis or degradation.

- A conserved haplotype block, on chromosome A02, contained a number of genes related to chlorophyll synthesis or degradation, whereby the multigene haplogroup was found to have a significantly greater contribution to variation for chlorophyll content than haplotypes for any single gene, suggesting positive effects of additive locus accumulation.

- Conserved haplotypes on chromosome A01 were observed to show opposite effects on leaf chlorophyll content and seed glucosinolate content.

- Introgression from winter rapeseed genetic diversity contributed to higher chlorophyll content.

# 6 Summary

*Brassica napus* (rapeseed, oilseed rape, canola) is the second most important oilseed crop in the world after soybean. Worldwide production of rapeseed was 61 million metric tons (MT) in 2011. *Brassica napus* (2n=38, AACC) is derived from interspecific hybridisation events between *B. rapa* (2n=20, AA) and *B. oleracea* (2n=18, CC), and was probably selected as an oilseed crop only 300–400 years ago.

Oilseed producing *B. napus* has only achieved economic importance in the past forty years following an intensive breeding programme to decrease nutritionally undesirable components of the oil and meal, and to increase yields. In some areas, rapeseed, which contains more than 40 % oil, has become more profitable than soybeans, which contain about 18 % oil. Attention of breeders was initially focused on reducing levels of erucic acid in the oil and glucosinolate (GSL) content in the meal. Four major quantitative trait locus (QTL) regions on chromosomes A09, C02, C07 and C09 relate to reduction in seed GSL content which has been achieved in worldwide rapeseed breeding during the past three decades. On the other hand, further reductions in seed GSL require the combination of these main effect loci with additional QTL that have less prominent effects. Such loci are often overlooked in genetic mapping studies that use parents carrying main-effect loci, making it difficult to implement them into breeding programmes. In this thesis, 43 QTL for seed GSL content were dissected over multiple environments in a doubled haploid (DH) population derived from two low-GSL parents. Two cleaved amplified polymorphic site markers were developed from within the confidence intervals of the detected QTL regions. These markers ware specific for *B. napus* orthologues of the glucosinolate biosynthesis genes IPMDH1 and APR3 and segregation of the marker alleles are explaining 3–8 % of the phenotypic variation for seed GSL. The use of these markers in marker–assisted breeding will facilitate breeding for ultra-low seed GSL content in canola.

Rapeseed is grown in different geographical regions of the world. It is adapted to different environments by modification of flowering time. Different *B. napus* gene populations have

undergone strict selection for flowering-related traits, including vernalisation requirement, winter survival and photoperiod-dependant flowering. Also they have undergone strict selection for essential seed quality traits (primarily low erucic acid and GSL contents). Chinese semi-winter rapeseed breeding has extensively used diploid Brassica species, particularly *B. rapa*, to enrich the genetic potential of the local gene pool. High-resolution genome analysis technologies provide an unprecedented level of insight into structural diversity across crop genomes. In this thesis, a high-density 60k Single Nucleotide Polymorphism (SNP) array is used to analyze linkage disequilibrium (LD) and haplotype structures in homologues QTL regions for erucic acid and GSL contents within the A- and C-subgenomes of 203 Chinese semi-winter rapeseed inbred lines. The result shows a strong selection of large chromosome regions associated with these important rapeseed seed quality traits conferred by C-subgenome QTL. This implies that an increase in genetic diversity and recombination within the C-genome is particularly important for breeding. The resolution of genome-wide association studies (GWAS) is also expected to vary greatly across different genomic regions.

GWAS examines co-transmission of phenotypes with genetic markers that provide a powerful insight into correlations of haplotype diversity to phenotype groups. Analysis of gene content in conserved haplotype blocks can further provide insight into co-selection for different quantitatively inherited traits. In this thesis a genome-wide analysis of haplotypes associated with the important physiological and agronomic traits leaf chlorophyll and seed GSL content was performed. Analysis of homologous haplotypes on chromosomes A05 and C05 revealed multiple orthologous copies of the chloroplast membrane protein gene PALMITOYL-MONOGALACTOSYLDIACYLGLYCEROL DELTA-7 DESATURASE (FAD5) associated with leaf chlorophyll content. Conserved haplotype blocks on chromosome A02 contained a number of genes related to chlorophyll synthesis or degradation. The multigene haplogroup had a significantly greater contribution to variation for leaf chlorophyll content than the haplogroups for any single gene. Furthermore, conserved haplotypes on chromosome A01 were observed to show opposite effects on leaf chlorophyll content and

seed GSL content. Haplotype network analysis across these four haplotype regions showed introgression from winter rapeseed contributing to genetic diversity and to higher chlorophyll content.

# 7 Zusammenfassung

*Brassica napus* (Raps) ist die zweitwichtigste Ölsaat liefernde Pflanze in der Welt nach Soja. Im Jahr 2011 betrug die weltweite Produktion von Ölssaat aus Raps 61 Millionen Tonnen. *Brassica napus* (2n = 38 Chromosomen, AACC) ist durch natürliche interszpezifische Hybridisierung aus den beiden Arten *B. rapa* (2n = 20, AA) und *B. oleracea* (2n = 18, CC) entstanden. Raps ist wahrscheinlich vor nur 300-400 Jahren vom Menschen als Ölpflanze selektiert worden.

Große wirtschaftliche Bedeutung hat Raps jedoch erst erlangt, nachdem in den letzten vierzig Jahren mittels intensiver Züchtung ernährungsphysiologisch unerwünschte Bestandteile des Öls und des Rapsschrots reduziert und die Erträge stark gesteigert wurden. In einigen Anbaugebieten ist mittlerweile die Produktion von Rapssamen, die mehr als 40 % Öl enthalten, profitabler geworden als die Produktion von Sojabohnen, die 18 % Öl enthalten. Züchtungsziele wurden zunächst auf die Reduktion des Erucasäure-Gehalts im Öl und des Glukosinolat-(GSL)-Gehalts im Schrot ausgerichtet. In den vergangenen drei Jahrzehnten wurde in der weltweiten Rapszüchtung eine starke Reduktion der Samen-GSL-Gehalte erreicht. Hieran sind vier Quantitative Trait Locus (QTL)-Regionen mit Haupt-Effekten auf den Chromosomen A09, C02, C07 und C09 beteiligt. Eine weitere Senkung der Samen-GSL-Gehalte erfordert jedoch die Kombination dieser Haupteffekt-Loci mit zusätzlichen QTL, die geringere Effekte aufweisen. Solche Loci werden in genetischen Kartierungsuntersuchungen, in denen beide Eltern Haupteffekt-QTL tragen, oft übersehen und dies macht es schwierig, sie erfolgreich in Zuchtprogramme einzubringen. In dieser Arbeit wurden 43 QTL für Samen-GSL-Gehalte in mehreren Umwelten in einer doppelt haploiden (DH) Kartierungspopulation detektiert, die durch Kreuzung zweier Eltern mit niedrig GSL-Samen-Gehalten produziert wurden. Zwei polymorphe Marker wurden entwickelt, die spezifisch sind für zwei orthologe, an der Glukosionlat-Biosynthese beteiligten Gene (IPMDH1 und APR3). Die indentifizierten Markerallele sind eng mit Glukosinolat-Gehalt gekoppelt und erklären 3-8% der phänotypischen Variation für den Samen-

Glukosinolat-Gehalt. Die Verwendung dieser Marker wird die Züchung von Niedrigen Samen-GSL-Gehalten erleichtern.

Raps wird in verschiedenen Regionen der Welt angebaut. Raps ist an verschiedene Umwelten duch Modifikation des Blühzeipunktes und -dauer angepasst. Verschiedene *B. napus*-Populationen sind einer strengen Selektion für Blüten-bezogene Eigenschaften unterzogen, einschließlich der Anforderung an die Vernalisation, die Winterhärte, und den Photoperiode-abhängigen Blühzeitpunkt. Eine strenge Selektion erfolgt auch für bedeutsame Saatgut-Qualitätsmerkmale (in erster Linie Erucasäure-Armut und GSL-Gehalt). Die Züchtung von chinesischen Semi-Winter-Typen verwendet in großem Ausmaß auch diploide *Brassica*-Arten, insbesondere *B. rapa*, um das genetische Potenzial des lokalen *B. napus* Genpools zu bereichern. Hochauflösende Genomanalyse-Technologien bieten ein beispielloses Maß an Einsicht in strukturelle Vielfalt in Pflanzengenomen. In dieser Arbeit wurde ein hochdichter 60k Single Nucleotide Polymorphism-(SNP) Array verwendet, um Kopplungs-Ungleichgewicht (Linkage Disequilibrium, LD) und Haplotyp-Struktur in homeologen QTL Regionen für Erucasäure- und GSL-Gehalte im A- und C-Subgenom von 203 chinesischen semi-Winterraps-Inzuchtlinien zu analysieren. Das Ergebnis zeigte, dass eine große Anzahl von großen Chromosomenregionen im C-Subgenom mit wichtigen Rapssamen-Qualitätsmerkmalen assoziiert ist (QTL für GSL und Erucasäure-Gehalt). Dies bedeutet, dass eine Erhöhung der genetischen Vielfalt und Rekombination innerhalb des C-Genoms für die Züchtung von besonderer Bedeutung ist. Es kann daher auch erwartet werden, dass die Auflösung von genomweiten Assoziationsstudien (GWAS) stark in den verschiedenen genomischen Regionen des A- und C-Genoms variieren wird.

GWAS untersucht die gemeinsame Vererbung von phänotypischen Merkmalen mit genetischen Markern und erlaubt einen Einblick in die Zusammenhänge von Haplotyp-Diversitäten mit Phänotyp-Gruppen. Die Analyse von Gen-Gehalten in konservierten Haplotyp-Blöcken kann weitere detailierte Einblicke in die Ko-Selektion verschiedener quantitativ vererbter Merkmale ermöglichen. In dieser Arbeit wurde eine genomweite Analyse

der Assoziation von Haplotypen mit dem wichtigen physiologischen und agronomischen Merkmale Blatt-Chlorophyll- und Samen-Glukosinolat-Gehalt durchgeführt. Die Analyse der homologen Haplotypen auf den Chromosomen A05 und C05 detektiert mehrere orthologe Kopien des Chloroplasten-Membran-Protein-Gens PALMITOYL-MONOGALACTOSYLDIACYLGLYCEROL DELTA-7 DESATURASE (FAD5), das mit Blattchlorophyll-Gehalt assoziiert ist. Konservierte Haplotyp-Blöcke auf dem Chromosom A02 enthalten eine Reihe von Genen, die an der Chlorophyll-Synthese oder am -Abbau beteiligt sind. Hierbei zeigen multigene Haplogruppen einen deutlich größeren Beitrag zur Variation für Blattchlorophyll-Gehalt als die Haplogruppen für ein einzelnes Gen allein. Darüber hinaus zeigen Haplotypen auf Chromosom A01 entgegengesetzte Auswirkungen auf den Blattchlorophyll- und den Samen-Glukosinolat-Gehalt. Eine Haplotyp-Netzwerkanalyse für diese vier Haplotyp-Regionen zeigte, dass eine Einkreuzung von Winterraps- in Semi-Winterraps-Typen zur Erhöhung der genetischen Vielfalt und des Blatt-Chlorophyll-Gehalts beigetragen hat.

# 8 References

Allender CJ and King GJ (2010) Origins of the amphiploid species *Brassica napus* investigated by chloroplast and nuclear molecular markers. *BMC Plant Biology* 10, 54.

Abraham V and Deman JM (1986) Hydrogenation of Canola Oil as affected by Chlorophyll. *Journal of the American Oil Chemists Society* 63, 1185-1188.

Atwell S et al. (2010) Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature* 465, 627–631.

Becker HC, Engqvist GM and Karlsson B (1995) Comparison of rapeseed cultivars and resynthesized lines based on allozyme and RFLP markers. *Theoretical Applied Genetics* 91, 62–67.

Bekele WA, Wieckhorst S, Friedt W and Snowdon RJ (2013) High-throughput genomics in sorghum: from whole-genome resequencing to a SNP screening array. *Plant Biotechnology Journal* 11, 1112-1125.

Bennett RA, Thiagarajah MR, King JR and Rahman MH (2008) Interspecific cross of *Brassica oleracea* var*. alboglabra* and *B.napus*: effects of growth condition and silique age on the efficiency of hybrid production, and inheritance of erucic acid in the self-pollinated backcross generation. *Euphytica* 164, 593-601.

Bouchet S, Pot D, Deu M, Rami JF, Billot C, Perrier X, Rivallan R, Gardes L, Xia L, Wenzl P, Kilian A and Glaszmann JC (2012) Genetic structure, linkage disequilibrium and signature of selection in Sorghum: lessons from physically anchored DArT markers. *PLoS One* 7, e33470.

Buckler E and Gore M (2007) An *Arabidopsis* haplotype map takes root. *Nature Genetics* 39, 1056-1057.

Buntjer JB, Sørensen AP and Peleman JD (2005) Haplotype diversity: the link between statistical and biological association. *Trends in Plant Science* 10, 466–471.

Bus A, Korber N, Snowdon RJ and Stich B (2011) Patterns of molecular variation in a species wide germplasm set of *Brassica napus*. *Theoretical Applied Genetics* 123, 1413-1423.

Chagné D, Batley J, Edwards D and Forster JW (2007) Single Nucleotide Polymorphisms genotyping in plants, in: NC Oraguzie, et al. (Eds.), Association mapping in plants, Springer, New York. pp 77-94.

Chalhoub B, Denoeud F, Liu S, Parkin IAP, Tang H, Wang X, Chiquet J, Belcram H, Tong C, Samans B, Corréa M, Da Silva C, Just J, Falentin C, Koh CS, Le Clainche I, Bernard M, Bento P, Noel B, Labadie K, Alberti A, Charles M, Arnaud D, Guo H, Daviaud C, Alamery S, Jabbari K, Zhao M, Edger PP, Chelaifa H, Tack D, Lassalle G, Mestiri I, Schnel N, Le Paslier M-C, Fan G, Renault V, Bayer PE, Golicz AA, Manoli S, Lee T-H, Thi VHD, Chalabi S, Hu Q, Fan C, Tollenaere R, Lu Y, Battail C, Shen J, Sidebottom CHD, Wang X, Canaguier A, Chauveau A, Bérard A, Deniot G, Guan M, Liu Z, Sun F, Lim YP, Lyons E, Town CD, Bancroft I, Wang X, Meng J, Ma J, Pires JC, King GJ, Brunel D, Delourme R, Renard M, Aury J-M, Adams KL, Batley J, Snowdon RJ, Tost J, Edwards D, Zhou Y, Hua

W, Sharpe AG, Paterson AH, Guan C and Wincker P (2014) Early allopolyploid evolution in the post-neolithic *Brassica napus* oilseed genome. *Science* 345, 950-953.

Chen S, Nelson MN, Ghamkhar K, Fu T and Cowling WA (2008) Divergent patterns of allelic diversity from similar origins: the case of oilseed rape (*Brassica napus* L.) in China and Australia. *Genome* 51, 1-10.

Chen YZ, Pang QY, He Y, Zhu N, Branstrom I, Yan XF and Chen S (2012) Proteomics and metabolomics of *Arabidopsis* responses to perturbation of glucosinolate biosynthesis. *Molecular plant* **5**, 1138-1150.

Ching A, Caldwell KS, Jung M, Dolan M, Smith OS, Tingey S, Morgante M and Rafalski AJ (2002) SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. *BMC Genetics* 3, 19.

Cook JP, McMullen MD, Holland JB, Tian F, Bradbury P, Ross-Ibarra J, Buckler ES and Flint-Garcia SA (2012) Genetic architecture of maize kernel composition in the nested association mapping and inbred association panels. *Plant Physiology* 158, 824–834.

Cornille A, Gladieux P, Smulders MJM, Roldan-Ruiz I, Laurens F, Le Cam B, Nersesyan A, Clavel J, Olonova M, Feugey L, Gabrielyan I, Zhang XG, Tenaillon MI and Giraud T (2012) New insight into the history of domesticated apple: secondary contribution of the European wild apple to the genome of cultivated varieties. *Plos Genetics* 8, e1002703.

Corradin O, Saiakhova A, Akhtar-Zaidi B, Myeroff L, Willis J, Cowper-Sal-lari R, Lupien M, Marlowitz S and Scacheri PC (2014) Combinatorial effects of multiple enhancer variants in linkage disequilibrium dictate levels of gene expression to confer susceptibility to common traits. *Genome Research* 24, 1–13.

Cruciani F, Santolamazza P, Shen PD, Macaulay V, Moral P, Olckers A, Modiano D, Holmes S, Destro-Bisol G, Coia V, Wallace DC, Oefner PJ, Torroni A, Cavalli-Sforza LL, Scozzari R and Underhill PA (2002) A back migration from Asia to sub-Saharan Africa is supported by high-resolution analysis of human Y-chromosome haplotypes. *The American Journal of Human Genetics* 70, 1197-1214.

Czyczyło-Mysza I, Tyrka M, Marcińska I, Skrzypek E, Karbarz M, Dziurka M, Hura T, Dziurka K and Quarrie SA (2013) Quantitative trait loci for leaf chlorophyll fluorescence parameters, chlorophyll and carotenoid contents in relation to biomass and yield in bread wheat and their chromosome deletion bin assignments. *Molecular Breeding* 32, 189-210.

Dalton-Morgan J, Hayward A, Alamery S, Tollenaere R, Mason AS, Campbell E, Patel D, Lorenc MT, Yi B, Long Y, Meng J, Raman R, Raman H, Lawley C, Edwards D and Batley J (2014) A high-throughput SNP array in the amphidiploid species *Brassica napus* shows diversity in resistance genes. *Functional & Integrative Genomics* 14, 643-55.

Delmas F, Sankaranarayanan S, Deb S, Widdup E, Bournonville C, Bollier N, Northey JG, McCourt P and Samuel MA (2013) ABI3 controls embryo degreening through Mendel's I locus. *Proceedings of the National Academy of Sciences* 110, 3888-3894.

Derycke G, Mabon N, Mandiki SN Bister JL, Wathelet JP, Marlier M and Paquay R (1999) Chemical changes and influence of rapeseed antinutrtional factors on lamb physiology

and performance. Part 1 Animal performance and thyroid physiology. *Animal Feed Science and Technology* 81, 81-91.

Ding Y, Mei J, Li Q, Liu Y, Wan H, Wang L, Becker HC and Qian W (2013) Improvement of *Sclerotinia sclerotiorum* resistance in *Brassica napus* by using *B. oleracea*. *Genetic Resources and Crop Evolution* 60, 1615-1619.

Downey RK, Klassen AL and Stringam GR (1980) Rapeseed and Mustard. Pages 495-509 in W. R . Fehr and H. Hadley, eds. Hybridization of crop plants. ASA, CSA, Madison, WI.

Ecke W, Clemens R, Honsdorf N and Becker HC (2010) Extent and structure of linkage disequilibrium in canola quality winter rapeseed (*Brassica napus* L.). *Theoretical and Applied Genetics* 120, 921-931.

Edwards D, Batley J and Snowdon RJ (2013) Accessing complex crop genomes with next-generation sequencing. *Theoretical and Applied Genetics* 126, 1-11.

Ersoz ES, Yu J and Buckler ES (2007) Applications of linkage disequilibrium and association mapping in crop plants. Genomics-assisted crop improvement Springer, Dordrecht, pp 97-120.

Fernandez L, Le Cunff L, Tello J, Lacombe T, Boursiquot JM, Fournier-Level A, Bravo G, Lalet S, Torregrosa L, This P and Martinez-Zapater JM (2014) Haplotype diversity of VvTFL1A gene and association with cluster traits in grapevine (V*. vinifera*). *BMC Plant Biology* 14, 209.

Flint-Garcia SA, Thornsberry JM and Buckler ESth (2003) Structure of linkage disequilibrium in plants. *Annual Review of Plant Biology* 54, 357-374.

Friedt W and Snowdon R (2009) Oilseed Rape. In: Handbook of Plant Breeding, vol. 4. Oil Crops, Vollman J and Rajcan J (eds.), Springer-Verlag, Dordrecht, Heidelberg, London, New York 91-126.

Friedt W and Snowdon RJ (2010) Oilseed rape. In: Vollmann J, Istvan R (eds) Handbook of Plant Breeding, Vol. 4: Oil crops breeding. Springer Verlag, Berlin/Heidelberg, pp 91-126.

Gómez-Campo C (1999) Biology of Brassica coenospecies. Elsevier Press, Netherlands, pp 33–58.

Gore MA, Chia JM, Elshire RJ, Sun Q, Ersoz ES, Hurwitz BL, Peiffer JA, McMullen MD, Grills GS, Ross-Ibarra J, Ware DH and Buckler ES (2009) A first-generation haplotype map of maize. *Science* 326, 1115-1117.

Gorelick R (2003) Transposable elements suppress recombination in all meiotic eukaryotes, including automictic ancient asexuals: a reply to Schön and Martens. *Journal of Natural History* 37, 903-909.

Gotoh E, Kobayashi Y and Tsuyama M (2010) The post-illumination chlorophyll fluorescence transient indicates the RuBP regeneration limitation of photosynthesis in low light in Arabidopsis. *FEBS Letter* 14, 3061-4.

Guryev V, Smits BMG, van de Belt J, Verheul M, Hubner N and Cuppen E (2006) Haplotype block structure is conserved across mammals. *Plos Genetics* 2, 1111-1118.

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

Halkier BA and Gershenzon J (2006) Biology and biochemistry of glucosinolates. *Annual Review Plant Biology* 57, 303-333.

Hamblin MT, Salas Fernandez MG, Casa AM, Mitchell SE, Paterson AH and Kresovich S (2005) Equilibrium processes cannot explain high levels of short- and medium-range linkage disequilibrium in the domesticated grass Sorghum bicolor. *Genetics* 171, 1247-1256.

Harper AL, Trick M, Higgins J, Fraser F, Clissold L, Wells R, HattoriC, Werner P and Bancroft I (2012) Associative transcriptomics of traits in the polyploid crop species *B. napus*. *Nature Biotechnology* 30, 798–802.

Hasan M, Friedt W, Pons-Kühnemann J, Freitag NM, Link K and Snowdon RJ (2008) Association of gene-linked SSR markers to seed glucosinolate content in oilseed rape (*Brassica napus* ssp. *napus*). *Theoretical and Applied Genetics* 116, 1035–1049.

Hasan M, Seyis F, Badani AG, Pons-Kuhnemann J, Lühs W, Friedt W and Snowdon RJ (2006) Surveying genetic diversity in the *Brassica napus* gene pool using SSR markers. *Genetic Resources and Crop Evoluation* 53, 793-802.

Heilmann I, Mekhedov S, King B, Browse J and Shanklin J (2004) Identification of the Arabidopsis palmitoyl-monogalactosyldiacylglycerol delta7-desaturase gene FAD5, and effects of plastidial retargeting of Arabidopsis desaturases on the FAD5 mutant phenotype. *Plant Physiology* 136, 4237-4245.

Heyndrickx KS and Vandepoele K (2012) Systematic identification of functional plant modules through the integration of complementary data sources. *Plant Physiology* **159**, 884-901.

Hirai MY, Sugiyama K, Sawada Y, Tohge T, Obayashi T, Suzuki A, Araki R, Sakurai N, Suzuki H, Aoki K, Goda H, Nishizawa OI, Shibata D and Saito K (2007) Omics-based identification of *Arabidopsis* Myb transcription factors regulating aliphatic glucosinolate biosynthesis. *Proceedings of the National Academy of Sciences* 104, 6478–6483.

Howell PM, Sharpe AG and Lydiate DJ (2003) Homoeologous loci control the accumulation of seed glucosinolates in oilseed rape (*B. napus*). *Genome* 46, 454–460.

Hutcheson DS, Falk KC and Rakow GFW (2000) TR4 summer turnip rape. *Canadian Journal of Plant Science* 80, 837–838.

Jensen SK, Liu YG and Eggum BO (1995) The effect of heat treatment on glucosinolates a nutritive value of rapseed meal in rat. *Animal Feed Science and Technology* 53, 17-28.

Jesske T, Olberg B, Schierholt A and Becker HC (2013) Resynthesized lines from domesticated and wild *Brassica* taxa and their hybrids with *B. napus* L.: genetic diversity and hybrid yield. *Theoretical and Applied Genetics* 126, 1053-1065.

Jin L, Lu Y, Xiao P, Sun M, Corke H and Bao J (2010) Genetic diversity and population structure of a diverse set of rice germplasm for association mapping. *Theoretical and Applied Genetics* 121, 475-487.

Kondra ZP and Stefansson BR (1970) Inheritance of major glucosinolates in rapeseed (*Brassica napus*) meal. *Canadian Journal of Plant Science* 50, 643-647.

Krzymanski J (1970) Inheritance of thioglucoside content by rapeseed (*B. napus*). *Journal International conference on oil rape Paris* 37, 213–218.

Lee Y, N. Jeong JH, Kim KJ, Lee KH, Kim et al, (2015) Development, validation and genetic analysis of a large soybean SNP genotyping array. *Plant Journal* 81, 625-636.

Leflon M, Grandont L, Eber F, Huteau V, Coriton O, Chelysheva L, Jenczewski E and Chèvre AM (2010) Crossovers get a boost in *Brassica* allotriploid and allotetraploid hybrids. *Plant Cell* 22, 2253-2264.

Li F, Chen B, Xu K, Wu J, Song W, Bancroft I, Harper AL, Trick M, Liu S, Gao G, Wang N, Yan G, Qiao J, Li J, Li H, Xiao X, Zhang T and Wu X (2014) Genome-wide association study dissects the genetic architecture of seed weight and seed quality in rapeseed (*Brassica napus* L.). *DNA Research* 21, 355-367.

Li P, Filiault D, Box MS, Kerdaffrec E, van Oosterhout C, Wilczek AM, Schmitt J, McMullan M, Bergelson J, Nordborg M and Dean C (2014) Multiple *FLC* haplotypes defined by independent cis-regulatory variation underpin life history diversity in *Arabidopsis thaliana*. *Genes & Development* 15, 1635-40.

Li PW, Zhao YG, Zhang W, Ding XX, Yang M, Wang XF, Xie CH and Fu TD (2005) Analysis of glucosinolate components and profiles in *Brassica napus*. *Scientia Agricultura Sinica* 38, 1346–1352.

Liu H (1985) Rapeseed genetics and breeding. Shanhai Science and Technology Press, Shanghai, pp 559-566.

Luo MC, Yang ZL, You FM, Kawahara T, Waines JG and Dvorak J (2007) The structure of wild and domesticated emmer wheat populations, gene flow between them, and the site of emmer domestication. *Theoretical and Applied Genetics* 114, 947-959.

Ma C, Kimura Y, Fujimoto H, Sakai T, Imamura J and Fu T (2000) Genetic diversity of Chinese and Japanese rapeseed (*Brassica napus* L.) varieties detected by RAPD markers. *Breeding Science* 50, 257–265.

Mace ES, Tai S，Gilding EK，Li Y, Prentils PJ，Bian L, Campbell BC, Hu W, Innes DJ,Han X,Cruickshank A,Dai C,Frere C,Zhang H,Hunt CH,Wang X,Shatte T,Wang M,Su Z,Li J, Lin X,Godwin ID,Jordan DR and Wang J (2013) Whole-genome sequencing reveals untapped genetic potential in Africa's indigenous cereal crop sorghum. *Nature Communication* 4, 2320.

Marchini J, Cardon LR, Phillips MS and Donnelly P (2004) The eVects of human population structure on large genetic association studies. *Nature Genetics* 36,512-517.

Mather KA, Caicedo AL, Polato NR, Olsen KM, McCouch S and Purugganan MD (2007) The extent of linkage disequilibrium in rice (*Oryza sativa* L.). *Genetics* 177, 2223-2232.

Mba C, Guimaraes EP and Ghosh K (2012) Re-orienting crop improvement for the changing climatic conditions of the 21st century. *Agriculture and Food Security* 1, 7.

McNaughton IH (1976) Swedes and rapes. In Evolution of Crop Plants, edited by N. W. Simmonds. Longman, London 1976, 53-56.

Mei J, Li Q, Qian L, Fu Y, Li J, Frauen M and Qian W (2011) Genetic investigation of the origination of allopolyploid with virtually synthesized lines: application to the C subgenome of *Brassica napus*. *Heredity* 106, 955-961.

Mullan BP, Pluske JR, Allen J and Harris DJ (2000) Evaluation of Western Australian canola meal for growing pigs. *Australian Journal of Agricultural Research* 51, 547-553.

Myles S, Boyko AR, Owens CL, Brown PJ, Grassi F, Aradhya MK, Prins B, Reynolds A, Chia JM, Ware D, Bustamante CD and Buckler ES (2011) Genetic structure and domestication history of the grape. *Proceedings of the National Academy of Sciences* 108, 3530-3535.

Murillo G and Mehta RG (2001) Cruciferous vegetables and cancer prevention. *Nutrition and Cancer*, 41, 17-28.

Nagaharu U (1935) Genome analysis in *Brassica* with special reference to the experimental formation of *B. napus* and peculiar mode of fertilisation. *The Journal of Japanese Botany* 7, 389-452.

Nordborg M, Hu TT, Ishino Y, Jhaveri J, Toomajian C, Zheng H, Bakker E, Calabrese P, Gladstone J, Goyal R, Jakobsson M, Kim S, Morozov Y, Padhukasahasram B, Plagnol V, Rosenberg NA, Shah C, Wall JD, Wang J, Zhao K, Kalbfleisch T, Schulz V, Kreitman M and Bergelson J (2005) The pattern of polymorphism in *Arabidopsis thaliana*. *PLoS Biology* 3, e196.

Olsson G (1960) Species crosses within the genus *Brassica*. *Hereditas* 46, 351.

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

Peterson CL, Reece DL, Hammond BL, Thompson J and Beck SM (1997) Processing, Characterization, and Performance of Eight Fuels from Lipids. *Applied Engineering in Agriculture* 13, 71-79.

Qian W, Meng J, Li M, Frauen M, Sass O, Noack J and Jung C (2006) Introgression of genomic components from Chinese *Brassica rapa* contributes to widening the genetic diversity in rapeseed (*B. napus* L.), with emphasis on the evolution of Chinese rapeseed. *Theoretical and Applied Genetics* 113, 49-54.

Qian W, Zhou H and Tang K (2014) Recent coselection in human populations revealed by protein-protein interaction network. *Genome Biology and Evolution* 7,136-53.

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

Remington DL, Thornsberry JM, Matsuoka Y, Wilson LM, Whitt SR, Doebley J, Kresovich S, Goodman MM and Buckler ESth (2001) Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proceedings of the National Academy of Sciences* 98, 11479-11484.

Rygulla W, Friedt W, Seyis F, Lühs W, Eynck C, von Tiedemann A and Snowdon RJ (2007a) Combination of resistance to *Verticillium longisporum* from zero erucic acid *Brassica oleracea* and oilseed *Brassica rapa* genotypes in resynthesized rapeseed (*Brassica napus*) lines. *Plant Breeding* 126, 596-602.

Sanchez MP, Tribout T, Iannuccelli N, Bouffaud M, Servin B, Tenghe A, Dehais P, Muller N, Del Schneider MP, Mercat MJ, Rogel-Gaillard C, Milan D, Bidanel JP and Gilbert H (2014) A genome-wide association study of production traits in a commercial population of Large White pigs: evidence of haplotypes affecting meat quality. *Genetics Selection Evolution* 46, 1-12.

Shiga T (1970) Rapa breeding by interspecific crossing between *Brassica napus* and *Brassica campestris* in japan. *Japan Agricultural Research Quarterly* 5, 5-10.

Slominski BA, Simbaya J, Campbell L, Rakow G and Guenter W (1999) Nutritive value for broilers of meals derived from newly developed varieties of yelow-seeded canola. *Animal Feed Science and Technology* 78, 249-262.

Stephens JC, Schneider JA, Tanguay DA, Choi J, Acharya T, Stanley SE, Jiang R, Messer CJ, Chew A, Han JH, Duan J, Carr JL, Lee MS, Koshy B, Kumar AM, Zhang G, Newell WR, Windemuth A, Xu C, Kalbfleisch TS, Shaner SL, Arnold K, Schulz V, Drysdale CM, Nandabalan K, Judson RS, Ruano G and Vovis GF (2001) Haplotype variation and linkage disequilibrium in 313 human genes. *Science* 293, 489-493.

Tautorus CL and Low NH (1993) Chemical Aspects of Chlorophyll Breakdown Products and Their Relevance to Canola Oil Stability. *Journal of the American oil Chemists Society* 70, 843-847.

Thomas H and Ougham H (2014) The stay-green trait. *Journal of Experimental Botany* 65, 3889-3900.

Uzunova M, Ecke W, Weissleder K and Röbbelen G (1995) Mapping the genome of rapeseed (*Brassica napus* L.). I. Construction of an RFLP linkage map and localization of QTLs for seed glucosinolate content. *Theoretical and Applied Genetics* 90, 194–204.

Voss-Fels K, Frisch M, Qian L, Kontowski S, Friedt W, Gottwald S and Snowdon RJ (2015) Subgenomic diversity patterns caused by directional selection in *bread wheat* gene pools. *Plant Genome* 8*,* doi:10.3835/plantgenome2015.03.0013*.*

Voss-Fels K and Snowdon RJ (2015) Understanding and utilizing crop genome diversity via high-resolution genotyping. *Plant Biotechnology Journal* DOI: 10.1111/pbi.12456.

Wall JD and Pritchard JK (2003) Assessing the performance of the haplotype block model of linkage disequilibrium. *The American Journal of Human Genetics* 73, 502-515.

Wang F, Wang G, Li X, Huang J and Zheng J (2008) Heredity, physiology and mapping of a chlorophyll content gene of rice (*Oryza sativa* L.). *Journal of Plant Physiology* 165, 324-330.

Wang N, Li F, Chen B, Xu K, Yan G, Qian J, Li J, Gao G, Bancroft I, Meng J, King GJ and Wu X (2014) Genome-wide investigation of genetic changes during modern breeding of *Brassica napus*. *Theoretical and Applied Genetics* 127, 1817-1829.

Wang S, Wong D, Forrest K, Allen A, Chao S, Huang BE, Maccaferri M, Salvi S, Milner SG, Cattivelli L, Mastrangelo AM, Whan A, Stephen S, Barker G, Wieseke R, Plieske J, International Wheat Genome Sequencing Consortium, Lillemo M, Mather D, Appels R, Dolferus R, Brown-Guedira G, Korol A, Akhunova AR, Feuillet C, Salse J, Morgante M, Pozniak C, Luo MC, Dvorak J, Morell M, Dubcovsky J, Ganal M, Tuberosa R, Lawley C, Mikoulitch I, Cavanagh C, Edwards KJ, Hayden M and Akhunov E (2014) Characterization of polyploid wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array. *Plant Biotechnology Journal* 12, 787–796.

Wittkop B, Snowdon RJ and Friedt W (2009) Status and perspectives of breeding for enhanced yield and quality of oilseed crops for Europe. *Euphytica* 170, 131–140.

Yan JB, Shah T, Warburton ML, Buckler ES, McMullen MD and Crouch J (2009) Genetic Characterization and Linkage Disequilibrium Estimation of a Global Maize Collection Using SNP Markers. *PLoS One* 4, e8451.

Yang J and Zhu ZJ (2009) Glucosinolates in Pakchoi (*Brassica Campestris* L. ssp. Chinensis) as Affected by Nutrient Status and Postharvest. PhD thesis.

Yang QZ, Li W, Li L, Ku C. Wang et al, (2013) CACTA-like transposable element in ZmCCT attenuated photoperiod sensitivity and accelerated the postdomestication spread of maize.Proc. *Proceedings of the National Academy of Sciences* 42, 16969–16974.

Zhao J and Meng J (2003) Detection of loci controlling seed glucosinolate content and their association with *Sclerotinia* resistance in *Brassica napus*. *Plant Breeding* 122, 19–23.

Zhao K, Tung CW, Eizenga GC, Wright MH, Ali ML, Price AH, Norton GJ, Islam MR, Reynolds A, Mezey J, McClung AM, Bustamante CD and McCouch SR (2011) Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. *Nature Communication* 2, 467.

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

# 9 Appendices

**Appendix I:** Electronic supplementary materials from Fu et al. (2015)

**Supplementary_Table S1:** Correlation of seed glucosinolate content between different environments in the DH population derived from 'EXPRESS' × 'SWU07'.

| Environment[1] | G2008 | G2012 | C2009 | C2010 | C2011 | C2012 |
|---|---|---|---|---|---|---|
| G2012 | 0.85** | | | | | |
| C2009 | 0.59** | 0.59** | | | | |
| C2010 | 0.57** | 0.67** | 0.67** | | | |
| C2011 | 0.48** | 0.55** | 0.71** | 0.70** | | |
| C2012 | 0.43** | 0.52** | 0.51** | 0.56** | 0.62** | |
| C2013 | 0.55** | 0.53** | 0.71** | 0.59** | 0.73** | 0.69** |

[1]: 'G' represents the environment in German, and 'C' represents the environment in China
** represents significance at $P = 0.01$ level

**Supplementary_Table S2:** QTL for seed glucosinolate (GSL) content and the microsynteny analysis of QTL against the reference genomes of *B. rapa* and *B. oleracea* for identification of GSL biosynthesis genes.

| QTL | Chr.[a] | Pos.[b] | Add.[c] | LOD[d] | $R^{2e}$ | Markers | Physical location | GSL biosynthesis gene orthologue |
|---|---|---|---|---|---|---|---|---|
| qGSLC09 | A02 | 3.2～11.2 | - | 2.53 | 4.69 | C2-141～A2-246 | R02: 0.75～3.23 | CYP79A2, **IPMDH1** |
| qGSLC11 | A02 | 3.2～10 | - | 4.58 | 5.87 | C2-141～A2-246 | R02: 0.75～3.23 | CYP79A2, **IPMDH1** |
| qGSLC08 | A02 | 6～11 | - | 3.5 | 5.99 | C2-143～A2-246 | R02: 0.81～3.23 | **IPMDH1** |
| qGSLG12 | A02 | 5.5～11 | - | 2.78 | 5.01 | C2-143～A2-246 | R02: 0.81～3.23 | **IPMDH1** |
| qGSLG08 | A02 | 12.6～ | - | 5.22 | 6.32 | A2-246～YD091 | R02: 3.23～6.20 | TFL2, ATMS1 |
| qGSLG12 | A02 | 12.7～ | - | 3.62 | 3.85 | A2-246～YD091 | R02: 3.23～6.20 | TFL2, ATMS1 |
| qGSLC13 | A02 | 29～32.4 | + | 8.68 | 14.74 | BRMS082～OI10C05 | R02: 9.59～13.14 | MYB122 |
| qGSLG08 | A02 | 37.6～39 | + | 3.29 | 4.02 | OI10C05～C2-606 | R02: 13.14～19.08 | - |
| qGSLC12 | A02 | 39.8～40 | + | 12.72 | 20.12 | C2-606～SWUC283 | R02: 15.92～19.08 | - |
| qGSLC13 | A02 | 39.8～ | + | 21.57 | 31.97 | C2-606～SWUC283 | R02: 15.92～19.08 | - |
| qGSLG12 | A02 | 41.5～ | + | 3.38 | 3.35 | SWUC283～CB10316 | R02: 13.88～15.92 | - |
| qGSLC09 | A02 | 42.7～ | + | 9.42 | 15.59 | SWUC283～CB10316 | R02: 13.88～15.92 | - |
| qGSLC10 | A02 | 42.7～ | + | 8.28 | 10.84 | SWUC283～CB10316 | R02: 13.88～15.92 | - |
| qGSLG08 | A02 | 42.9～ | + | 4.18 | 5.01 | SWUC283～CB10316 | R02: 13.88～15.92 | - |
| qGSLC11 | A02 | 43～44.9 | + | 15.96 | 20.99 | SWUC283～CB10316 | R02: 13.88～15.92 | - |
| qGSLC12 | A02 | 43.4～ | + | 12.73 | 19.71 | SWUC283～CB10316 | R02: 13.88～15.92 | - |
| qGSLC10 | A03 | 19.9～32 | + | 4.92 | 7.29 | BRMS176～CNU146 | R03: 22.08～31.04 | **APR3**, GSH1, GGP1 |
| qGSLG08 | A03 | 23.3～ | + | 4.78 | 6.25 | CNU316～CNU146 | R03: 22.58～31.04 | **APR3**, GSH1, GGP1 |
| qGSLG12 | A03 | 23.6～ | + | 7.92 | 9.46 | CNU316～CNU146 | R03: 22.58～31.04 | **APR3**, GSH1, GGP1 |
| qGSLC09 | A03 | 25.4～ | + | 3.75 | 11.49 | CNU316～CNU146 | R03: 22.58～31.04 | **APR3**, GSH1, GGP1 |
| qGSLC09 | A03 | 34.8～ | + | 3.91 | 13.5 | CNU146～A3-1487b | R03: 24.93～31.04 | **APR3**, GSH1, GGP1 |
| qGSLG08 | A03 | 34.6～ | + | 2.35 | 4.1 | CNU146～A3-1487b | R03: 24.93～31.04 | **APR3**, GSH1, GGP1 |
| qGSLG12 | A03 | 35～40.4 | + | 4.87 | 8.92 | CNU146～YD065 | R03: 24.19～31.04 | **APR3**, GSH1, GGP1 |
| qGSLG08 | A04 | 1.8～8 | + | 13.71 | 18.22 | CNU254～YD061b | R02: 23.47～26.13 | MYB28, MYB34, MAM1, MAM3, CHY1, TGG1, CYSD2, APK3 |
| qGSLC10 | A04 | 2.6～9.3 | + | 9.06 | 12.03 | CNU254～YD061b | R02: 23.47～26.13 | MYB28, MYB34, MAM1, MAM3, CHY1, TGG1, CYSD2, APK3 |
| qGSLG12 | A04 | 3～8.7 | + | 17.84 | 24.87 | CNU254～YD061b | R02: 23.47～26.13 | MYB28, MYB34, MAM1, MAM3, CHY1, TGG1, CYSD2, APK3 |

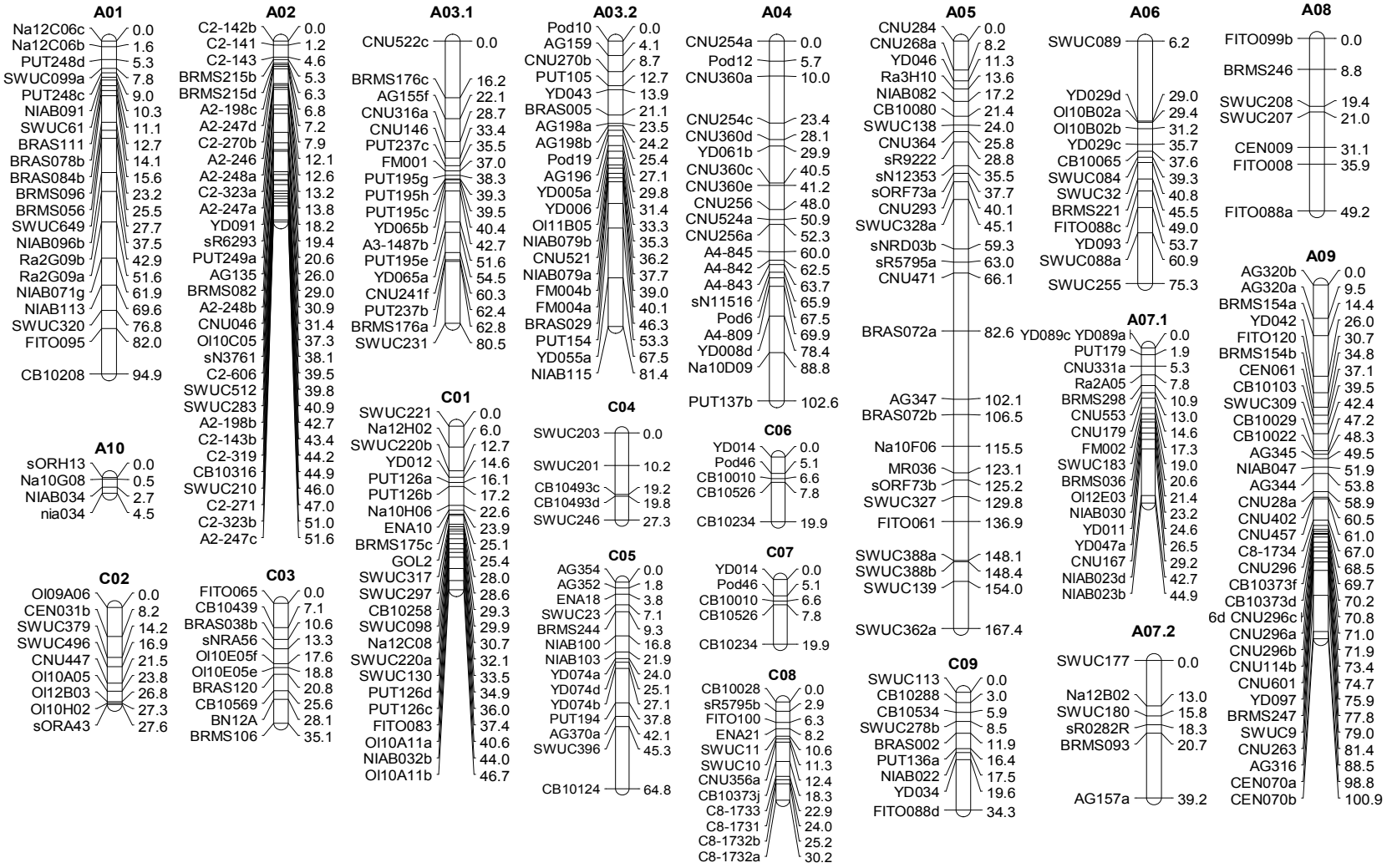
| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| qGSLC09 | A04 | 2.7～13 | + | 7.63 | 13.3 | CNU254～YD061b | R02: 23.47～26.13 | MYB28, MYB34, MAM1, MAM3, CHY1, TGG1, CYSD2, APK3 |
| qGSLC11 | A04 | 1～14.5 | + | 5.1 | 6.54 | CNU254～YD061b | R02: 23.47～26.13 | MYB28, MYB34, MAM1, MAM3, CHY1, TGG1, CYSD2, APK3 |
| qGSLC12 | A04 | 0～16.8 | + | 4.79 | 7.07 | CNU254～YD061b | R02: 23.47～26.13 | MYB28, MYB34, MAM1, MAM3, CHY1, TGG1, CYSD2, APK3 |
| qGSLC13 | A04 | 1.5～15.4 | + | 5.41 | 7.23 | CNU254～YD061b | R02: 23.47～26.13 | MYB28, MYB34, MAM1, MAM3, CHY1, TGG1, CYSD2, APK3 |
| qGSLC09 | A07 | 10.3～ | - | 4.74 | 7.4 | Ra2A05～BRMS036 | R07: 19.93～22.76 | BZO1p1 |
| qGSLG08 | A09 | 67.1～ | - | 8.43 | 10.83 | C8-1734～CB10373 | R09: 29.98～30.80 | - |
| qGSLG12 | A09 | 67～70.8 | - | 6.69 | 7.36 | C8-1734～CB10373 | R09: 29.98～30.80 | - |
| qGSLC11 | A09 | 71～73.4 | - | 3.19 | 3.77 | CNU296～CNU114 | R09: 30.76～31.71 | CYSC1 |
| qGSLG08 | A09 | 75.9～81 | - | 9.57 | 12.17 | CNU601～CNU263 | R09: 31.81～33.71 | - |
| qGSLG12 | A09 | 75.9～81 | - | 7.29 | 7.97 | CNU601～CNU263 | R09: 31.81～33.71 | - |
| qGSLC13 | A09 | 76.9～ | - | 3.73 | 4.58 | CNU601～CNU263 | R09: 31.81～33.71 | - |
| qGSLG08 | A09 | 81.4～ | - | 8.56 | 12.41 | CNU263～CEN070 | R09: 33.71～38.33 | Dof1.1, FMOGS-OX5, MYB51, SUR1 |
| qGSLG12 | A09 | 81.4～ | - | 5.34 | 7.32 | CNU263～AG316 | R09: 33.71～35.94 | MYB51, SUR1 |
| qGSLC10 | A09 | 81.5～ | - | 3.75 | 4.71 | CNU263～CEN070 | R09: 33.71～38.33 | Dof1.1, FMOGS-OX5, MYB51, SUR1 |
| qGSLC13 | A09 | 85.3～ | - | 5.31 | 6.4 | CNU263～CEN070 | R09: 33.71～38.33 | Dof1.1, FMOGS-OX5, MYB51, SUR1 |
| qGSLG12 | C03 | 22.8～ | + | 3.93 | 4.17 | BRAS120～BN12A | O03: 7.16～8.60 | - |
| qGSLC13 | C08 | 0～10.4 | + | 3.28 | 3.88 | CB10028～SWUC11 | O08: 35.25～39.01 | SUR1, TGG4, FMOGS-OX5 |

[a] Chromosome
[b] Length of 2-LOD score confidence interval
[c] Additive effect. The direction of additive effect is from the allele of 'EXPRESS'.
[d] Peak effect of the QTL (LOD, limit of detection)
[e] Percentage of the phenotypic variance explained by each QTL

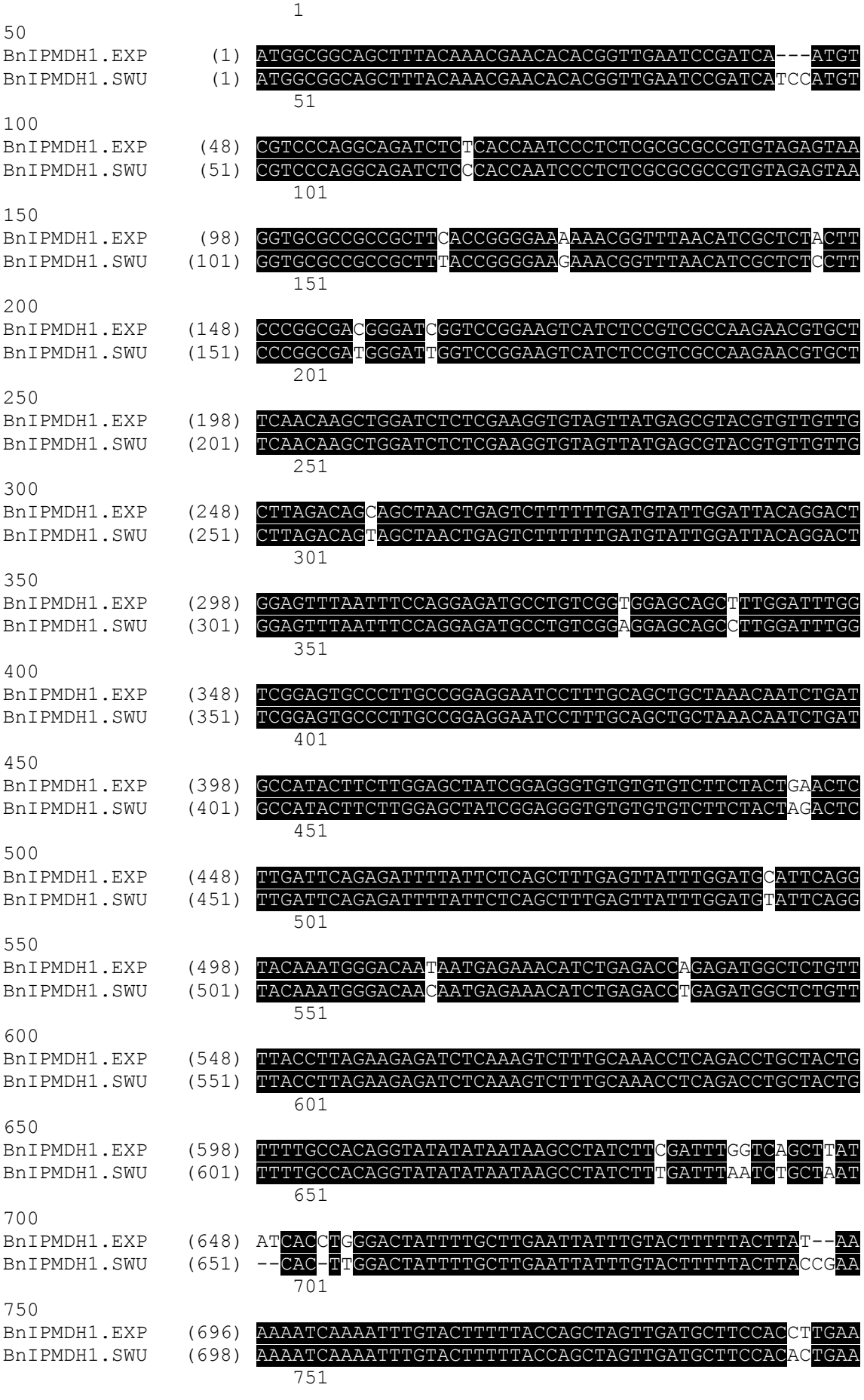
**Supplementary_Table S3:** Information of the 23 genes involved in glucosinolate mechalism detected in this study.

| Gene | Code in tair | Code in *B. rapa* | Location (Mb) | Function |
|---|---|---|---|---|
| CYP79A2 | At5g05260 | Bra028764 | A02: 0.77 | Core structure formation |
| IPMDH1 | At5g14200 | Bra023450 | A02: 2.39 | Amino acid side chain elongation |
| TFL2 | At5g17690 | Bra023629 | A02: 3.23 | Transcription factors |
| ATMS1 | At5g17920 | Bra023645 | A02: 3.32 | primary S-metabolism |
| MYB122 | At1g74080 | Bra008131 | A02: 12.33 | Transcription factors |
| APR3 | At4g21990 | Bra019406 | A03: 24.39 | primary S-metabolism |
| GSH1 | At4g23100 | Bra019332 | A03: 24.95 | Co-substrate pathways |
| GGP1 | At4g30530 | Bra024068 | A03: 27.90 | Core structure formation |
| MYB28 | At5g61420 | Bra029311 | A02: 25.40 | Transcription factors |
| MYB34 | At5g60890 | Bra029349 | A02: 25.18 | Transcription factors |
| MAM1 | At5g23010 | Bra029355 | A02: 25.12 | Side-chain elongation |
| MAM3 | At5g23020 | Bra029356 | A02: 25.10 | Side-chain elongation |
| CHY1 | At5g65940 | Bra031802 | A02: 26.92 | Co-substrate pathways |
| TGG1 | At5g26000 | Bra020551 | A02: 24.08 | Breakdown |
| CYSD2 | At5g28020 | Bra020605 | A02: 23.71 | Cysteine biosynthesis |
| APK3 | At3g03900 | Bra031965 | A05: 24.59 | Co-substrate pathways |
| BZO1p1 | At1g65880 | Bra004132 | A07: 20.29 | Co-substrate pathways |
| CYSC1 | At3g61440 | Bra007604 | A09: 31.50 | Cysteine biosynthesis |
| Dof1.1 | At1g07640 | Bra031588 | A09: 37.44 | Transcription factors |
| FMOGS-OX5 | At1g12140 | Bra026988 | A09: 36.39 | Secondary modification |
| MYB51 | At1g18570 | Bra031035 | A09: 34.59 | Transcription factors |
| SUR1 | At2g20610 | Bra031132 | A09: 34.10 | Core structure formation |
| TGG4 | At1g47600 | Bol044759 | C08: 35.48 | Breakdown |

## Supplementary_Fig S1



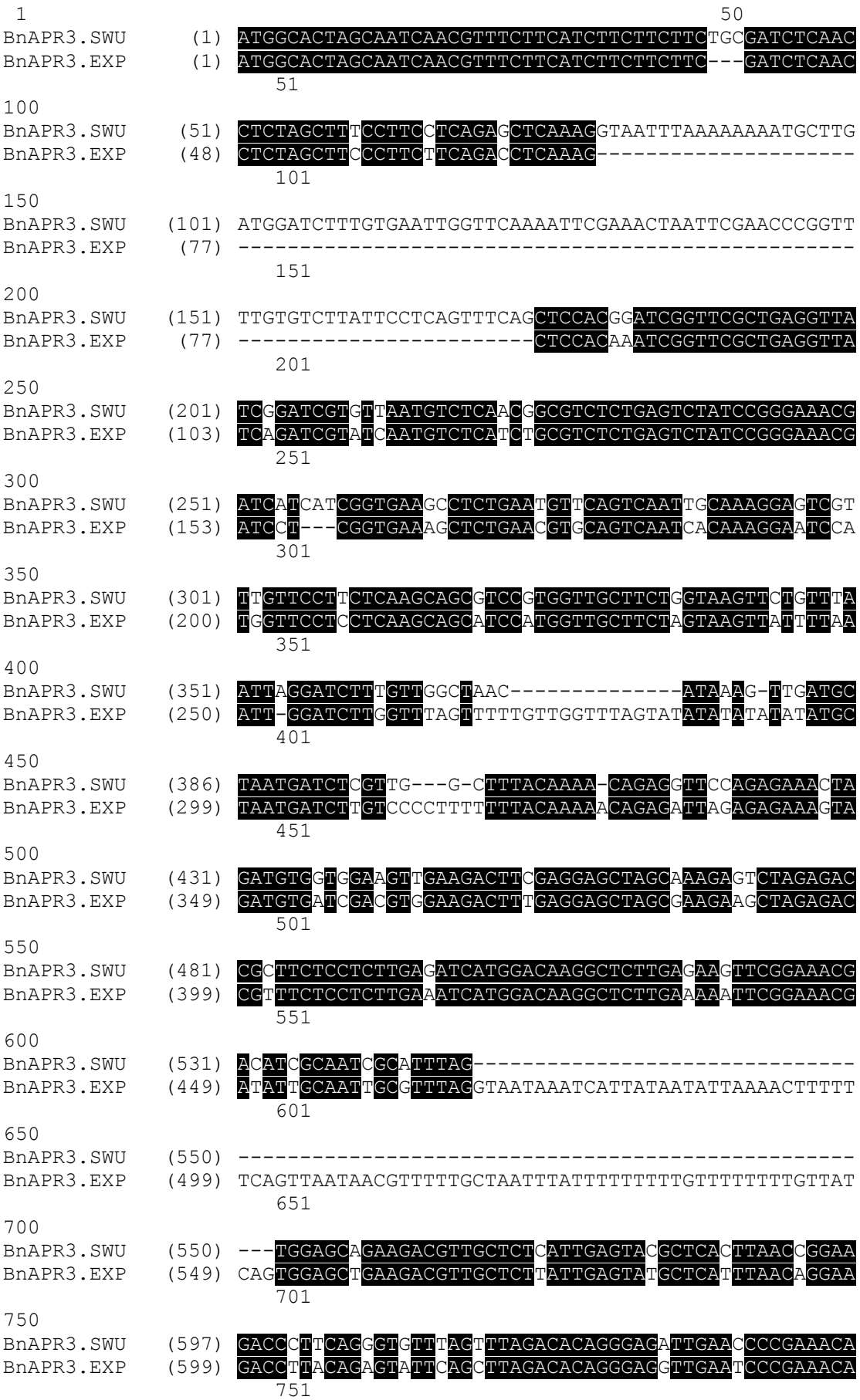
**Supplementary_Fig S2**



```
800
BnIPMDH1.EXP   (746) GAGAGAAGTGGCAGAAGGTGTTGATATGATGATTGTAAGGGAGCTTACTG
BnIPMDH1.SWU   (748) GAGAGAAGTGGCAGAAGGCGTTGATATGATGATTGTAAGGGAGCTTACAG
                        801
850
BnIPMDH1.EXP   (796) GAGGTATCTACTTTGGAGAGCCAAGGGGCATTAAAACCGATGAAAATGGC
BnIPMDH1.SWU   (798) GAGGTATCTACTTTGGAGAGCCAAGGGGAATCAAGACCAACGAAAATGGC
                        851
900
BnIPMDH1.EXP   (846) GAAGAAGTCGGCTTTAATACAGAGTTCTACGCTGCTCATGAGGTGCTTAC
BnIPMDH1.SWU   (848) GAAGAAGTCGGCTTTAATACAGAGCTCTACGCTGCTCATGAGGTACTTAC
                        901
950
BnIPMDH1.EXP   (896) TAATACGACTACTCTTCACAATGTTCTCTCCATGACATTGGTTGTATAGT
BnIPMDH1.SWU   (898) TA---------CTCTTCACAATGTTCTCTCCATGACATTGGTTATATAGT
                        951
1000
BnIPMDH1.EXP   (946) AA----TGTGTTTCTCATGTTTTTCAGATCGATAGAATTGCTCGCGTTGC
BnIPMDH1.SWU   (939) AAGTAATGTGTTTCTCATGTTCTTCAGATTGATAGAATTGCTCGTGTTGC
                       1001
1050
BnIPMDH1.EXP   (992) ATTCGAGACTGCTAGGAAACGGCGTGGCAAGCTGTGTTCTGTCGACAAAG
BnIPMDH1.SWU   (989) GTTCGAGACTGCTAGGAAACGGCGTGGCAAGCTGTGTTCTGTCGACAAAG
                       1051
1100
BnIPMDH1.EXP  (1042) CCAATGTGTTGGATGTACGTTTGATACTTGTGTTTGAATCTGTTGTGTGT
BnIPMDH1.SWU  (1039) CCAATGTGTTGGATGTACGTTTGATACTTGTGTTTGAATCTGTTGA---T
                       1101
1150
BnIPMDH1.EXP  (1092) GGGCATTTCTAGTGATGCTAATTTATTTTTTGGTCATATAGGCATCAATA
BnIPMDH1.SWU  (1086) GTGTATTTCTAGTGATGCTAATTTGTTTTTTGGTCATATAGGCATCAATA
                       1151
1200
BnIPMDH1.EXP  (1142) TTGTGGAGGAGAAGAGTAACAGCGTTAGCCTCTGAGTATCCAGACGTTGA
BnIPMDH1.SWU  (1136) TTGTGGAGGAAAAGAGTAACAGCGTTAGCCTCTGAGTATCCAGACGTTGA
                       1201
1250
BnIPMDH1.EXP  (1192) ACTAACACATATGTATGTCGACAATGCTGCAATGCAGCTTATTCGTGACC
BnIPMDH1.SWU  (1186) ACTAACACATATGTATGTCGACAATGCTGCAATGCAGCTTATTCGTGACC
                       1251
1300
BnIPMDH1.EXP  (1242) CCAAACAGGTGATCTTCTCGTCCATTGGTTAAAGTGGTAACTTAGTAGTA
BnIPMDH1.SWU  (1236) CCAAACAGGTGCTCTTCTCGTCCATTGGTTAAAGTGGTAACTTAGTAGTA
                       1301
1350
BnIPMDH1.EXP  (1292) ACTCAATGTGTTTTTTAATATTTTTTAAAAAACATTTGCAGTTTGACACA
BnIPMDH1.SWU  (1286) ACTCAATGTGTTTTTTAATATTTTTTAAAAAACATTTGCAGTTTGACACA
                       1351
1400
BnIPMDH1.EXP  (1342) ATAGTCACCAATAACATTTTTGGTGATATATTGTCTGATGAGGCTTCAAT
BnIPMDH1.SWU  (1336) ATAGTCACCAATAACATTTTTGGTGATATATTGTCTGATGAGGCTTCAAT
                       1401
1450
BnIPMDH1.EXP  (1392) GATCACTGGAAGCATTGGGATGCTTCCATCTGCTAGTCTCGGTGTATCGG
BnIPMDH1.SWU  (1386) GATCACTGGAAGCATTGGGATGCTTCCATCTGCTAGTCTCGGTGTATCGG
                       1451
1500
BnIPMDH1.EXP  (1442) TAAAATAAAATTAAAATCTGAAAATCTTCCAAAGGTTCCTACGTACAACA
BnIPMDH1.SWU  (1436) TAAAATAAAATTAAAATCTAAAAATGTTCCAAAGGTTCCTA----CAACA
                       1501
1550
BnIPMDH1.EXP  (1492) ATATCTTGATTGTTAATTAACTTGTGTTGGAAT--CAGGGACCTGGACTG
BnIPMDH1.SWU  (1482) ATATCTTGATTGTTAA--AACTTGTGTTGAAATATCAGGGACCTGGACTG
```

```
                      1551
1600
BnIPMDH1.EXP  (1540) TTTGAGCCTATACATGGTTCAGCACCAGATATAGCTGGCCAGGACAAGGC
BnIPMDH1.SWU  (1530) TTTGAGCCTATACATGGTTCAGCACCAGATATAGCTGGTCAGGACAAGGC
                      1601
1650
BnIPMDH1.EXP  (1590) AAACCCATTGGCCACCATTCTCAGCGCTGCAATGCTTCTGAAATATGGAC
BnIPMDH1.SWU  (1580) AAACCCATTGGCCACCATTCTCAGCGCTGCAATGCTTCTGAAATATGGAC
                      1651
1700
BnIPMDH1.EXP  (1640) TTGGAGAAGAAAAGGCTGCAAAGAGGATCGAAGACGCGGTGTTGGATACT
BnIPMDH1.SWU  (1630) TTGGAGAAGAAAAGGCTGCAAAGAGGATCGAAGACGCGGTGTTGGATACT
                      1701
1750
BnIPMDH1.EXP  (1690) TTGAACAAAGGGTTTAGAACCGGTGACATCTACTCCCCTGGAAATGTATG
BnIPMDH1.SWU  (1680) TTGAACAAAGGATTTAGAACCGGTGACATCTACTCCCCTGGAAATGTATG
                      1751
1800
BnIPMDH1.EXP  (1740) TATGCTGATTTTCATTTTTTT-ATAAGTTAGTAAGCACAGTTATTAATAT
BnIPMDH1.SWU  (1730) TATGCTGATTTTCATTTTTTTTATAAGTTAGTAAGCACAGTTATTAATAT
                      1801
1850
BnIPMDH1.EXP  (1789) TTACTTAAAAGATCTTGTGTTTTTGATAATATGAAACAGAAACTGGTGGG
BnIPMDH1.SWU  (1780) TTACTTAAAAGATCTTGTGTTTTTGATAATATGAAACAGAAACTGGTGGG
                      1851
1900
BnIPMDH1.EXP  (1839) ATGCAAGGAGATGGGTGAGGAAGTGCTAAAATCAGTGGACTCCAAAGTTA
BnIPMDH1.SWU  (1830) ATGCAAGGAGATGGGTGAGGAAGTGCTAAAATCAGTGGACTCCAAAGTTA
                      1901        1914
BnIPMDH1.EXP  (1889) CAGCTACTGTTTAA
BnIPMDH1.SWU  (1880) CAGCTACTGTTTAA
```

**Supplementary_Fig S3**



```
800
BnAPR3.SWU   (647) TACAGACTCTTCGACACCGTGGAGAAGCACTACGGTATTCGAATCGAGTA
BnAPR3.EXP   (649) TACAGACTCTTCGACACCGTGGAGAAACACTACGGTATTCGAATCGAGTA
                      801
850
BnAPR3.SWU   (697) CATGTTTCCCGACGCTGTTGAGGTCCAAGCTCTGGTTAGAAACAAAGGTT
BnAPR3.EXP   (699) CATGTTTCCCGACGCAGTTGAGGTTCAAGCTTTGGTTAGAAACAAAGGTC
                      851
900
BnAPR3.SWU   (747) TGTTCTCTTTCTACGAAGACGGTCACCAGGAGTGTTGCCGCATCAGAAAG
BnAPR3.EXP   (749) TGTTCTCTTTCTACGAAGACGGTCACCAAGAGTGTTGCCGTATCAGAAAA
                      901
950
BnAPR3.SWU   (797) GTTAGACCACTGAGGCGTGCGTCGAAGGGCTTACGCGCTTGGATCACTGG
BnAPR3.EXP   (799) GTGAGACCGTTGAGGCGTGCTTTAAAAGGTTTACGCGCTTGGATCACCGG
                      951
1000
BnAPR3.SWU   (847) ACAAAGGAAAGATCAGTCACCAGGGACGAGATCAGAGATCCCCGTTGTTC
BnAPR3.EXP   (849) ACAAAGAAAAGATCAATCACCAGGGACAAGATCAGAGATCCCCGTTGTTC
                     1001
1050
BnAPR3.SWU   (897) AAGTTGATCCGGTGTTTGAAGGGTTAGACGGTGGAGCTGGTAGTTTGGTG
BnAPR3.EXP   (899) AAGTCGATCCGGTGTTCGAAGGACTAGACGGTGGAGCTGGTAGCTTGGTG
                     1051
1100
BnAPR3.SWU   (947) AAGTGGAATCCGGTTGCGAATGTCGAAGGGAACGATGTTTGGAACTTCTT
BnAPR3.EXP   (949) AAGTGGAACCCCGTTGCGAACGTGGAAGGGAACGATGTTTGGAACTTCTT
                     1101
1150
BnAPR3.SWU   (997) GAGGACTATGGATGTGCCCGTGAACACGCTTCACGCTGCGGGGTATGTTT
BnAPR3.EXP   (999) GAGGACTATGGGTGTTCCCGTCAACACGCTGCACGCTGCGGGGTATGTTT
                     1151
1200
BnAPR3.SWU  (1047) CTATAGGATGTGAGCCGTGCACGAGAGCGGTTTTGCCTGGGCAGCACGAG
BnAPR3.EXP  (1049) CCATAGGGTGCGAGCCGTGCACGAGGGCGGTTTTGCCAGGTCAGCACGAG
                     1201
1250
BnAPR3.SWU  (1097) AGAGAAGGGAGATGGTGGTGGGAAGACGCTAAGGCTAAAGAGTGTGGACT
BnAPR3.EXP  (1099) AGGGAAGGGAGGTGGTGGTGGGAAGACGCTAAGGCGAAAGAGTGTGGGCT
                     1251
1300
BnAPR3.SWU  (1147) TCACAAAGGGAACATCAAGGAGAACAGCAACGGGAAC---GCTAATGCTA
BnAPR3.EXP  (1149) TCACAAAGGGAACATCAAAGAGAGTAGCAACGGAAACAACGCAGCGGTTA
                     1301
1350
BnAPR3.SWU  (1194) ATGTCAATGGGACGTCATCCACGGTTGCTGATATCTTCAAGAGCGAGAAT
BnAPR3.EXP  (1199) ATGGCAATGGGACAACGTCAACGGTCGATGATATTTTCAAAAGCGAGAAC
                     1351
1400
BnAPR3.SWU  (1244) GTTGTGAGCTTGAGCAGGCAAGGGATTGAGAATCTGATGAAGCTGGAGAA
BnAPR3.EXP  (1249) GTTGTGAGCTTGAGCAGGCAAGGGATTGAGAATCTGATGAAGCTGGAGAA
                    1401
1450
BnAPR3.SWU  (1294) TCGTAAAGAGGCTTGGATCGTTGTGCTTTACGCGCCTTGGTGCCCGTTTT
BnAPR3.EXP  (1299) CAGGAAGGAGGCGTGGATCGTTGTGCTTTACGCGCCTTGGTGCCCGTTTT
                     1451
1500
BnAPR3.SWU  (1344) GTCAGGCGATGGAAGGTTCTTTTGATGAGTTGGCGGATAAGTTG--GGTG
BnAPR3.EXP  (1349) GCCAAGCGATGGAGGCTTCGTTTGATGAGTTGGCGGATAAGTTGAAGGGA
                     1501
1550
BnAPR3.SWU  (1392) GGAG-TGGCGTGAAGGTGGCGAAGTTTAGAGCTGATGGTGACCAGAAGGA
BnAPR3.EXP  (1399) GGAGATGGCGTGAAGGTGGCTAAGTTTAGAGCGGACGGTGACCAAAAAGA
```

```
                    1551
1600
BnAPR3.SWU  (1441) GTTTGCTAAAAGGGAGTTGCAGCTTGGGAGCTTCCCGACGATGCTCGTGT
BnAPR3.EXP  (1449) GTTTGCTAAGAGTGAGTTGCAGCTAGGGAGCTTTCCGACGATACTTGTGT
                    1601
1650
BnAPR3.SWU  (1491) TCCCGAAGAACTCTTCACGACCAATCAAGTATCCGTCAGAGAAGAGGGAT
BnAPR3.EXP  (1499) TTCCGAAGAACTCTTCGAGACCTATCAAGTATCCATCTGAGAAGAGGGAT
                    1651                                        1689
BnAPR3.SWU  (1541) GTTGATTCTTTGACATCTTTCTTGAATCTTGTTAGGTAA
BnAPR3.EXP  (1549) GTTGATTCTTTGACA
```

**Appendix II:** Electronic supplementary materials from Qian et al. (2014)

**Additional file 1:** Genotype matrix, flanking sequences and genomic positions (best BLAST hit against the Darmor-*Bzh B. napus* V4.1 reference genome) for the subset of 24,994 high-quality, polymorphic, single-locus *Brassica napus* SNP markers with MAF ≥0.05, as used for the LD analyses. SNP allele calls were generated in 203 *B. napus* breeding lines using the Illumina 60kSNP Infinium Brassica Consortium Array (Illumina Inc., San Diego, USA.) (on the attached CD ROM).

**Additional file 2:** Proportional memberships in subpopulations as defined by Structure.

| Code | Subpopulations | Q1 | Q2 | Q3 | Source |
|---|---|---|---|---|---|
| SW001 | Q1 | 0.640 | 0.329 | 0.031 | (96-331)-7-41-8-8-3 |
| SW003 | Q1 | 0.773 | 0.066 | 0.161 | (Zhengchun ×Qin2-2)F5 |
| SW118 | Q1 | 1.000 | 0.000 | 0.000 | winter |
| SW128 | Q1 | 0.621 | 0.101 | 0.278 | winter |
| SW132 | Q1 | 0.624 | 0.280 | 0.097 | spring |
| SW004 | Q1 | 0.600 | 0.000 | 0.400 | [(Zhengchun ×Qin2-2) ×Yuyou No.2]F6 |
| SW012 | Q1 | 0.776 | 0.224 | 0.000 | (821 Xuan ×Pin 93-498) F8 |
| SW013 | Q1 | 0.746 | 0.254 | 0.000 | Ningyou 10 |
| SW016 | Q1 | 0.639 | 0.361 | 0.000 | 88(2)-1-2-1-2-3-2 |
| SW018 | Q1 | 0.622 | 0.047 | 0.331 | 98009 restorer-8-3 |
| SW019 | Q1 | 0.778 | 0.222 | 0.000 | A-35-1 |
| SW020 | Q1 | 0.715 | 0.101 | 0.184 | B62-4-5-32-7-11 |
| SW021 | Q1 | 0.696 | 0.136 | 0.168 | C022-30-8-3-2 |
| SW023 | Q1 | 0.773 | 0.047 | 0.180 | C179-3-36-8 |
| SW026 | Q1 | 0.623 | 0.283 | 0.094 | H98-18-30-7-17-4 |
| SW027 | Q1 | 0.625 | 0.282 | 0.094 | R13 Xuan a-11-35-4-8 |
| SW031 | Q1 | 0.778 | 0.222 | 0.000 | Fuyu No.1 |
| SW033 | Q1 | 0.659 | 0.201 | 0.140 | Ganyou No.13 |
| SW034 | Q1 | 0.704 | 0.296 | 0.000 | Guinong 78-6-112 |
| SW036 | Q1 | 0.810 | 0.190 | 0.000 | Huayou 6 |
| SW038 | Q1 | 0.931 | 0.069 | 0.000 | Youyou 14 |
| SW042 | Q1 | 0.722 | 0.278 | 0.000 | Huiyou 50-8-1 |
| SW046 | Q1 | 0.620 | 0.132 | 0.247 | Liangyou586 -6-3 |
| SW047 | Q1 | 0.630 | 0.001 | 0.369 | Longyou 2 |
| SW048 | Q1 | 0.766 | 0.211 | 0.022 | Huyou 15 |
| SW049 | Q1 | 0.676 | 0.000 | 0.323 | Huyou 16 |
| SW051 | Q1 | 0.690 | 0.093 | 0.217 | Ningyou 12 |
| SW057 | Q1 | 0.702 | 0.240 | 0.058 | Qingyou 14 |
| SW059 | Q1 | 0.600 | 0.028 | 0.372 | Shilifeng |
| SW062 | Q1 | 0.721 | 0.175 | 0.104 | Wanyou 5 |
| SW063 | Q1 | 1,000 | 0.000 | 0.000 | Wanyouzao |
| SW064 | Q1 | 0.782 | 0.022 | 0.196 | Xiangyou 13 |
| SW065 | Q1 | 0.954 | 0.046 | 0.000 | Xiangyou 16 |
| SW069 | Q1 | 0.805 | 0.089 | 0.106 | Yuyou 5 |
| SW070 | Q1 | 0.899 | 0.024 | 0.077 | Yunyou 21-11 |
| SW072 | Q1 | 0.869 | 0.017 | 0.115 | Zhenzao1-2-30-6-6 |
| SW080 | Q1 | 0.652 | 0.221 | 0.127 | Zhongyou 821 |
| SW084 | Q1 | 0.617 | 0.061 | 0.322 | Xiangfu 112 |
| SW089 | Q1 | 0.689 | 0.311 | 0.000 | (Zhongshuang220 ×8424016) F8 |
| SW091 | Q1 | 0.660 | 0.016 | 0.324 | (Aisipeifax74-317)x3529-5 F8 |
| SW092 | Q1 | 0.607 | 0.208 | 0.185 | (D57 X Oro) x Youyan 2 F6 |
| SW093 | Q1 | 0.820 | 0.000 | 0.180 | (D57 X Oro) x 85-64 F9 |
| SW094 | Q1 | 0.631 | 0.000 | 0.369 | Youyan No.2 ×84-24016 |
| SW095 | Q1 | 0.886 | 0.000 | 0.113 | Xinongchangjiao ×((D57 X Oro)x 85-64)F7 |
| SW096 | Q1 | 0.763 | 0.062 | 0.175 | Altex ×74-317 F7 |
| SW097 | Q1 | 0.948 | 0.001 | 0.051 | AR 12 |

| | | | | | |
|---|---|---|---|---|---|
| SW098 | Q1 | 0.968 | 0.032 | 0.000 | Xiangnongyou-1 |
| SW099 | Q1 | 0.637 | 0.363 | 0.000 | Ganyou-5 |
| SW100 | Q1 | 0.888 | 0.018 | 0.094 | Zheyou 7 |
| SW101 | Q1 | 0.808 | 0.192 | 0.000 | Aijiazao |
| SW102 | Q1 | 0.941 | 0.059 | 0.000 | Chuanyoufuguo |
| SW103 | Q1 | 0.738 | 0.262 | 0.000 | Yunyou 5 |
| SW104 | Q1 | 0.836 | 0.164 | 0.000 | Tezao 16 |
| SW105 | Q1 | 0.837 | 0.163 | 0.000 | Puyou 3 |
| SW106 | Q1 | 0.710 | 0.280 | 0.010 | Huayou 5 |
| SW107 | Q1 | 1.000 | 0.000 | 0.000 | Ningyou 7 |
| SW108 | Q1 | 1.000 | 0.000 | 0.000 | Wanyou 5 |
| SW109 | Q1 | 0.733 | 0.267 | 0.000 | Huayou 3 |
| SW110 | Q1 | 0.823 | 0.037 | 0.141 | Zhenyou-1 |
| SW112 | Q1 | 0.769 | 0.231 | 0.000 | Tieganqing |
| SW113 | Q1 | 1.000 | 0.000 | 0.000 | Youyan 2 |
| SW116 | Q1 | 0.626 | 0.374 | 0.000 | Inbreed line |
| SW119 | Q1 | 0.772 | 0.228 | 0.000 | Inbreed line |
| SW121 | Q1 | 0.625 | 0.098 | 0.277 | Inbreed line |
| SW123 | Q1 | 0.975 | 0.025 | 0.000 | Inbreed line |
| SW125 | Q1 | 0.702 | 0.258 | 0.040 | Inbreed line |
| SW135 | Q1 | 0.609 | 0.391 | 0.000 | Inbreed line |
| SW136 | Q1 | 0.670 | 0.330 | 0.000 | Inbreed line |
| SW142 | Q1 | 0.652 | 0.348 | 0.000 | Inbreed line |
| SW145 | Q1 | 0.691 | 0.042 | 0.267 | Inbreed line |
| SW147 | Q1 | 0.788 | 0.212 | 0.000 | Inbreed line |
| SW153 | Q1 | 0.620 | 0.380 | 0.000 | Inbreed line |
| SW156 | Q1 | 0.794 | 0.020 | 0.186 | Inbreed line |
| SW158 | Q1 | 0.623 | 0.000 | 0.377 | Inbreed line |
| SW159 | Q1 | 0.656 | 0.000 | 0.344 | Inbreed line |
| SW160 | Q1 | 0.780 | 0.060 | 0.160 | Inbreed line |
| SW239 | Q1 | 0.827 | 0.173 | 0.000 | Inbreed line |
| SW205 | Q1 | 0.935 | 0.065 | 0.000 | Inbreed line |
| SW164 | Q1 | 0.938 | 0.062 | 0.000 | Inbreed line |
| SW169 | Q1 | 0.780 | 0.000 | 0.220 | Inbreed line |
| SW170 | Q1 | 0.821 | 0.000 | 0.179 | Inbreed line |
| SW173 | Q1 | 0.742 | 0.258 | 0.000 | Inbreed line |
| SW175 | Q1 | 0.681 | 0.078 | 0.241 | Inbreed line |
| SW186 | Q1 | 0.821 | 0.179 | 0.000 | Inbreed line |
| SW187 | Q1 | 0.997 | 0.003 | 0.000 | Inbreed line |
| SW191 | Q1 | 0.731 | 0.071 | 0.198 | Inbreed line |
| SW192 | Q1 | 0.946 | 0.011 | 0.043 | Inbreed line |
| SW193 | Q1 | 0.715 | 0.137 | 0.147 | Inbreed line |
| SW195 | Q1 | 0.727 | 0.031 | 0.242 | Inbreed line |
| SW206 | Q2 | 0.009 | 0.991 | 0.000 | Inbreed line |
| SW210 | Q2 | 0.205 | 0.691 | 0.104 | Inbreed line |
| SW214 | Q2 | 0.000 | 0.923 | 0.077 | Inbreed line |
| SW215 | Q2 | 0.000 | 0.923 | 0.077 | Inbreed line |
| SW224 | Q2 | 0.096 | 0.904 | 0.000 | Inbreed line |

| | | | | | |
|---|---|---|---|---|---|
| SW226 | Q2 | 0.114 | 0.886 | 0.000 | Inbreed line |
| SW235 | Q2 | 0.338 | 0.662 | 0.000 | Inbreed line |
| SW008 | Q2 | 0.344 | 0.626 | 0.030 | 1278-1-1-12 |
| SW035 | Q2 | 0.344 | 0.656 | 0.000 | Huxi |
| SW039 | Q2 | 0.000 | 0.824 | 0.176 | Huashuang 4 |
| SW043 | Q2 | 0.000 | 0.694 | 0.306 | Jingyou 4 -10 |
| SW044 | Q2 | 0.043 | 0.957 | 0.000 | Jujia 1 -5-13-3 |
| SW045 | Q2 | 0.000 | 0.861 | 0.139 | Keyou818-21 |
| SW054 | Q2 | 0.000 | 0.675 | 0.325 | Peng 201 |
| SW055 | Q2 | 0.000 | 1.000 | 0.000 | Pin 2-1-44-6-2 |
| SW056 | Q2 | 0.000 | 1.000 | 0.000 | Pinxi 3-4 |
| SW074 | Q2 | 0.000 | 1.000 | 0.000 | Zhongshuang 4 |
| SW078 | Q2 | 0.000 | 1.000 | 0.000 | Zhongshuang 8 |
| SW079 | Q2 | 0.000 | 1.000 | 0.000 | Zhongshuang 9 -11-9-101 |
| SW137 | Q2 | 0.001 | 0.676 | 0.323 | Inbreed line |
| SW138 | Q2 | 0.014 | 0.663 | 0.323 | Inbreed line |
| SW143 | Q2 | 0.057 | 0.943 | 0.000 | Inbreed line |
| SW149 | Q2 | 0.221 | 0.779 | 0.000 | Inbreed line |
| SW150 | Q2 | 0.321 | 0.679 | 0.000 | Inbreed line |
| SW157 | Q2 | 0.229 | 0.771 | 0.000 | Inbreed line |
| SW165 | Q2 | 0.011 | 0.989 | 0.000 | Inbreed line |
| SW167 | Q2 | 0.223 | 0.712 | 0.064 | Inbreed line |
| SW171 | Q2 | 0.168 | 0.832 | 0.000 | Inbreed line |
| SW176 | Q2 | 0.000 | 0.789 | 0.210 | Inbreed line |
| SW177 | Q2 | 0.161 | 0.650 | 0.189 | Inbreed line |
| SW178 | Q2 | 0.000 | 1.000 | 0.000 | Inbreed line |
| SW179 | Q2 | 0.001 | 0.909 | 0.091 | Inbreed line |
| SW213 | Q3 | 0.000 | 0.363 | 0.637 | Inbreed line |
| SW232 | Q3 | 0.299 | 0.000 | 0.701 | Inbreed line |
| SW002 | Q3 | 0.000 | 0.211 | 0.789 | (Youyan No.9 ×Zhongyouza No.3)F5 |
| SW007 | Q3 | 0.000 | 0.358 | 0.642 | 1226-7-6-9 |
| SW011 | Q3 | 0.009 | 0.360 | 0.631 | 268 Youcai-4 |
| SW014 | Q3 | 0.381 | 0.000 | 0.619 | 533-34-5-8 |
| SW015 | Q3 | 0.385 | 0.000 | 0.615 | 86178-2-34-1-4 |
| SW017 | Q3 | 0.001 | 0.159 | 0.840 | 914 Youcai-3 |
| SW024 | Q3 | 0.229 | 0.000 | 0.771 | C182-2-34-8-5 |
| SW025 | Q3 | 0.382 | 0.000 | 0.618 | C384-4-38-1-21 |
| SW029 | Q3 | 0.133 | 0.000 | 0.867 | T4-2 Nucleic male sterulity xuan a-7B-2-6-4 |
| SW050 | Q3 | 0.366 | 0.001 | 0.633 | Huyou 18 |
| SW053 | Q3 | 0.214 | 0.143 | 0.643 | Ningza 3 -4-33-5-1-2 |
| SW058 | Q3 | 0.000 | 0.000 | 1,000 | Shanyou 8- 5-6-13-2 |
| SW061 | Q3 | 0.255 | 0.124 | 0.621 | Suyou 3 |
| SW068 | Q3 | 0.002 | 0.245 | 0.753 | Yuyou 2 |
| SW081 | Q3 | 0.069 | 0.000 | 0.931 | Zhongyouza 1 -51-35-5-17 |
| SW115 | Q3 | 0.333 | 0.000 | 0.667 | Inbreed line |
| SW120 | Q3 | 0.159 | 0.132 | 0.709 | Inbreed line |
| SW148 | Q3 | 0.000 | 0.362 | 0.638 | Inbreed line |
| SW152 | Q3 | 0.001 | 0.327 | 0.672 | Inbreed line |

| | | | | | |
|---|---|---|---|---|---|
| SW162 | Q3 | 0.001 | 0.172 | 0.827 | Inbreed line |
| SW163 | Q3 | 0.000 | 0.176 | 0.823 | Inbreed line |
| SW166 | Q3 | 0.033 | 0.176 | 0.791 | Inbreed line |
| SW168 | Q3 | 0.000 | 0.363 | 0.636 | Inbreed line |
| SW180 | Q3 | 0.298 | 0.055 | 0.646 | Inbreed line |
| SW189 | Q3 | 0.287 | 0.001 | 0.712 | Inbreed line |
| SW127 | Q3 | 0.002 | 0.000 | 0.998 | winter |
| SW129 | Q3 | 0.269 | 0.000 | 0.731 | spring |
| SW131 | Q3 | 0.022 | 0.000 | 0.978 | spring |
| SW133 | Q3 | 0.260 | 0.000 | 0.740 | spring |
| SW005 | Mixed | 0.290 | 0.141 | 0.569 | [Zhongshuang No.4 ×(B5 ǯ 76)]F6 |
| SW006 | Mixed | 0.283 | 0.241 | 0.477 | {[(Zhengchun ×Qin2-2)×Hui 3]×Hui 4}F6 |
| SW009 | Mixed | 0.160 | 0.402 | 0.438 | 1492-1-5-1-3 |
| SW010 | Mixed | 0.418 | 0.582 | 0.000 | 1492-1-7-1 |
| SW022 | Mixed | 0.343 | 0.368 | 0.289 | C148-4-36-8-10 |
| SW028 | Mixed | 0.449 | 0.135 | 0.416 | R7-14 Yue 7-8 Xuan-7-34-6-5-23 |
| SW030 | Mixed | 0.587 | 0.216 | 0.197 | Chuanyou 18 |
| SW032 | Mixed | 0.360 | 0.404 | 0.236 | Fuyou No.2-15-3 |
| SW037 | Mixed | 0.209 | 0.394 | 0.397 | Huashuang 3 |
| SW040 | Mixed | 0.455 | 0.545 | 0.000 | Huashuang 5 |
| SW041 | Mixed | 0.553 | 0.447 | 0.000 | Hui 110-1-40-7-7-2 |
| SW052 | Mixed | 0.418 | 0.257 | 0.326 | Ningyou 14 |
| SW060 | Mixed | 0.103 | 0.418 | 0.479 | Suyou 5 -1-8-6 |
| SW066 | Mixed | 0.474 | 0.200 | 0.325 | Xiangyou 17 |
| SW067 | Mixed | 0.575 | 0.086 | 0.339 | Yangyou 6 |
| SW071 | Mixed | 0.573 | 0.416 | 0.011 | Zheshuang 72 |
| SW073 | Mixed | 0.518 | 0.472 | 0.009 | Huangzifenxuan-5-3 |
| SW075 | Mixed | 0.513 | 0.487 | 0.000 | Zhongshuang 5 |
| SW076 | Mixed | 0.506 | 0.494 | 0.000 | Zhongshuang 6 |
| SW077 | Mixed | 0.432 | 0.568 | 0.000 | Zhongshuang 7 |
| SW082 | Mixed | 0.494 | 0.03 | 0.476 | Zhongyouza 2 Restorer -20-3-15-1 |
| SW083 | Mixed | 0.521 | 0.479 | 0.000 | Zhongshuang 10 |
| SW085 | Mixed | 0.000 | 0.408 | 0.591 | Xiwang 106 |
| SW086 | Mixed | 0.471 | 0.150 | 0.379 | 9636 Xuan |
| SW087 | Mixed | 0.419 | 0.001 | 0.581 | Ganyou 17 xuan |
| SW088 | Mixed | 0.516 | 0.484 | 0.000 | Huyou 17 |
| SW090 | Mixed | 0.512 | 0.064 | 0.424 | 97V38×(72-2Xlijnte) F6 |
| SW111 | Mixed | 0.457 | 0.298 | 0.245 | Huanshuang 1 |
| SW114 | Mixed | 0.438 | 0.000 | 0.562 | Inbreed line |
| SW117 | Mixed | 0.431 | 0.078 | 0.491 | Inbreed line |
| SW122 | Mixed | 0.555 | 0.219 | 0.226 | Inbreed line |
| SW124 | Mixed | 0.518 | 0.141 | 0.341 | Inbreed line |
| SW134 | Mixed | 0.248 | 0.268 | 0.484 | Inbreed line |
| SW139 | Mixed | 0.506 | 0.494 | 0.000 | Inbreed line |
| SW140 | Mixed | 0.415 | 0.585 | 0.000 | Inbreed line |
| SW141 | Mixed | 0.457 | 0.543 | 0.000 | Inbreed line |
| SW144 | Mixed | 0.529 | 0.416 | 0.055 | Inbreed line |
| SW146 | Mixed | 0.547 | 0.000 | 0.453 | Inbreed line |

| | | | | | |
|---|---|---|---|---|---|
| SW151 | Mixed | 0.377 | 0.100 | 0.523 | Inbreed line |
| SW154 | Mixed | 0.570 | 0.078 | 0.352 | Inbreed line |
| SW155 | Mixed | 0.517 | 0.483 | 0.000 | Inbreed line |
| SW161 | Mixed | 0.486 | 0.233 | 0.280 | Inbreed line |
| SW172 | Mixed | 0.482 | 0.030 | 0.488 | Inbreed line |
| SW174 | Mixed | 0.147 | 0.486 | 0.366 | Inbreed line |
| SW181 | Mixed | 0.312 | 0.250 | 0.439 | Inbreed line |
| SW182 | Mixed | 0.001 | 0.541 | 0.458 | Inbreed line |
| SW183 | Mixed | 0.162 | 0.475 | 0.363 | Inbreed line |
| SW184 | Mixed | 0.419 | 0.337 | 0.244 | Inbreed line |
| SW185 | Mixed | 0.53 | 0.329 | 0.141 | Inbreed line |
| SW188 | Mixed | 0.486 | 0.276 | 0.238 | Inbreed line |
| SW190 | Mixed | 0.570 | 0.091 | 0.339 | Inbreed line |
| SW194 | Mixed | 0.585 | 0.000 | 0.415 | Inbreed line |
| SW200 | Mixed | 0.411 | 0.589 | 0.000 | Inbreed line |
| SW209 | Mixed | 0.426 | 0.574 | 0.000 | Inbreed line |
| SW211 | Mixed | 0.142 | 0.292 | 0.565 | Inbreed line |
| SW212 | Mixed | 0.490 | 0.448 | 0.062 | Inbreed line |
| SW126 | Mixed | 0.559 | 0.058 | 0.383 | winter |
| SW130 | Mixed | 0.599 | 0.000 | 0.401 | spring |
| SW220 | Mixed | 0.457 | 0.513 | 0.030 | Inbreed line |
| SW240 | Mixed | 0.484 | 0.187 | 0.329 | Inbreed line |

**Additional file 3:** Integrated map showing genomic positions of SSR and SNP marker sequences from QTL for seed glucosinolate (GLS) and erucic acid content, identified by BLAST searches onto A- and C-subgenome chromosomes from the *Brassica napus* Darmor-Bzh reference genome. (on the attached CD ROM)

**Additional file 4:** Details of synteny alignments for QTL positions for seed glucosinolate (chromosomes A02 and A09) and erucic acid (A08) into homoeologous regions of chromosomes C02, C09 and C03, respectively.

| GLS and erucic acid QTL regions | | | Sequence alignment to C02, C09 and C03 homoeologous region | | |
|---|---|---|---|---|---|
| **A-subgenome homoeologue** | **Seq_start position (bp)** | **Seq_end position (bp)** | **C-subgenome homoeologue** | **Seq_start position (bp)** | **Seq_end position (bp)** |
| A02 | 19715788 | 23166916 | C02 | 37490998 | 44879337 |
| A09 | 777017 | 3828580 | C09 | 290804 | 5121974 |
| A08 | 9513648 | 11743490 | C03 | 53509468 | 58400326 |

**Additional file 5:** Detailed information on putative functional candidate genes and LD (haplotype block) analysis within the investigated QTL intervals for seed glucosinolate (GLS) and erucic acid content.

| Homologous GLS QTL region (A02-C02 and A09-C09) | Homologous erucic acid QTL region (A08-C03) | ID (B.napus) | Chr. | Seq.start | Seq.end | Seq.Description | Genes coverage size in QTL region (Mb) (form first to last gene in QTL region) | Haplotype block (D') | | Ldheat-map ($r^2$) | Best *A.thaliana* hit | Best *A.thaliana* hit annotation | Best *A.thaliana* hit reference |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| GLS QTL region (A02-C02 ) | | GSBRNA2G093766001 | A02 | 21283484 | 21286561 | monodehydroascorbate reductase 4 | 0.50 | | | | AT3G27820 | glucosinolate biosynthetic process | [74] |
| | | GSBRNA2G093756001 | A02 | 21342180 | 21344807 | nodulin 21 -like transporter family protein | | | | | AT3G28080 | glucosinolate biosynthetic process | [74] |
| | | GSBRNA2G080110001 | A02 | 21559193 | 21560606 | aig2-like protein | | | | | AT3G28940 | glucosinolate biosynthetic process | [74] |
| | | GSBRNA2G091634001 | A02 | 21786532 | 21787583 | kelch repeat-containing protein | | | | | AT5G48180 | glucosinolate catabolic process | [71] |
| | | GSBRNA2G108334001 | C02 | 38996854 | 38999944 | cytochrome p450 79c2 | 5.71 | √ | 38 | * 4.54 | AT1G58260 | glucosinolate biosynthetic process | |
| | | GSBRNA2G117557001 | C02 | 40162109 | 40164375 | nodulin 21 -like transporter family protein | | √ | 81 | | AT3G28080 | glucosinolate biosynthetic process | [74] |
| | | GSBRNA2G117640001 | C02 | 40591848 | 40593128 | aig2-like protein | | * | 326 | | AT3G28940 | glucosinolate biosynthetic process | [74] |
| | | GSBRNA2G117642001 | C02 | 40600753 | 40602211 | aig2-like protein | | | | | AT3G28941 | glucosinolate biosynthetic process | [74] |
| | | GSBRNA2G120710001 | C02 | 41387856 | 41389461 | kelch repeat-containing protein | | | | | AT5G48180 | glucosinolate catabolic process | [71] |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | GSBRNA2G00071732001 | C02 | 44597927 | 44599979 | methylthioalkylmalate synthase | | | | | AT5G23020 | glucosinolate biosynthetic process | [82] |
| | GSBRNA2G00071741001 | C02 | 44703129 | 44706876 | myb transcription factor | | √ | 157 | | AT5G60890 | indole glucosinolate biosynthetic process | [72] |
| GLS QTL region (A09-C09) | GSBRNA2G00094719001 | A09 | 2698974 | 2700122 | myb34 | | √ | 24 | | AT5G60890 | indole glucosinolate biosynthetic process | [72] |
| | GSBRNA2G00094633001 | A09 | 3074814 | 3077979 | 1-deoxy-d-xylulose 5-phosphate reductoisomerase | | | | | AT5G62790 | glucosinolate biosynthetic process | [74] |
| | GSBRNA2G00005776001 | A09 | 3400066 | 3406197 | cellulose synthase catalytic subunit | 1.12 | | | | AT4G39350 | glucosinolate biosynthetic process | [74] |
| | GSBRNA2G00005668001 | A09 | 3815355 | 3817167 | 33 kda polypeptide of oxygen-evolving complex in photosystem ii | | | | | AT3G50820 | glucosinolate biosynthetic process | [74] |
| | GSBRNA2G00146147001 | C09 | 3099232 | 3099846 | BnaA.HAG1/myb28 | | | | | AT5G61420 | glucosinolate biosynthetic process | [73],[75],[74],[81] |
| | GSBRNA2G00146148001 | C09 | 3099905 | 3100976 | mby transcription factor | | | | | AT5G61420 | glucosinolate biosynthetic process | [73],[75],[74],[81] |
| | GSBRNA2G00146187001 | C09 | 3294955 | 3296509 | isopropyl malate isomerase large subunit 1 | 1.98 | √ | 31 | * 0.59 | AT4G13430 | glucosinolate biosynthetic process | [74],[80] |
| | GSBRNA2G00134656001 | C09 | 3680496 | 3681837 | pyruvate kinase | | * | 113 | | AT5G08570 | glucosinolate biosynthetic process | [74] |

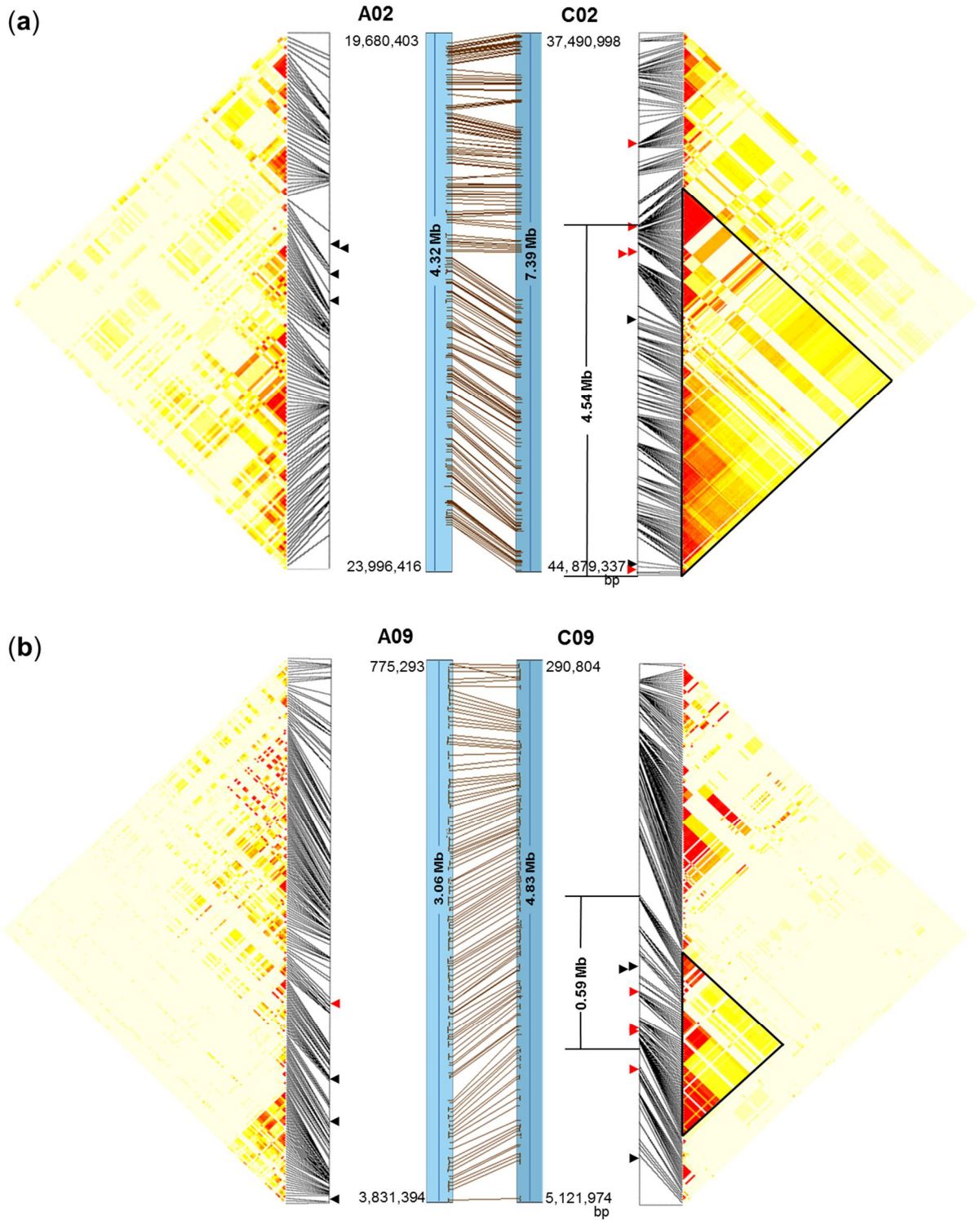
| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | GSBRNA2G00134657001 | C09 | 3681893 | 3682874 | pyruvate kinase | | | | | AT5G08570 | glucosinolate biosynthetic process | [74] |
| | | GSBRNA2G00134740001 | C09 | 4010674 | 4016818 | cellulose synthase catalytic subunit | | √ | 27 | | AT4G39350 | glucosinolate biosynthetic process | [74] |
| | | GSBRNA2G00134966001 | C09 | 5078986 | 5080826 | 33 kda polypeptide of oxygen-evolving complex in photosystem ii | | | | | AT3G50820 | glucosinolate biosynthetic process | [74] |
| | Erucic acid QTL region (A08-C03 ) | GSBRNA2G00126652001 | A08 | 10187601 | 10189121 | BnaA.FAE1 | | | | | AT4G34520 | very long-chain /fatty acid biosynthetic process | [76] |
| | | GSBRNA2G00126651001 | A08 | 10193689 | 10195113 | 3-ketoacyl-synthase 4-like | | | | | AT2G16280 | very long-chain /fatty acid biosynthetic process | [77], [74] |
| | | GSBRNA2G00021891001 | A08 | 10994261 | 10997355 | 3-hydroxyisobutyryl - hydrolase-like protein mitochondrial-like | | | | | AT4G31810 | fatty acid beta-oxidation | - |
| | | GSBRNA2G00021869001 | A08 | 11101075 | 11102843 | at4g31410 f8f16_230 | 1.93 | | | * 0.30 | AT4G31410 | fatty acid beta-oxidation | [74] |
| | | GSBRNA2G00021863001 | A08 | 11123283 | 11125509 | Encodes 20S proteasome subunit PBA1 | | * | 137 | | AT4G31300 | fatty acid beta-oxidation | [74] |
| | | GSBRNA2G00021832001 | A08 | 11285185 | 11287741 | chloroplast omega-6 fatty acid desaturase | | √ | 13 | | AT4G30950 | unsaturated/ fatty acid biosynthetic process | [78],[74] |

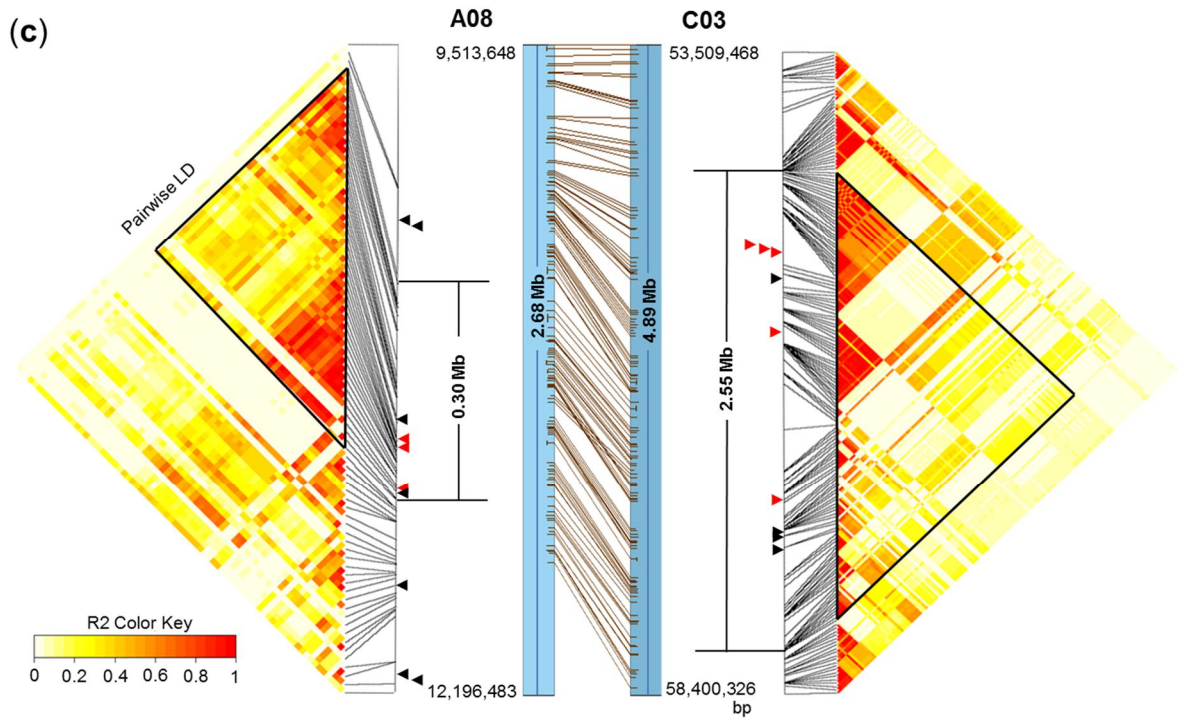
| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | GSBRNA2G00021829001 | A08 | 11293088 | 11296404 | chloroplast omega-6 fatty acid desaturase | | | | | | AT4G30950 | unsaturated/ fatty acid biosynthetic process | [78],[74] |
| | GSBRNA2G00078793001 | A08 | 11698681 | 11701496 | 26s proteasome subunit 4 | | | | | | AT4G29040 | fatty acid beta-oxidation | [74] |
| | GSBRNA2G00078880001 | A08 | 12104471 | 12105358 | kua-ubiquitin conjugating enzyme hybrid localisation domain-containing protein | | | | | | AT4G27030 | unsaturated fatty acid biosynthetic process | [70] |
| | GSBRNA2G00078884001 | A08 | 12117243 | 12118100 | kua-ubiquitin conjugating enzyme hybrid localisation domain-containing protein | | | | | | AT4G27030 | unsaturated fatty acid biosynthetic process | [70] |
| | GSBRNA2G00072955001 | C03 | 54924531 | 54930430 | tyrosine transaminase like protein | | | | | | AT4G23600 | unsaturated fatty acid biosynthetic process | [74] |
| | GSBRNA2G00072956001 | C03 | 54930616 | 54931447 | tyrosine transaminase like protein | | * | 477 | | | AT4G23600 | unsaturated fatty acid biosynthetic process | [74] |
| | GSBRNA2G00072957001 | C03 | 54931522 | 54932070 | tyrosine transaminase like protein | 2.54 | | | * | 2.55 | AT4G23600 | unsaturated fatty acid biosynthetic process | [74] |
| | GSBRNA2G00095710001 | C03 | 55173180 | 55174831 | Peroxisomal nad-malate dehydrogenase 1 | | | | | | AT2G22780 | fatty acid catabolic process/regulation of fatty acid beta-oxidation | [74],[79] |
| | GSBRNA2G00095638001 | C03 | 55684252 | 55685679 | BnaC.FAE1 | | √ | 421 | | | AT4G34520 | very long-chain /fatty acid biosynthetic process | [76] |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | GSBRNA2G00068715001 | C03 | 57058069 | 57061103 | 3-hydroxyisobutyryl - hydrolase-like protein mitochondrial-like | | √ | 281 | | AT4G31810 | fatty acid beta-oxidation | - |
| | | GSBRNA2G00062078001 | C03 | 57203619 | 57205528 | at4g31410 f8f16_230 | | | | | AT4G31410 | fatty acid beta-oxidation | [74] |
| | | GSBRNA2G00062096001 | C03 | 57281968 | 57284222 | 20s proteasome subunit pba1 | | | | | AT4G31300 | fatty acid beta-oxidation | [74] |
| | | GSBRNA2G00062114001 | C03 | 57461243 | 57463872 | chloroplast omega-6 fatty acid desaturase | | | | | AT4G30950 | unsaturated/ fatty acid biosynthetic process | [78],[74] |

" √" respresent the gene in haplotype block.
"*" respresent the adjacent two or more genes in the same block.

**Additional file 6:**




(c)
A08
C03
9,513,648
53,509,468
Pairwise LD
0.30 Mb
2.68 Mb
4.89 Mb
2.55 Mb
R2 Color Key
0 0.2 0.4 0.6 0.8 1
12,196,483
58,400,326
bp

**Appendix III:** Electronic supplementary materials from Qian et al. (2016)

**Figure S1**

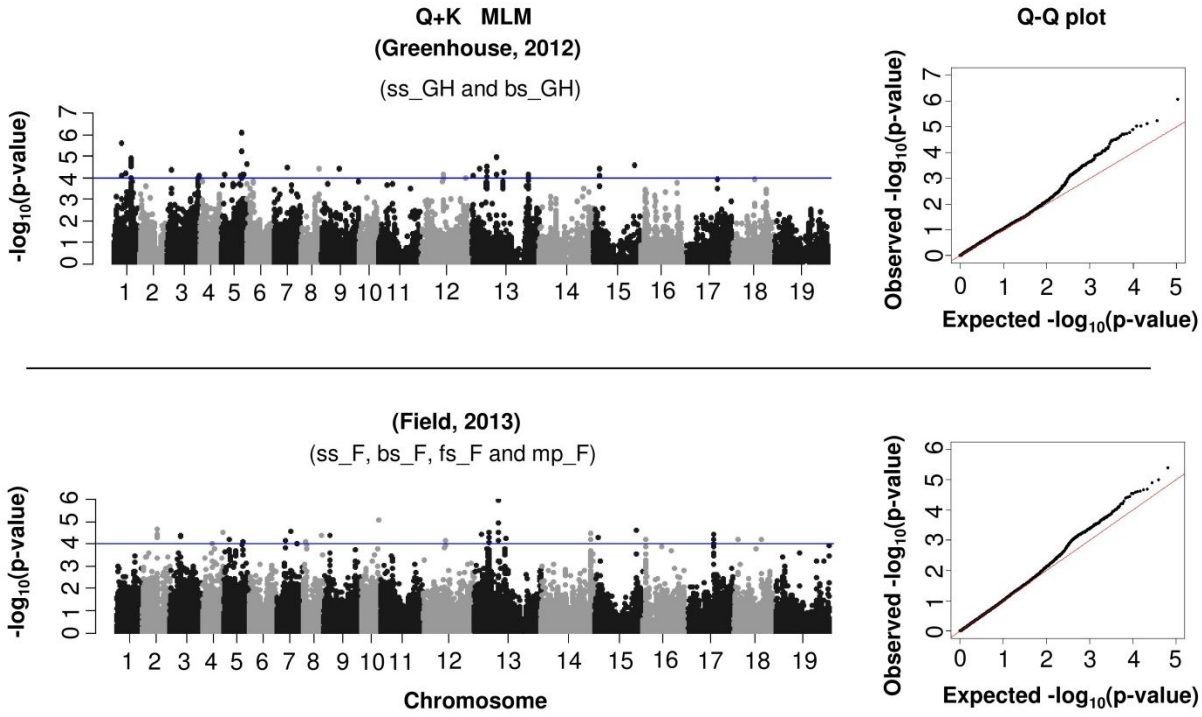


**Figure S2**

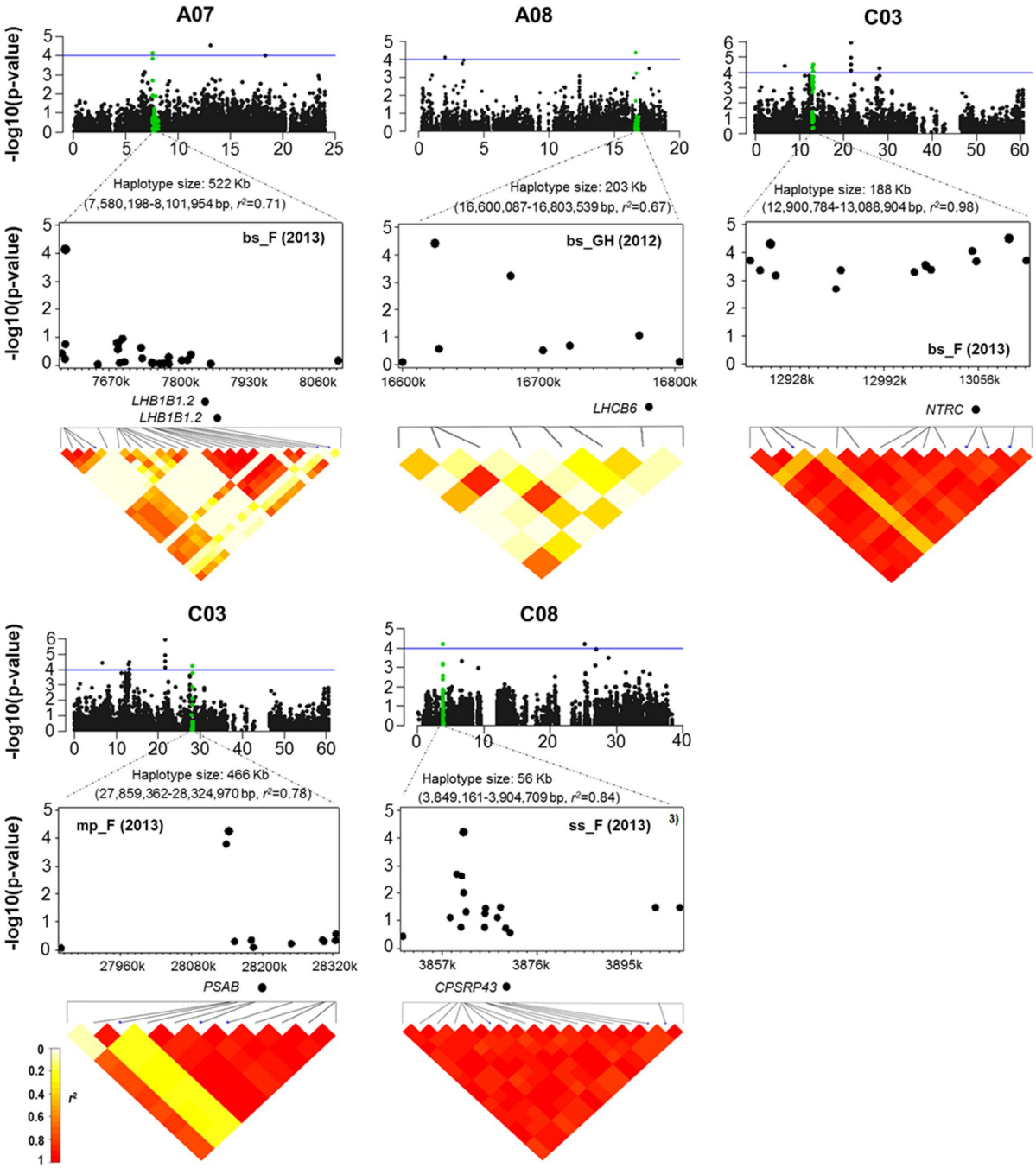

A07
A08
C03
-log10(p-value)
Haplotype size: 522 Kb
(7,580,198-8,101,954 bp, r²=0.71)
bs_F (2013)
Haplotype size: 203 Kb
(16,600,087-16,803,539 bp, r²=0.67)
bs_GH (2012)
Haplotype size: 188 Kb
(12,900,784-13,088,904 bp, r²=0.98)
bs_F (2013)
LHB1B1.2
LHB1B1.2
LHCB6
NTRC
C03
C08
Haplotype size: 466 Kb
(27,859,362-28,324,970 bp, r²=0.78)
mp_F (2013)
Haplotype size: 56 Kb
(3,849,161-3,904,709 bp, r²=0.84)
ss_F (2013) 3)
PSAB
CPSRP43
r²

**Figure S3**

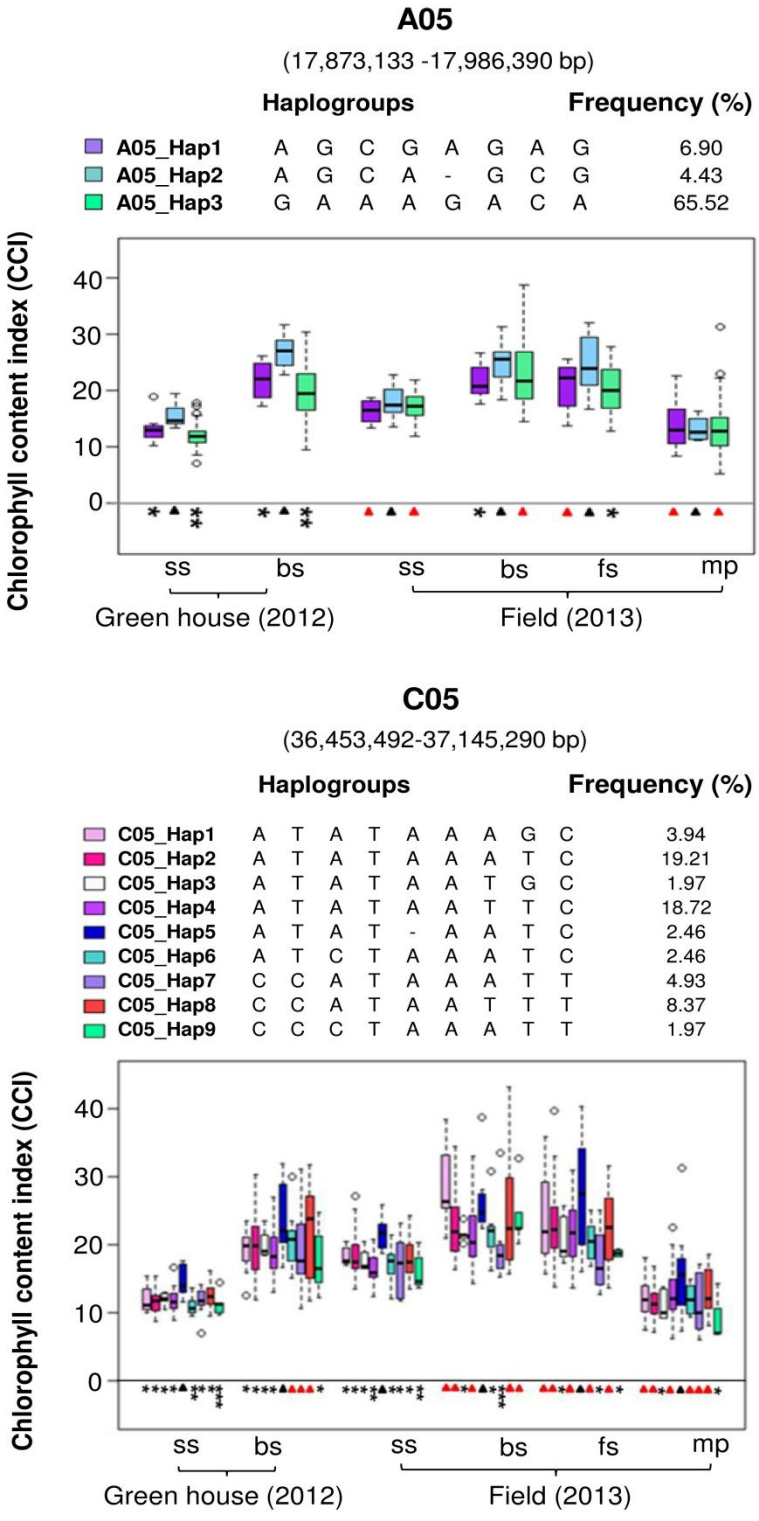

A05
(17,873,133 -17,986,390 bp)
Haplogroups
Frequency (%)
A05_Hap1 A G C G A G A G 6.90
A05_Hap2 A G C A - G C G 4.43
A05_Hap3 G A A A G A C A 65.52
Chlorophyll content index (CCI)
0
10
20
30
40
ss
bs
ss
bs
fs
mp
Green house (2012)
Field (2013)
C05
(36,453,492-37,145,290 bp)
Haplogroups
Frequency (%)
C05_Hap1 A T A T A A A G C 3.94
C05_Hap2 A T A T A A A T C 19.21
C05_Hap3 A T A T A A T G C 1.97
C05_Hap4 A T A T A A T T C 18.72
C05_Hap5 A T A T - A A T C 2.46
C05_Hap6 A T C T A A A T C 2.46
C05_Hap7 C C A T A A A T T 4.93
C05_Hap8 C C A T A A T T T 8.37
C05_Hap9 C C C T A A A T T 1.97
Chlorophyll content index (CCI)
0
10
20
30
40
ss
bs
ss
bs
fs
mp
Green house (2012)
Field (2013)

**Figure S4**

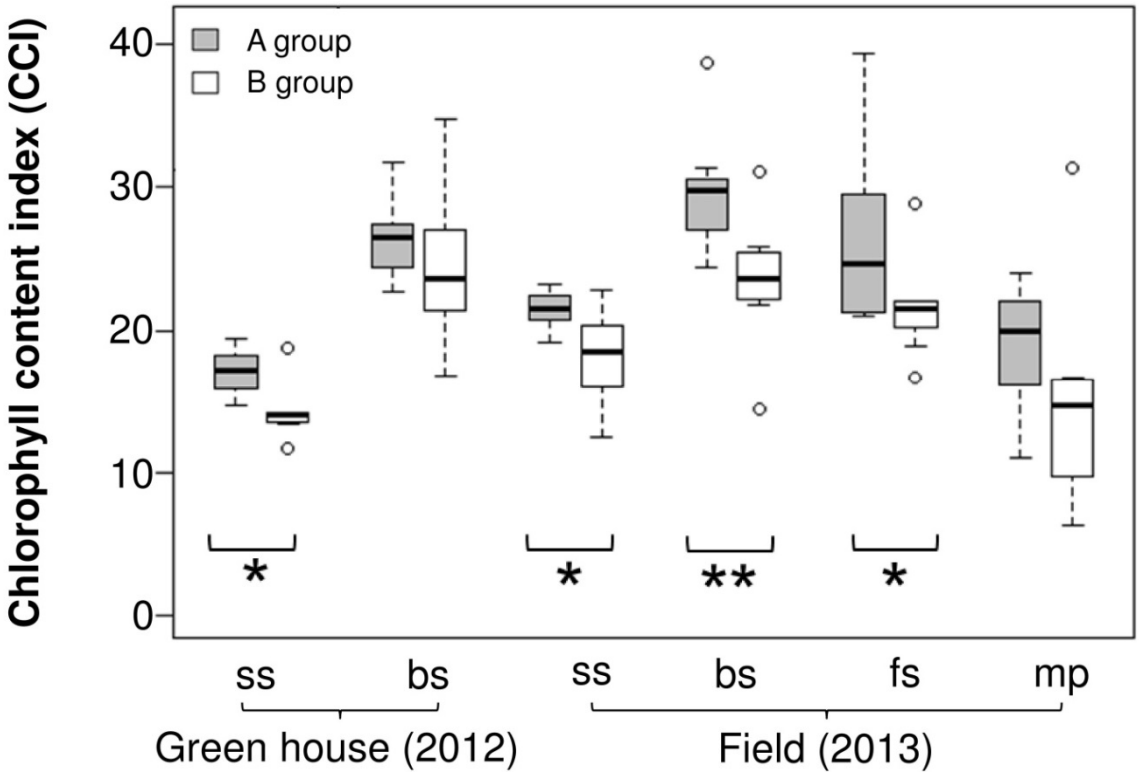

Chlorophyll content index (CCI)
40
30
20
10
0
A group
B group
*
*
**
*
ss
bs
ss
bs
fs
mp
Green house (2012)
Field (2013)

**Figure S5**

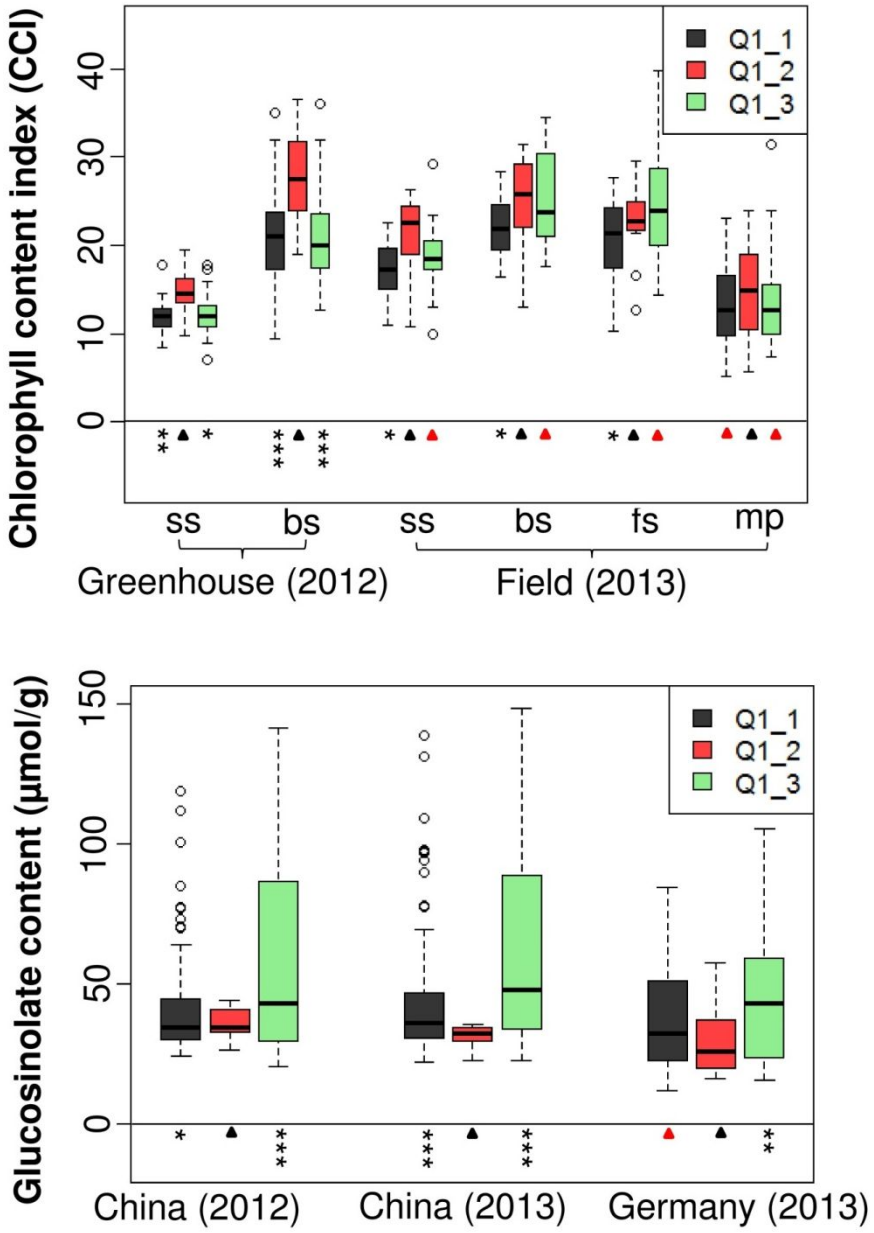

Chlorophyll content index (CCI)
0
10
20
30
40
Q1_1
Q1_2
Q1_3
ss
bs
ss
bs
fs
mp
Greenhouse (2012)
Field (2013)
Glucosinolate content (µmol/g)
0
50
100
150
Q1_1
Q1_2
Q1_3
China (2012)
China (2013)
Germany (2013)

**Table S1** Source, population structure, heritability (leaf CCI and seed GSL) and associated haplogroups in 203 Chinese semi-winter rapeseed accessions. (on the attached CD ROM)

**Table S2** Detailed description of five haplotype regions significantly associated with leaf CCI. (on the attached CD ROM)

**Table S3** Comparative analysis of haplogroups related to leaf CCI, along with gene content in homologous haplotype regions on chromosomes A05 and C05. (on the attached CD ROM)

**Table S4** Comparative analysis of six haplogroups related to leaf CCI, along with gene information in the A02 haplotype region. (on the attached CD ROM)

**Table S5** Comparative analysis of three haplogroups corresponding to leaf CCI in BnaA02g20650D gene region. (on the attached CD ROM)

**Table S6** Comparative analyses of five haplogroups corresponding to leaf CCI and seed GSL and gene information in chromosome A01 co-selection haplotype region. (on the attached CD ROM)

**Table S7** Comparative analyses of three subgroups corresponding to leaf CCI and seed GSL. (on the attached CD ROM)

# List of Abbreviations

| | |
|---|---|
| bp | base pairs |
| CAPS | Cleaved Amplified Polymorphic Sequences |
| CCI | Chloropyll content index |
| CIM | Composite interval mapping |
| cM | centi Morgans |
| DH | Doubled haploid |
| GLM | General linear model |
| GSL | Glucosinolate content |
| GWAS | Genome-wide association studies |
| LD | Linkage disequilibrium |
| MAF | Minor allele frequency |
| NIRS | Near-infrared reflectance spectroscopy |
| PCA | Principal component analysis |
| PIC | Polymorphic information content |
| QTG | Quantitative trait-related genes |
| QTL | Quantitative trait locus |
| SNP | Single nucleotide polymorphism |
| UPGMA | Unweighted pair group matrix algorithm |

# Declaration

I, Lunwen Qian, declare that the dissertation and work presented in it are my own and has been generated by me as the result of my own original research. I have not copied from any other students' work or from any other sources. I have indicated that some parts of this thesis have previously been published. Wherever contributions of others are involved, every effort is made to indicate this clearly, with due reference to the literature, and acknowledgement of collaborative research and discussions. I have followed the principles of good scientific practice as defined in the "Statutes of the Justus Liebig University for the Safeguarding of Good Scientific Practice".

# Acknowledgments

Firstly, I would like to express my sincere gratitude to my advisor Prof. Dr. Rod Snowdon who gave me the opportunity to work in his lab as a PhD student. He gave me intellectual freedom in my work, supported my participation at various conferences, engaged me in developing new ideas, and demanded a high quality of work throughout all my endeavors. His guidance helped me all the time during my research and the writing of this thesis. His academic achievements, diligent research style, deeply affects my work and my life now and in the future.

I am also thankful to my second supervisor Prof. Matthias Frisch, Institute for Agronomy and Plant Breeding II, Justus Liebig University, Giessen, Germany for evaluating my work and his valuable suggestions to improve the manuscript.

I am also thankful to Prof. Dr. Dr. h.c. Wolfgang Friedt for his guidance and support during my stay in Giessen.

My sincere thanks go to Prof. Dr. Wei Qian who helped me to search for a PhD position in Germany and provided me with the opportunity to work and learn abroad. I am also grateful for his patience, understanding and close supervision during my postgraduate education in China.

The work described in this thesis was accomplished with the help and support of fellow lab-mates and collaborators. I would like to thank Dr. Benjamin Wittkop for his help in phenotyping and field trials. Thanks to Dr. Wubishet Abebe Bekele, Dr. Birgit Samans and Dr. Christian Obermeier for their help in data analysis. My colleagues, Stefanie Lück, Kai Voss-Fels and Habib Jan, thank you all for providing an enriching environment for English speaking. Thanks to Habib Jan and Johann Krautz for proof reading my thesis manuscript. I want to thank technical assistants Stavros Tzigos, Liane Renno, Petra Degen, Petra Kretschmer, Birgit Keiner and Annette Plank who helped me during my experiments in the lab and in the greenhouse.

My deep appreciation goes to my parents. Without their encouragement and assistance, I would never have been able to finish my studies.

Finally, I am very grateful to the China Scholarship Council (CSC) for financing my studies.