Exploring a *Brassica rapa* RIL population: Phenotypic and genotypic analysis

Master Thesis

Inge van Oorschot 870524-623-020

Supervisor: Hedayat Bagheri MSc

Examinors: Dr. Mark Aarts
Prof. Dr. Maarten Koornneef

Laboratory of Genetics, Wageningen University

Date: December 2009

Table of contents

Preface

This master thesis will give information about the aims and activities of my research in the Laboratories of Genetics and partially of plant Plant Breeding. In this project, a RIL population of 160 lines of *Brassica rapa* will be phenotyped for 12 morphological traits. SSR markers will be added to the RIL linkage map to extend the map and to anchor the linkage map to reference genetic map of *B. rapa*. QTL analysis will be performed to detect QTLs for the studied traits.

The reason for choosing this subject has been my long lasting interest in plants and the potential they have to serve as a source of food or beauty. With my educational background in ecology, doing a thesis in the worlds of genetics and plant breeding was not the easiest path to take. Yet, I feel that by performing my research in the Lab of Genetics I have gained much insight into this field. Phenotyping the plants and doing markers analysis with different kinds of markers allowed me to become familiar with some techniques I had not used before. I have learned that with doing research not everything will go as planned, to have patience when doing experiments and to think them out carefully before starting.
I enjoyed my stay in the Lab of Genetics and was lucky to be in such a friendly group of people with different cultural and scientific backgrounds. I would like to thank the people that helped me in any way during my thesis. Much thanks goes out to my supervisor Hedayat Bagheri, your knowledge and advice has helped me a lot. Also I would like to thank the experts on this topic, Mark Aarts, Maarten Koornneef and Joost Keurentjes, your advice and help was very welcome. Working in the laboratory of genetics proved to be much to my liking, also thanks to the fun times or good advice from technical assistants Corrie Hanhart and Frank Beckers. During my time in the laboratory of plant breeding Koen Pelgrom, Petra van den Berg, Gert van Arkel and Johan Bucher were kind enough to help me when needed with different problems during marker analysis. Thank you all, I really appreciate your help. Lastly I would like to say thanks to my fellow master students Joseline Houwman, Thomas Geydan-Riveira, Gerjon Ikink, and Florien Gorter who helped me a lot by creating a fun working-environment.

Abstract

A recombinant inbred population (F7) of 160 *B. rapa* lines was made by repeated selfing F2 plants derived from a cross between genotype L58, which is a rapid-flowering, self-compatible Caixin line, and R-o-18, a self-compatible Indian doubled haploid spring oil line. This RIL population was used to create a genetic linkage map with 127 markers, including 85 SNP markers, 25 AFLP markers and 17 publicly available SSR markers. 20.3% of markers showed distortion (≥2:1) based on a $\chi 2$ test for goodness-of-fit to the expected 1:1 mendelian segregation ratio. 4.9% of them are highly distorted (≥3:1). An overall preference for the R-o-18 genome was detected. This population showed heterozygosity for 1.3% of all loci. A linkage map was made in which all 10 linkage groups of *B. rapa* could be identified and anchored to the reference linkage map with publicly available SSR markers. The total map covers 585.5 cM and has an average resolution of 4.6 cM. QTL analysis revealed 5 major and 7 minor QTLs for 7 traits, including plant height, number of seeds per silique, silique height top, number of leaves in first 15 cm, silique ripening time, seed coat colour and vivipary. This study shows the usefulness of the *B. rapa* RIL population and corresponding linkage map for future studies on the genetics of breeding related traits.

1. Introduction

1.1 Background *Brassica rapa*

Brassica rapa belongs to the Brassicaceae (Cruciferae) family. This family includes the well-known species *Arabidopsis thaliana*, which is a genetic model organism. The *Brassica* genus contains many economically important species, amongst others the three elementary diploid species *B. rapa* (2n = 20), *B. nigra* (2n = 16) and *B. oleracea* (2n = 18). *B. rapa* can be divided into three different types based on morphology. Due to division in cultivation there are now turnip, oleiferous and leafy types (Zhao, 2007).

B. rapa has been used as a crop for many centuries from the Western Mediterranean region to Central Asia, where it is still present. Several studies done on morphology, geographic distribution, isozymes and molecular data show that *B. rapa* has two places of origin (Gomez Campo, 1999). Europe is considered to be the place where oil and turnip types originated and East Asia is considered to be the place where Indian oil types and Chinese leafy types originated. There are also other cultivars of *B. rapa* which probably originated from different morphotypes that lie within the two centers of origin and thereafter evolved apart from one another (Zhao, 2007).

Although there are different morphotypes, *B. rapa* generally has smooth green leaves of which the upper ones partially clasp the stem. The stem is fairly branched, although the amount of branching depends on variety and environmental conditions. Branches originate in the axils of the upper leaves, and end in an inflorescence. Lower leaves are somewhat toothed or lobed and petioled, while upper leaves are sessile, elongated and lance-shaped. The inflorescence is an elongated raceme with yellow flowers growing clustered at the top (Downey *et al.*, 1980; CFIA, 1999).

As a crop, *B. rapa* provides food for people in many countries around the world, which makes it a crop of economic importance. Research performed in several international research groups shows that the *Brassica* vegetables have notably high levels of vitamins A, C and E, folate and potassium, which indicates that consumption of *B. rapa* products can be beneficial for peoples' health (http://www.brassica.info/).
B. rapa gives rise to many different vegetables and oil products, especially if crossed with related species (Zhao, 2007). The crop was the first *Brassica* species to be domesticated and so it has been cultivated for many centuries to this day forth (Gomez Campo, 1999). Its long cultivation history has allowed people to select for many different traits causing many varieties to arise, which vary in morphology as well as genetics (Zhao, 2007).
B. rapa has three morphotypes, of which mostly the oleiferous and leafy types are still cultivated. The oleiferous type has seeds which contain oils and proteins. The oil can be used as biofuel, as cooking oil, in salads or for making margarine. The seeds can also be grinded to get meal containing valuable protein to use as livestock feed (Zhao, 2007). The other morphotype has a more leafy appearance and is grown to produce leafy vegetables like Chinese cabbage and Chinese Pak choi. The turnip and also the vegetable types are important as vegetable sources of fodder to feed sheep and cattle (CFIA, 1999).

Because *B. rapa* is such an important crop, a lot of effort is put into research in order to find ways of improving the different traits. For the improvement of a crop, knowledge is required about the genetic variation underlying the traits and molecular markers linked to relevant genes. The knowledge gained can then be used to optimize plant breeding programs. The research needed includes assessing traits on a phenotypic and genotypic level, which is what will be done for this thesis.

1.1.1 Aim of the thesis

For the research in this thesis, first plant morphological traits will be scored on a segregating population of *B. rapa*. Thereafter marker analysis will be performed to detect differences in the DNA sequence between the recombinant inbred lines. To determine the distance between and the relative order of the markers, a genetic linkage map will be made. This linkage map will be anchored with publicly available SSR markers to a *B. rapa* reference map made by Choi *et al.* (2007). QTL analysis can than be performed to identify markers that are associated with important qualitative and quantitative agronomic traits. With the acquired knowledge on traits important in breeding, breeding programs can be made more efficient. A more detailed explanation of the above mentioned terms and methods will be given in the following sections of this thesis.

1.2 Quantitative trait study

1.2.1 Mapping populations

For making a linkage map a segregating population is required. This is a population derived from sexual reproduction that typically segregates for different traits. The parents selected for the mapping population are chosen because they differ for one or more interesting traits (Collard *et al.*, 2005). When parents in a crossing are less related their genetic make-up and consequently their characteristics will also be less similar. Therefore the population will segregate to a greater amount, which facilitates the finding of QTLs. But attention also has to be paid to the offspring because if the F1 turns out to be mostly sterile this can give problems.

There are several different kinds of mapping populations that can be used for QTL analysis. Each population type has its own advantages and disadvantages. A simple way of creating a mapping population is by backcrossing. This is done by crossing the F1 back to one of the parents, and then take the F2 for mapping purposes. This quickly produces a population segregating for different traits. The downside is that this population is still largely heterozygous and will thus segregate with reproduction. The consequence is that the exact same population can never be obtained again and used in the same way again, thus creating a mortal population (Quijada *et al.*, 2007). One way of solving this problem is by using doubled-haploid (DH) populations or inbred populations which are immortal. DH populations can be developed by taking the haploid microspores or anthers and culturing treating them in order to get spontaneous chromosome doubling. If this does not work, colchicine can be used to induce chromosome doubling. However, this treatment does not seem to work well for *B. rapa* (Kole *et al.*, 1997). Also, these populations tend to have a relatively high level of segregation distortion, where the genotype of one parent is more often present in the offspring than the other parent's genotype. This is especially the case in DHs, since lethal genes then

become homozygous, leading to expression and elimination of that line (Xian-Liang *et al.*, 2006).
A mapping population that is under the influence of this phenomenon to a lesser extent and that does have the advantage of being immortal is a population of recombinant inbred lines. A RIL population is produced by inbreeding individual F2 plants. After multiple generations (typically more than seven), this produces a series of (almost) homozygous lines, each having a different combination of chromosomal parts from the original parents (Collard *et al.*, 2005). For this study the F7 generation of a RIL population will be used.

1.2.2 Phenotyping

The plants used for this thesis come from a cross made between genotype L58, which is a rapid-flowering, self-compatible Caixin line and R-o-18, a European doubled haploid spring oil genotype. Crossing this vegetable variety of *B. rapa* with an oil variety produces a population segregating for different plant traits that are typical for these varieties.
The way to improve certain traits is to do research on the underlying genes for these traits. Therefore, first each individual in the recombinant inbred population has to be phenotyped where after the specific traits of each plant are compared to its genotype.

In breeding of *Brassica rapa*, the general aim is to increase yield, improve agronomic characteristics and improve quality. The specific goals and priorities differ for each breeder, depending on the vegetable type and its specific characteristics. Often, market demands are taken into account when a breeding program is designed (Zhao, 2007). Traits desired by farmers, processing companies and shop owners are also considered by the breeder.
The traits that are studied for this thesis have mainly been chosen for the reason of being responsible for important plant characteristics like yield and growth. Some traits have previously been studied but for others QTLs and underlying genes are still unknown. Here follows a list of the studied traits with an explanation of why they are important:

Plant height is one of the agronomic characteristics that is of importance to the farmer because it can influence plant growth and development in many ways. Take oilseed rape for example; breeding for a small plant would mean it would easily be overgrown by competitive weeds, but a very tall plant would invest too much in growth instead of in its seed. Although every cultivar requires a different height, depending on its product, the ability to detect where the genes responsible for plant height are located would accelerate breeding. For this thesis research, plant height will be measured from the base to the end of the inflorescence after flowering.
Within plant height, also **main inflorescence length** is included. The reason for looking at the length of the main inflorescence as a separate trait is that a longer inflorescence might give more siliques (the seed containing seedpods) and thus a higher yield.
Since inflorescence length might influence silique number, the latter trait was studied by Bagheri (2009), but no QTL was found. Therefore, in this thesis research only the **number of siliques in the main inflorescence** will be taken into account. Another silique related trait studied is the **number of seeds per silique**. Together these traits determine yield in oil varieties of *B. rapa*. A high number of siliques with a high number of seeds per silique would give a high yield, which is one of the most important goals in *B. rapa* breeding.

Height at top and bottom of the silique is defined as the height of the silique 7 mm from the top and bottom. A segregation for this trait can be expected since it can be seen by eye that parent R-o-18 has bigger siliques, both at top and bottom, than does parent L58. The use of siliques with a greater height at the bottom or top might by that they carry bigger seeds. This would have a positive influence on yield. Also the **ratio between the height at top and bottom of the silique** might give information about seed size or other seed related characteristics.
An important trait in *B. rapa* is the **seed coat colour**, which can range from yellow to black. This trait is of high significance one because the colour of the seed coat is related to the quality of the seed. Yellow seeds have a seed coat that is thinner than the one from black seeds. This lower percentage of seed coat in the product means that more oil and protein are produced. Another advantage is that the oil is more transparent and that the meal has a lower fiber content. This makes yellow seeds more suitable for feeding livestock (Tang *et al.*, 1997; Zhang *et al.*, 2009).
For oilseed rape, another trait of significance is **seed ripening time**. This trait is defined as the time it took from planting to when most siliques on the main stem were ripe and dry. Since the seed ripening time determines when the plants can be harvested a short ripening time is preferred. The parents chosen for the RIL population in this study are both fast flowering, so seed ripening time should therefore be relatively short. Still the population should segregate for this trait so QTLs and eventually the underlying genes can be detected.
A trait that might be correlated with seed ripening time is the **germination rate**. Some seeds will germinate faster or slower than others and fast germination will contribute to a shorter growth period. Both parents chosen for the RIL population were fast flowering but still a difference in germination rate was seen. To score this trait number of seeds that have germinated 15 hours after sowing will be scored.
Vivipary is the phenomenon where seed germinates while still inside the silique. This causes a decline in quality of oil and meal products of oilseed rape. Resistance to vivipary is therefore an important goal in oilseed breeding programs.
Vegetable types of *B. rapa* like Chinese cabbage and Pakchoi have other traits of agricultural significance. For these varieties the position of the leaves on the stem is very characteristic. The vegetable types derive their typical morphology from the leaves that are positioned very close to each other on the bottom part of the stem. That is why the trait **number of leaves in the first 15 cm** is taken into account, where all leaves on the main stem are counted from the base up to 15 cm.

1.2.3 Genotyping

One way of evaluating if a plant is suitable for further use in breeding, is to select for desired characteristics by looking at the plants' phenotype. This method is relatively simple, but it has some disadvantages. The procedure is rather time consuming and selection can only take place for one trait at a time. Also, the way a trait is expressed can be influenced by the environment, making the selection less effective. Lastly, traits like auxin or fatty acid content are not easily scored. That is where genotyping comes in handy.

Genotyping is done by using molecular markers or genetic markers. Molecular markers are pieces of DNA that represent genetic differences between individual organisms of a species. They sometimes represent the target genes themselves but more can also act as a signal for genes that are close by and that the markers are linked to. Since the genetics (together with the

environment) is an important factor in determining plant traits, and markers are linked to genes, it is possible to see which traits a plant has just by looking at the markers (Collard *et al.*, 2005). Selecting plants with a suitable phenotype for breeding purposes or research with the help of markers is called marker assisted selection.

With marker assisted selection, molecular markers can be used to identify major genes and quantitative trait loci (QTLs) responsible for quantitative plants characteristics. This analysis has some major advantages. Many traits can be scored at once and in a reliable way. Also, since the plant's DNA is already present at the seedling stage, pre-selection can be done to pick out the most desirable genotypes. So marker assisted selection has the advantages of being efficient, effective, reliable and cost-effective compared to the more conventional methodology used in plant breeding (Collard *et al.*, 2005).

Since molecular markers are thought of as handy tools for plant breeding, multiple kinds of molecular markers have been developed and used for plant breeding and genetic studies (Suwabe *et al.*, 2002). The most commonly used markers and their advantages and disadvantages are presented in Collard *et al.* (2005). These are: Restriction fragment length polymorphism (RFLP), amplified fragment length polymorphisms (AFLPs), Cleaved Amplified Polymorphic Sequences (CAPS) and simple sequence repeats (SSRs).
RFLPs are co-dominant. This means that they are able to indicate size differences whereas dominant markers are either present or absent. So co-dominant markers may have many different alleles in contrast to dominant markers, which only have two alleles. Other advantages of the RFLPs are their reliability, robustness and their transferability across populations. Downsides are the fact that they are time-consuming, laborious and expensive. Also large amounts of DNA are required and there is limited polymorphism, meaning that generally two genotypes are not different and thus not polymorphic for many markers.
CAPS polymorphisms are differences in the length of restriction fragments caused by single nucleotide polymorphisms (SNPs), insertions or deletions (INDEL). First the restriction enzymes are used to digest the DNA resulting in DNA fragments of different lengths. Then PCR is performed to amplify the polymorphisms with locus-specific primers. When the PCR products are fractioned by agarose or acrylamide gel electroforesis, the digested PCR products will be visible as clearly distinguishable bands. The advantage of CAPS markers is that they are co-dominant, locus-specific and easy to score and interpret. Also the markers are well exchangeable between other laboratories (http://www.ncbi.nlm.nih.gov/).

The types of markers used for this thesis are SSR, AFLP and SNP markers. SSR markers or microsatellites are short repeating units of DNA (1-6 bp) that are present throughout the genome. The nucleotide sequences next to the repeated parts are used to design primers. SSRs have several advantages, like being abundant in most species and also being polymorphic for the number of repeated units that they have in the piece of DNA investigated. Furthermore, SSRs are co-dominant, robust, reliable and can be analyzed by a simple PCR-based method that allows screening for a large number of individuals at the same time (Suwabe *et al.*, 2002). Downsides are that producing primers takes much time and labour, and that polyacrylamide gel electrophoresis using a sophisticated setup is usually required (Collard *et al.*, 2005).
In this thesis the F7 population will be screened with 14 SSRs with a known mapping position. This way the SSR are not only used to detect polymorphisms but also to anchor each linkage group to a known linkage group from the reference genetic linkage map made by Choi *et al.* (2007).

AFLP or Amplified Fragment Length Polymorphisms is a type of marker that has the great advantage that it can be used for many loci at the same time, which will speed up marker analysis a great amount. Further more, the reaction is locus specific within a species, suitable for high throughput and reproducible. Also no prior genetic knowledge is necessary. Downsides are the complicated methodology and that the markers are mostly dominant.

To detect single-nucleotide polymorphisms the Complexity Reduction of Polymorphic Sequences (CRoPS®) technology will be used as a new way for discovering large-scale polymorphisms in complex genomes. Polymorphisms, which are found in the DNA sequence, are mostly all single-base-pair differences (Quijada *et al.*, 2007). SNPs have the ability to detect these tiny differences in DNA sequence between the two alleles of a gene. This makes these markers particularly useful, because in this way they can reveal differences between individuals of the same or different species (Collard *et al.*, 2005). Since SNPs are able to detect tiny differences in DNA sequence, they are also used to fine-tune existing maps that have already mapped the locations of QTLs (Hayashi *et al.*, 2004).
To speed up the genotyping process SNPs, high throughput genotyping platforms can be used. For genotyping of the RIL population the recently developed Illumina® BeadXpress™ platform will be used in combination with the GoldenGate assay. This enables efficient and high-throughput genotyping.

1.2.4 Genetic linkage map

After marker analysis with the SSR, SNP and AFLP markers will be completed, the data can be put together to make a genetic linkage map that is anchored to a reference map. A linkage map indicates the position and relative genetic distance between markers along the chromosomes. Linkage maps are used to see where chromosomal regions are that contain genes controlling simple traits (controlled by a single gene) and quantitative traits using QTL analysis (Mohan *et al.*, 1997; Collard *et al.*, 2005).
A linkage map is constructed by taking the DNA marker data for each individual of the population. On this data, linkage analysis is conducted using computer programs. The distance between, and the relative order of the markers can be determined by measuring the number of recombinations. The distance on a linkage map can be calculated by taking the number of times recombination has taken place between markers. The relative order of the markers can be calculated by looking at the number of times recombination has taken place. If little recombination has taken place, two markers will be closer on the same chromosome. On the other hand, the chance of recombination during meiosis is higher when markers are further apart on a chromosome (Collard *et al.*, 2005).
For making a linkage map, in this thesis the computer program JoinMap4 will be. This program uses odds ratios for calculating the linkage. These represent the ratio of linkage versus no linkage, and are also called the logarithm of odds (LOD) score. For making linkage maps, a LOD score of 3 or more is often used. This means that linkage between two markers is 1000 times more likely than no linkage.

® *The AFLP®, KeyGene®, CRoPS® technologies are covered by patents and/or patent applications owned by Keygene N.V. AFLP, KeyGene and CRoPS, are (registered) trademarks of Keygene N.V.*

1.2.5 QTL analysis

After marker analysis and the construction of a linkage map, QTL analysis can be performed. The goal is to find a quantitative trait locus by establishing the relation between a quantitative trait locus (marked by a marker) and a phenotyped quantitative trait. QTL analysis is performed by taking the phenotypic and genotypic (marker) data and calculating the probability that they are associated (P>0.05). Because of the large amount of phenotypic and genotypic data, special computer software has been developed that can systematically search the genome for genes and QTLs (Collard *et al.*, 2005).
The computer program MapQTL5 will do QTL analysis by performing interval mapping and multiple QTL mapping (MQM mapping). With QTL analysis, the previously produced genetic linkage map is used to produce a so called "QTL likelihood map". The method works by calculating for each location in the genome, one by one, the possibility that there is a segregating QTL present. Simultaneously it is calculated how large the genetic effect of the QTL is and how much the residual variance is. The likelihood that the calculations are true can be determined by the LOD score. To see where the significance threshold is for the LOD score, the permutation test can be performed. When the LOD score is higher that this threshold value, a segregating QTL is found on this linkage group. The location of the highest LOD score on the linkage group is taken as the QTL position (Van Ooijen, 2006).
After interval mapping, multiple QTL mapping can be performed. This is basically performing interval mapping again but now with a subset of the markers that gave a high LOD score as covariates. This way, less significant QTLs for the same trait will also appear because the residual variation will be reduced. After performing MQM mapping for the first time, the positions of some QTLs can be different from the locations in the phase of cofactor selection because the power has been enhanced. Now the selection of cofactors should be switched a number of times to find the best possible result.

Some plant traits that will be studied for this thesis were studied before, where sometimes one or more QTLs were detected. Other traits were not studied in the same way before and so an attempt was made to look for traits that were quite similar.

Plant height was studied by Lou where plant height for both studies was defined as the "height from ground to the apical point of plant at flowering stage" (Lou *et al.*, 2009). Three QTLs for plant height were found on R02, R03 and R07. Another QTL for plant height was also detected on linkage group A5 by Song (1995), where plant height was measured from the base to the first open flower.
For **main inflorescence length** no QTLs have yet been found in *B. rapa*. In *Arabidopsis* however, 4 QTLs for inflorescence length (El-Lithy *et al.*, 2004) have been detected.
Number of siliques in the main inflorescence and **number of seeds per silique** together greatly influence the yield in oil varieties of *B. rapa*. Consequently, much research has already been done in order to find QTLs for seed yield. However, the only research done on the number of seeds per silique in *B. rapa* was performed by Bagheri (2009). Here, 3 QTLs were detected on linkage groups A6, A9 and A10. In *B. napus* seed number has also been examined and a QTL was found on linkage group A2 (Shi *et al.*, 2009). For number of siliques in the main inflorescence, no study has yet found a QTL, although a QTL for total silique number was found on linkage group A9 (Bagheri, 2009).

For the traits **silique height at top and at the bottom of the silique**, a quite similar study was previously performed by Lou. The silique height was there defined as the "Width at the lengthwise midpoint of each silique" and one QTL was found on R07 (Lou *et al.*, 2009). The **ratio between height at the top and bottom of the silique** was not studied before.
Since **seed coat colour** is related to the seed quality, much research has been performed to find QTLs and underlying genes. One locus controlling seed coat colour was mapped on linkage group A5 by Teunonico and Osborn (1994). Lao mapped another QTL for seed colour in *B. rapa* to the middle of linkage group R09 (Lou *et al.*, 2009). Recently Bagheri found this QTL on linkage group A9 as well as a QTL on an unknown group that later turned out to be linkage group A3 (Bagheri, 2009).
So far, no QTLs were found for **seed ripening time** as it is defined here. However, many studies have tried to find QTLs for flowering time, which is the time it takes the plant from sowing to open the first flower. This trait has a large part in the seed ripening time so the underlying genes might partially be the same. Lou (2007) found eight QTLs for flowering time, on R01, R02, R03, R06, R07, R08, and R10. These QTLs were studied in four different populations and evaluated in different growing seasons. Possibly, these QTLs will coincide with QTLs for seed ripening time.
Seed ripening time might be correlated with germination rate, since some seeds will germinate faster than others. To find out how fast seed germinate the **germination rate** per line was scored 15 hours after sowing. A similar experiment is not known. Even though both parents are already fast flowering they are still expected to segregate for the germination speed.
For the phenomenon of **vivipary** in *B. rapa* Bagheri (2009) has revealed 1 QTL on linkage group "G3" which later mapped to linkage group A9.
In other research, a whole range of different leaf related traits has been examined so far, like the number of leaves at a certain time after planting, or on a certain place of the stem.
The most closely related research performed **on number of leaves in the first 15 cm** on the main stem was the number of leaves in the first 20 cm, were no QTL was found (Bagheri, 2009). No other research has been done yet with the purpose of finding a QTL for the number of leaves in the first 15 cm of the stem after plant growth has stopped. Lou (2009) however did find 12 markers that were associated with leaf number, defined as the number of leaves before flowering. Of these 12 markers, 4 had a known map position and could be found distributed over linkage groups R02, R03, R05 and R07.

2. Material and methods

2.1 Development of the RIL population

A population of recombinant inbred lines was made by a cross of *B. rapa* between the rapid-flowering, self compatible Caixin line L58 (♀) and the self-compatible Indian doubled haploid spring oil line R-o-18 (♂). To create an almost completely homozygous RIL population segregating for a number of different traits, the F2 generation was propagated until the F7 generation using single seed descent. One of the most obvious segregating traits was seed coat colour, which ranged from dark black (L58) to light yellow (R-o-18) (Figure 1).
In April 2007 the plants were grown in separate pots in a greenhouse with controlled temperature and artificial day light extension to 16 hours. After approximately four weeks of growing under optimal growing conditions, the first lines started to flower. To prevent cross-pollination plastic bags were put over the inflorescences. In case seed set was poor, pollination was done by hand. In June 2009, a F7 population of 160 individuals was obtained and considered as recombinant inbred population for genetic studies (Bagheri, 2009).

L58 (♀) R-o-18 (♂)

Figure 1. *Both parents of the RIL population, with L58 carrying black seeds and R-o-18 yellow ones.*

2.2 Phenotyping

A RIL population of 160 individuals was created by crossing genotype L58 and R-o-18 and propagated by repeated selfing. The female parent in this crossing was L58, which is a flowering Chinese cabbage (*B. rapa* var. *pekinensis*), vegetable type and it was provided by Dr. Jian Wu (The Chinese Academy of Agricultural Sciences). The male parent in the crossing was R-o-18 (*B. rapa* var. *trilocularis*), which is a Yellow Sarson, double haploid oil type and it was obtained from Dr. Lars Ostergaard (John Inns Centre). Both lines are fast flowering and self compatible.
Crossing L58 and R-o-18 of led to offspring varying for a range of traits inherited to a certain extend from either one of the parents. Noticeable was the range of colours of the F8 seed coat where parent L58 had black-brown seeds and R-o-18 yellow seeds.

The RIL population was phenotyped for 12 traits, including plant height, main inflorescence length, number of siliques in the main inflorescence, number of seeds per silique, height at top and bottom of the silique, ratio between height at top and bottom of the silique, seed coat colour, seed ripening time, germination rate, vivipary and number of leaves in the first 15 cm. For each trait measurements were conducted for the 2 plants per line that were present and the average was taken. For the parent lines, there were three F7 lines present from which the average was taken. In case of missing data for one or the F7 lines, only the available data was taken into account.

Plant height was measured from the lowest leaf on the stem to the highest point of the main inflorescence. Length of main inflorescence was measured from the lowest seed containing silique on the main stem, to the highest point of the main inflorescence. For the number of siliques in the main inflorescence all seed containing seed pods in the main inflorescence were counted. The number of seeds per silique were determined for 5 siliques of each line and the average was taken. The height at top and bottom of the silique was measured 7 mm from the beginning and end of the silique for 5 siliques of each line (Figure 2 and Figure 3). From the previous trait the ratio of height between bottom and top of the silique could be calculated by taking the height at the top and dividing it by the height at the bottom of the silique. Seed coat colour of the F8 seeds was scored by eye and grouped into 9 classes from yellow (1) to black (9). For seed ripening time the dates of when the plants were sown and when they were ready for harvest were taken to calculate the time it takes for the plants to complete a full cycle. The plants were harvested when the siliques on the main stem are ripe and dry. The germination rate for the F8 seeds was obtained by sowing 30 seeds of every genotyped line and looking at the percentage of germination 15 ours after sowing. The seeds were sterilized by putting them in sodium hypochlorite (2% active chlorite) for 2 minutes. After rinsing 2 times in sterile water distilled water, the now sterile seeds were sown in two rows of 15 seeds on square plates with 50 ml of sterile 0.8% agar with half strength MS medium and vitamins (Murashige and Skoog, 1962). The petri dishes plates were placed vertically in a 25° C growing chamber with a 16 h light and 8 h dark photoperiod and scored for germination 15 h later (Figure 4). Vivipary was scored by eye where the seeds from each line were classified into four groups based on the number of seeds with radicles. The number of seeds already germinated (with radicle) per group; group 1: 0-1%, group 2: 1-30%, group 3: 30-60% and group 4: > 60% of germinated seeds. The number of leaves below 15 cm on the main stem were counted from the lowest leaf up to 15 cm above this point.

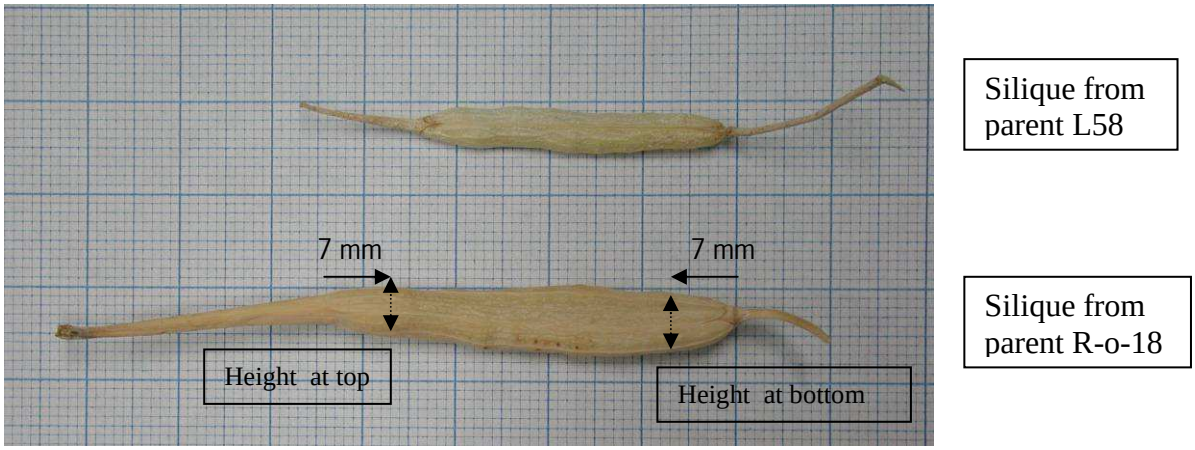


Figure 2. *Pictorial representation of measurement of silique.*

Figure 3. *Picture of siliques from 3 different F7 plants.*

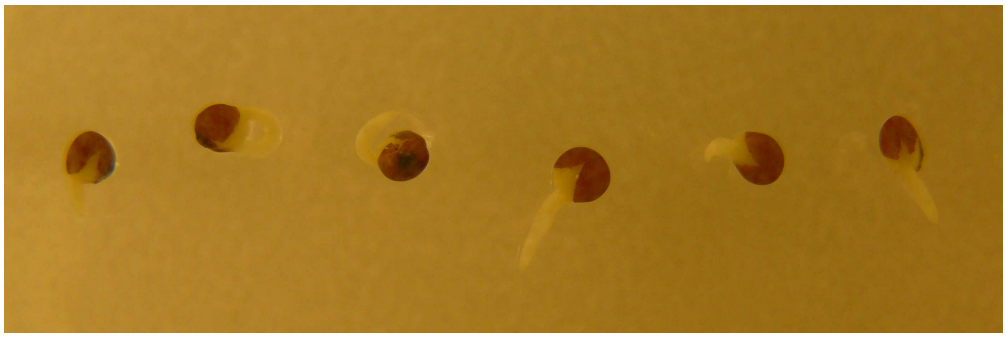


Figure 4. *Germinated seeds 15 h after sowing.*

2.3 Genotyping

2.3.1 DNA extraction

DNA of 160 F7 plants of the RIL population was extracted mostly from frozen leaves and sometimes from flower buds, as described by Zhao *et al.* (2005). In the end 100 µl of DNA was obtained per sample. To check the quality of the DNA, 30 samples were put on a 2% agarose gel. The concentration of the DNA was checked for all samples on a NanoDrop™ machine. Based on the results, DNA was extracted again for 10 samples according to the same procedure.

2.3.1 Marker analysis

In total, 3 types of markers were used for making the linkage map including SNP markers, SSR markers, AFLP markers. Single-nucleotide polymorphisms (SNPs) were detected using the Complexity Reduction of Polymorphic Sequences (CRoPS®) technology. To genotype the RIL population, the Illumina® BeadXpress™ platform that was recently developed was combined with the GoldenGate assay.

For SSR marker analysis a mastermix was made with forward primer that were labeled with IRD-700 or IRD-800 at the 5' end. For distributing the mastermix in portions of 10 µl over the eliza plates a 125 ml distributing pipet was used. Then 3 µl of the extracted DNA was added from each individual of the RIL population. After performing PCR for selective amplification, the product was mixed with 10 µl of formamide-loading buffer and put in -20°C for a minimum of half an hour. Next the samples were denatured at 94°C for 1 minute and cooled on ice to prepare them for running on a 5.5% denaturing polyacrylamide gel with a LI-COR (Lincoln, Neb) 4200 DNA sequencer (Myburg *et al.*, 2001) (Bagheri, 2009). The gel was loaded with 0.5-0.7 µl of the sample, depending on how strong the signal was.
The RIL population was screened with 17 publicly available SSR primer pairs, from which 12 mapped to a linkage group (Choi *et al.*, 2007). For every linkage group at least 1 marker was found (Table 1). Markers BRMS-096-A1, BRMS-042-A3 and BRMS-007-A5 were run on the LI-COR sequencer but did not give a good result.
Also, data from 21 SSR markers obtained from the F2 population were screened across the F7 population by changing the scoring. Of the 13 markers that could be mapped this way, 5 were used for linkage mapping (47BRMS-042-A3, 181/1ENA10,173BRMS-040-A7, 178KS50420uA10 and 186EJU_A10) and were mapped to linkage groups A3, A5, A7 and A10 respectively.

® *The AFLP®, KeyGene®, CRoPS® technologies are covered by patents and/or patent applications owned by Keygene N.V. AFLP, KeyGene and CRoPS, are (registered) trademarks of Keygene N.V.*

Table 1. *List of SSR markers used for linkage mapping, where 12 out of 14 markers could be mapped.*

SSR marker	Mapped	Linkage group
BRMS-031-A1	Yes	A1
BC7-A1	No	A1
Chsssr13-A2	Yes	A2
BRMS-043-A3	Yes	A3
Na10D09-A4	Yes	A4
BRMS-034-A5	Yes	A5
Na12H07-A6	Yes	A6
Ra2A01-A7	Yes	A7
BRMS-018-A7	Yes	A7
Ra2E12-A8	Yes	A8
Ol12F02-A9	Yes	A9
BRMS-051-A9	Yes	A9
Ol10D08-A9	No	A9
BRH80A08flc1-A10	Yes	A10

For AFLP marker analysis the extracted DNA was digested using the combination of restriction enzymes P*st* I/*Mse* I, and ligated to adaptors. Pre-amplification and selective amplification were performed as described by Zhao *et al.* (2005). Selective amplification was done for 2 combinations of P*st* I/*Mse* I primers; P23M48 and P21M47. The P*st* I primer was labeled with IRD-700 at the 5'end for selective amplification (Zhao *et al.*, 2005). The rest of the procedure is the same as mentioned above.
Similarly as described above, 97 AFLP markers that were run on the F2 population tried for mapping in the F7 population by changing the scoring of the loci.

The SSR and AFLP gel images were analyzed by visual estimation. Every distinguishable band between 50 and 500 base pairs was used for data analysis. The SSR bands were scored as 'a' when representing the genome of parent plant L58 or scored, 'b' when representing R-o-18 or 'h' when the locus was heterozygous for both parents. For the marker data from the F2 population, every heterozygote locus 'h' was changed to an unknown locus 'u' before linkage mapping in the F7. In the F7, the AFLP bands were scored for 'a' and 'b'. With the F2 population, scores 'a', 'b' and homozygous for were left unchanged for the parents and the rest of the loci (c, d) converted to 'u' for unknown. This resulted in many loci scored as unknown. All weak or doubtful bands were also scored as 'u' for unknown.

2.4 Construction of genetic linkage map

For performing linkage analysis and constructing a genetic linkage map the program JoinMap 4 was used (Van Ooijen, 2006). For all linkage groups except for A7 the markers were grouped by a LOD score ranging from 2 to 10 to identify the linkage groups. For these linkage groups regression mapping was used as mapping algorithm. Linkage groups A7, A9 and A10 were forced into separate linkage groups by performing linkage analysis for each group separately. For linkage group A7 a LOD score of 1.5 was used to identify the linkage group and maximum likelihood was used as mapping algorithm.

Mapping distances expressed in centi Morgan (cM) were obtained by converting recombination frequencies in cM using the kosambi mapping function.

2.5 QTL analysis

QTL analysis was carried out using the program MapQTL 5 (Van Ooijen, 2006). To discover putative QTLs the interval mapping test was performed. Also the permutation test with 5000 repetitions was done to find the genome wide LOD thresholds per trait (P=0.05). LOD scores of 3.1 (plant height), 3.1 (number of seeds per silique), 6.1 (silique height top), 16.2 (number of leaves in first 15 cm), 9.4 (seed ripening time), 3.1 (seed coat colour) and 4.9 (vivipary) were used as the significance threshold values to indicate the presence of a candidate QTL. Major QTLs are defined as having a LOD score equal to or higher than the threshold value. Minor QTLs After the selection of cofactors, Multiple-QTL model (MQM) mapping was performed to indicate where the QTL was located. Thereafter the 1-LOD and 2-LOD confidence intervals (95%) were indicated by boxes and whiskers, respectively. Major QTLs were indicated by black boxes and minor QTLs by striped green boxes. To visualize the linkage map together with the found QTLs, MapChart5 was used (Voorrips, 2002).

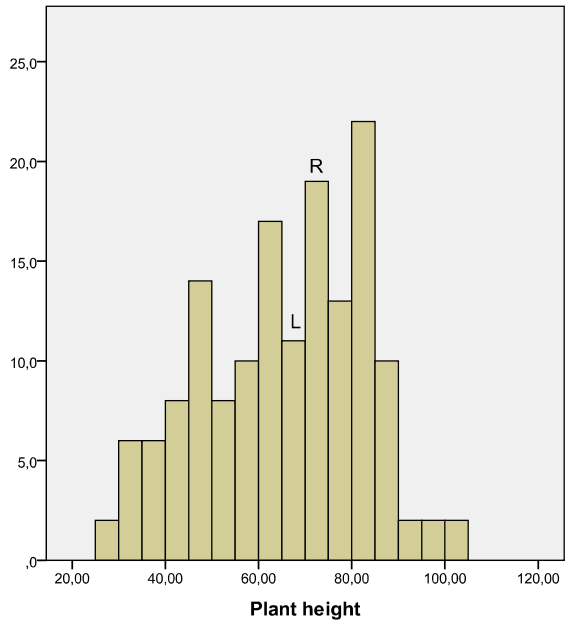
3. Results

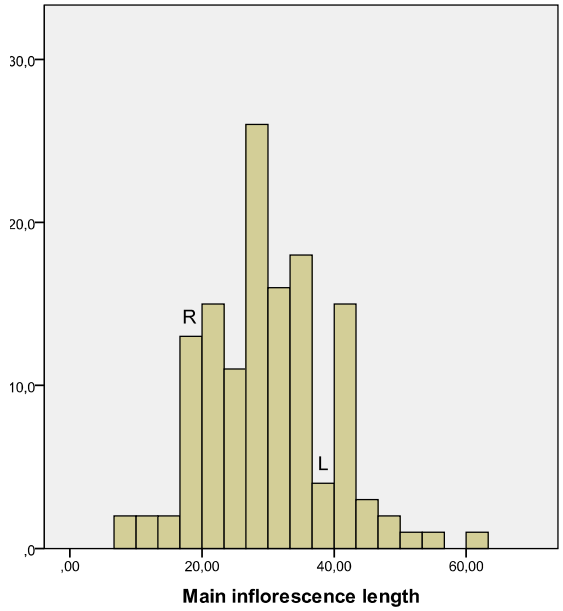
3.1 Phenotyping

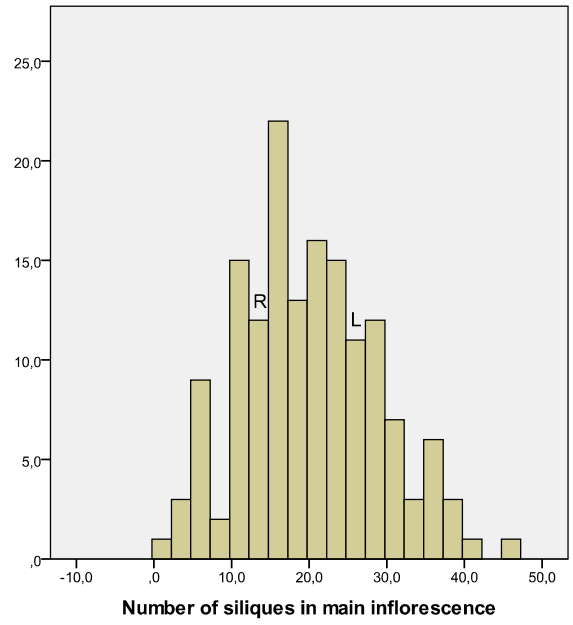
As many as 12 different phenotypic traits were measured in the F7 generation and F8 seeds of the L58 x R-o-18 recombinant inbred population.
Phenotypes that were clearly segregating were seed coat colour and germination rate. Parent L58 has almost black seeds which did not germinate yet 15 hours after sowing. On the contrary, parent R-o-18 has yellow seeds from which 93% had germinated 15 hours after being sown.

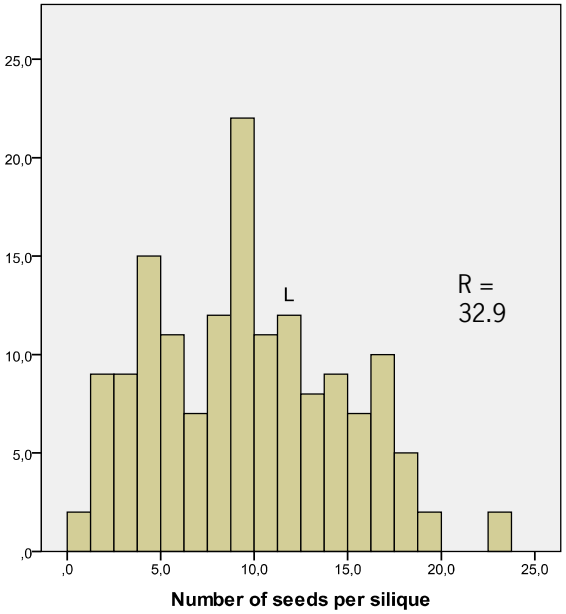
Figure 5 visualizes the frequency distributions for plant height, main inflorescence length, number of siliques in main inflorescence, number of seeds per silique, silique height top, silique height bottom, silique ratio, number of leaves in first 15 cm, seed ripening time, seed coat colour, germination rate and vivipary.
With nearly all traits transgression took place beyond the parental values, except for germination rate. In case of the seed coat colour, there was very little regression beyond the value of parent L58. For the traits number of seeds per silique and silique height at the bottom of the silique, transgression is only beyond the L58 parent. For the rest of the traits, the transgression is to both sides but the transgression beyond L58 is relatively high. Because of this, the data for some traits does not show a normal distribution. Especially the traits number of leaves in the first 15 cm, silique ratio and silique height top show skewedness of data and also have some outliers. Also for vivipary the data is skewed.

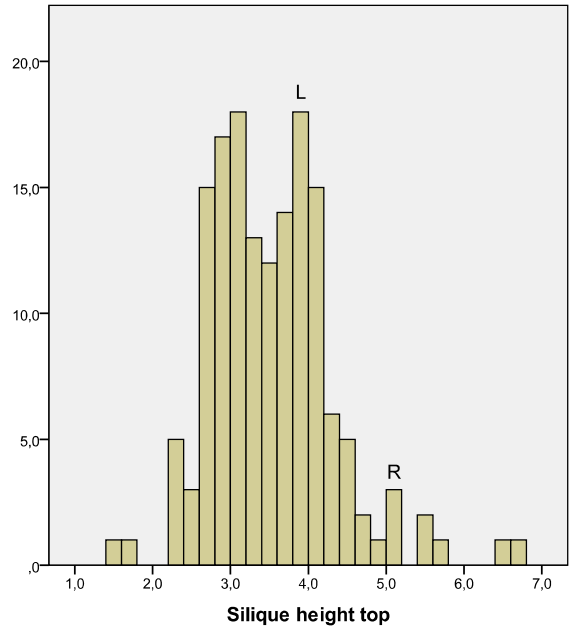
Correlation analysis shows a very high positive correlation between silique height at the top and bottom of the silique (Table 2). Also plant height was highly positively correlated with main inflorescence length and number of siliques in the main inflorescence but there was no significant correlation with vivipary. The number of siliques was highly positively correlated with main inflorescence length, the number of seeds per silique and the number of leaves in the first 15 cm, although the latter traits were not significantly correlated with each other. The number of seeds per silique was however negatively correlated with silique ratio, silique ripening time and vivipary, and positively correlated with seed coat colour and silique height at the bottom of the silique. There was a positive correlation between the number of leaves in the first 15 cm and the seed ripening time, and a negative correlation between the first mentioned trait and vivipary. No correlation at all was found for germination rate. Vivipary showed a high positive correlation between silique height at top as well as at the bottom of the silique, indicating bigger siliques might have more viviparous seeds.

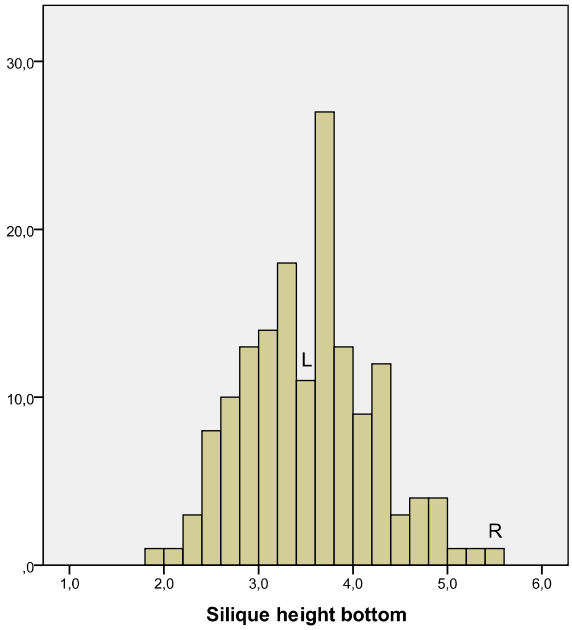

25,0
20,0
15,0
10,0
5,0
,0
R
L
20,00
40,00
60,00
80,00
100,00
120,00
Plant height

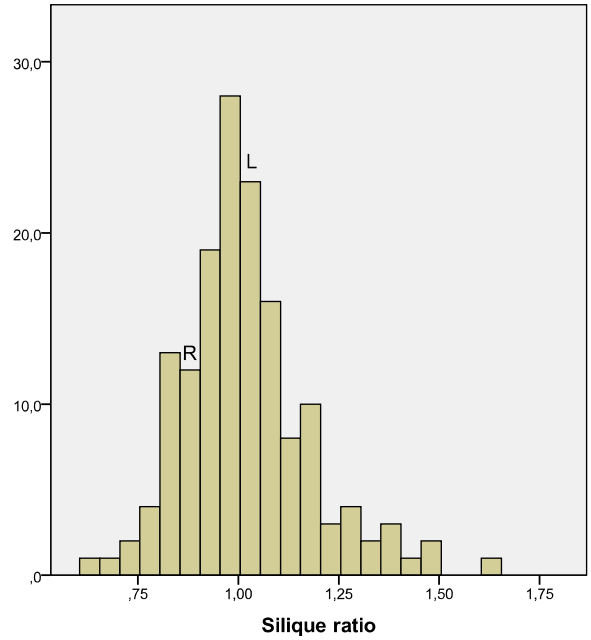

30,0
20,0
10,0
,0
R
L
,00
20,00
40,00
60,00
Main inflorescence length

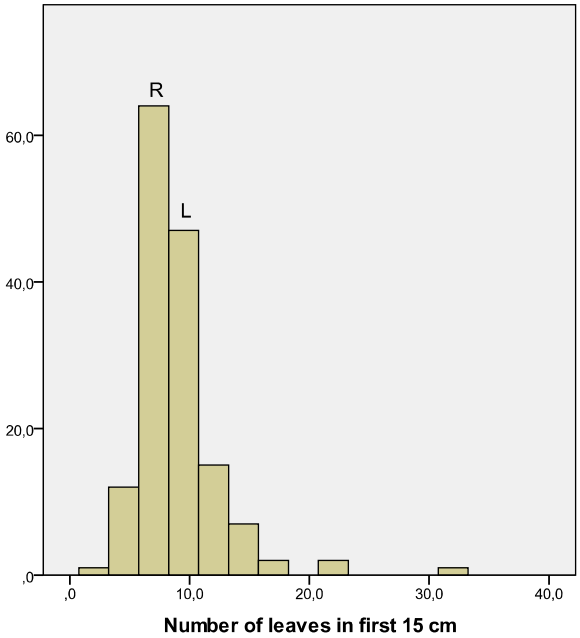

25,0
20,0
15,0
10,0
5,0
,0
R
L
-10,0
,0
10,0
20,0
30,0
40,0
50,0
Number of siliques in main inflorescence


25,0
20,0
15,0
10,0
5,0
,0
L
R =
32.9
,0
5,0
10,0
15,0
20,0
25,0
Number of seeds per silique


20,0
15,0
10,0
5,0
,0
L
R
1,0
2,0
3,0
4,0
5,0
6,0
7,0
Silique height top


30,0
20,0
10,0
,0
L
R
1,0
2,0
3,0
4,0
5,0
6,0
Silique height bottom


30,0
20,0
10,0
,0
L
R
,75
1,00
1,25
1,50
1,75
Silique ratio


60,0
40,0
20,0
,0
R
L
,0
10,0
20,0
30,0
40,0
Number of leaves in first 15 cm

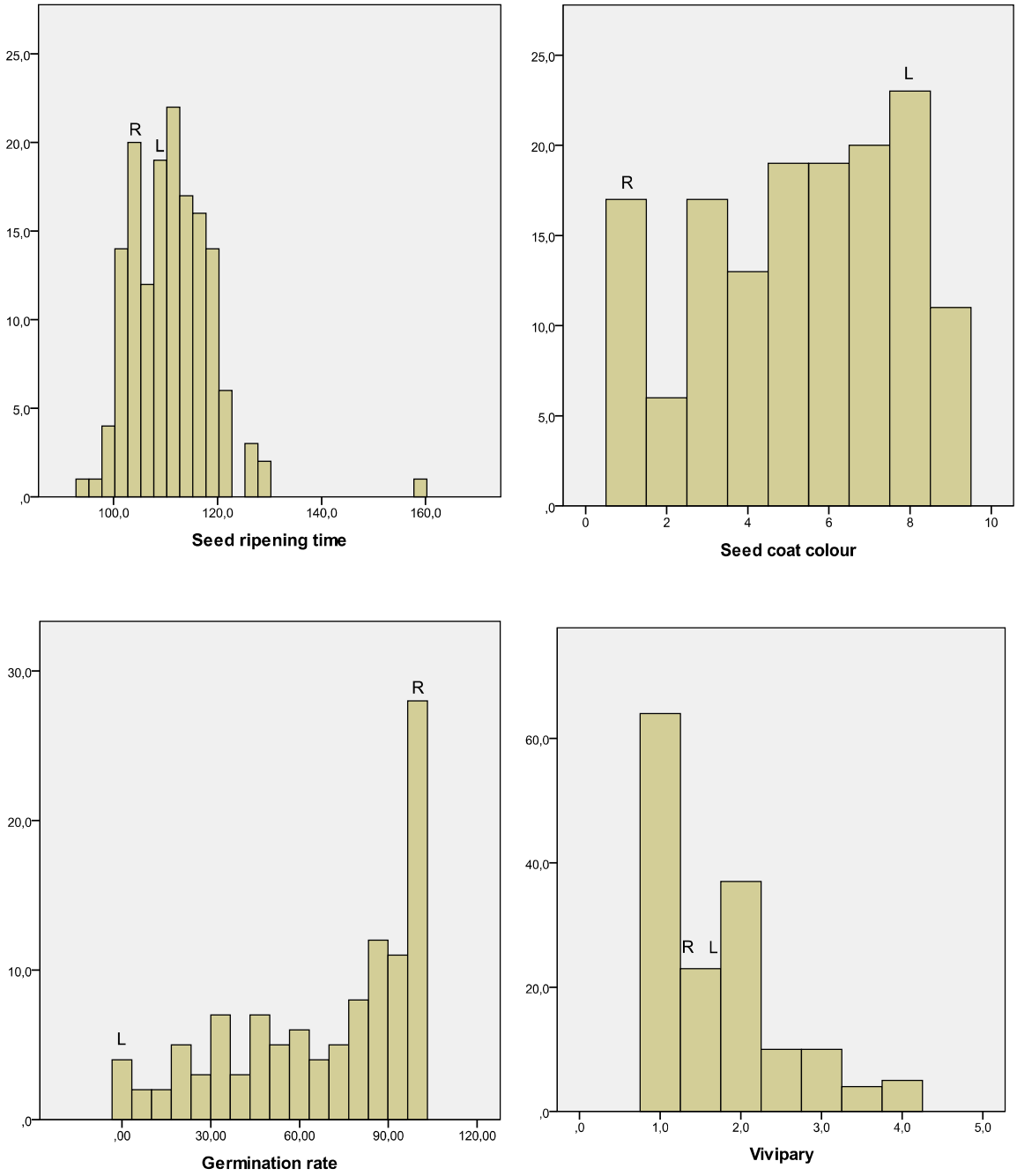


Figure 5. *Frequency distributions of non-normalized data of twelve traits in the L58 × R-o-18 F7 population. On the vertical axis the number of lines per trait value class is shown, on the horizontal axis the different trait value classes are given. Plant height (cm); main inflorescence length (cm); number of siliques in the main inflorescence; number of seeds per silique; height at the top of the silique (mm); height at the bottom of the silique (mm); ratio between height at top and bottom of the silique; number of leaves in the first 15 cm; seed ripening time (days after sowing); seed coat colour; germination rate (% germinated seeds 15 h after sowing); vivipary (amount of viviparous seeds). The parental values are the average of three replicates, indicated as L for L58 and R for R-o-18.*

Table 2. *Pearson correlation table of the measured traits, * Significant at P < 0.05, ** significant at P < 0.01, indicated in light and dark grey respectively. PH: Plant height; InfLe: Main inflorescence length; NrSil: Number of siliques in the main inflorescence; NrSeeds: number of seeds per silique; SR: ratio between height at top and bottom of the silique; NrL15: number of leaves in the first 15 cm; SRT: seed ripening time; SCC: seed coat colour; GR: germination rate; Viv: vivipary; SHT: height at top of the silique; SHB: height at the bottom of the silique.*

	PH	InflLe	NrSil	NrSeed	SR	NrL15	SRT	SCC	GR	Viv	SHT	SHB
PH												
InflLe	0,637**											
NrSil	0,423**	0,573**										
NrSeeds	0,135	0,092	0,398**									
SR	-0,030	-0,045	-0,164*	-0,416**								
NrL15	0,100	0,104	0,212**	0,052	-0,114							
SRT	-0,114	-0,034	-0,105	-0,372**	0,195*	0,180*						
SCC	-0,036	0,026	0,096	0,170*	-0,057	-0,124	-0,105					
GR	0,012	0,054	0,171	0,077	-0,175	-0,028	-0,121	-0,112				
Viv	-0,149	-0,003	0,043	-0,195*	0,038	-0,161*	0,114	-0,072	0,182			
SHT	-0,071	-0,070	-0,056	-0,052	0,369**	-0,313**	0,075	-0,139	-0,011	0,291**		
SHB	-0,089	-0,084	0,037	0,195*	-0,191*	-0,269**	-0,062	-0,063	0,083	0,248**	0,766**	

3.2 Genotyping

3.2.1 Marker analysis

In total 127 markers from all 254 markers used, mapped on the genetic linkage map for the F7 RIL population. Of these markers, the 138 markers run in the F7 population included 90 SNPs, 34 AFLP and 14 SSR markers, from which 122 markers mapped to the genetic linkage map. This constitutes 88 % of the markers used for linkage analysis in the F7 population (Table 3).
A 116 markers run in the F2 of the RIL population were used for linkage analysis in the F7 population. Of these, none of the 97 AFLP markers mapped and 13 out of the 21 SSR markers mapped, of which 5 were used in the linkage map.

Table 3. *Markers run in the F7 population and used in linkage analysis*

Marker type	Total	Mapped	% mapped
SNP	90	85	94.4
AFLP	34	25	73.5
SSR	14	12	85.7

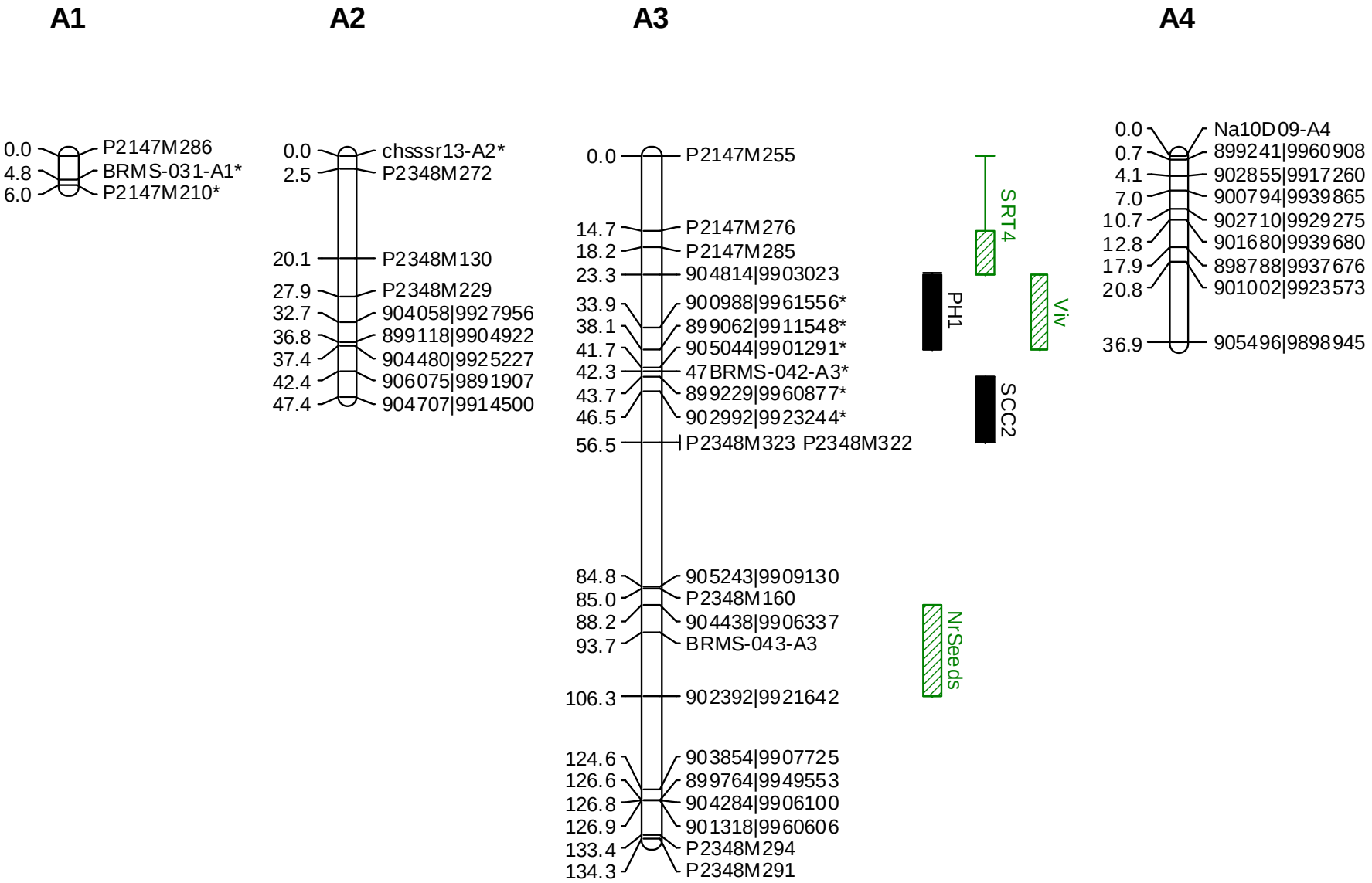
3.2.2 Linkage mapping and QTL analysis

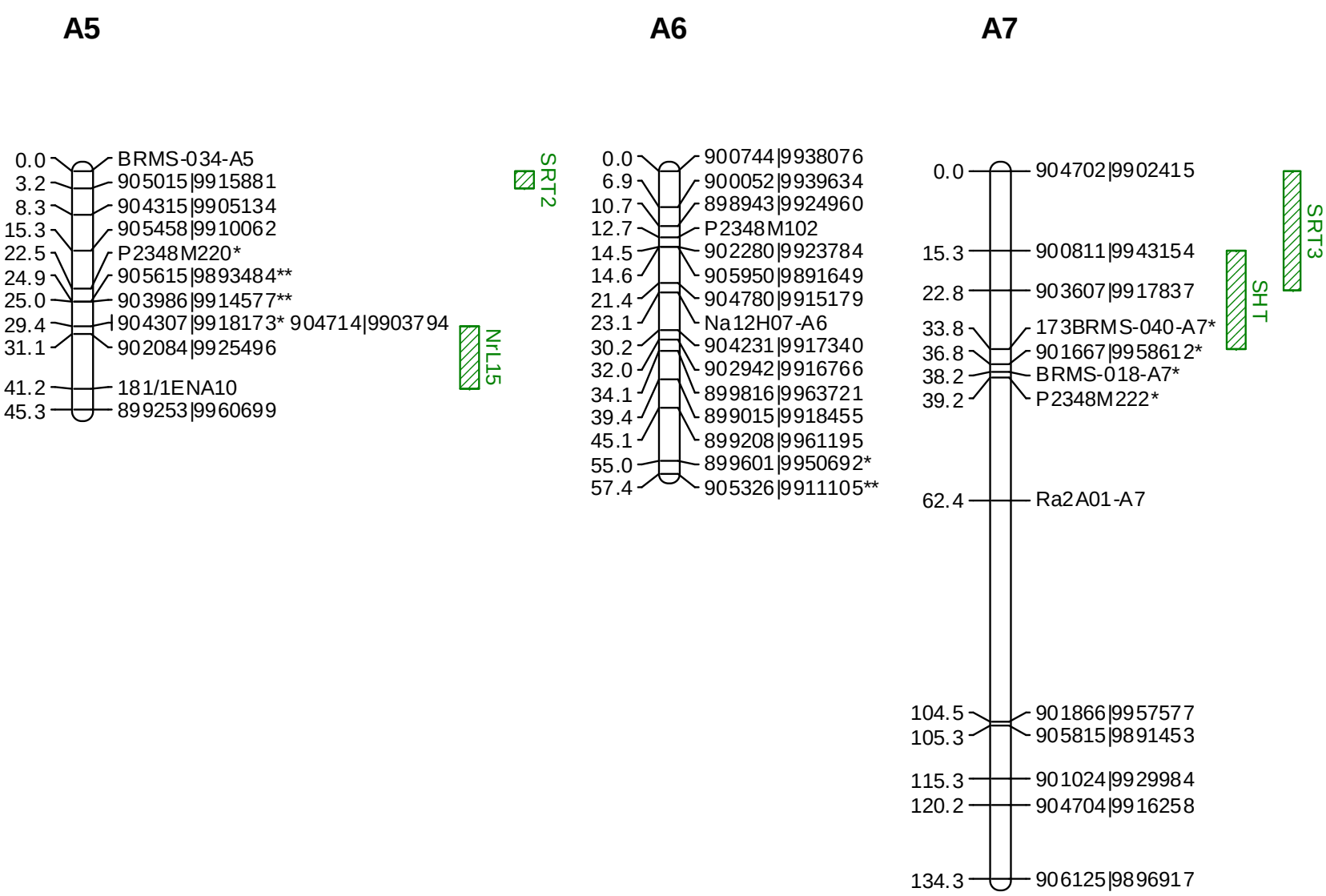
For the L58 x R-o-18 F7 population counting 160 lines, a linkage map was constructed using 85 SNP markers, 25 AFLP markers and 17 publicly available SSRs. The constructed map covered a total distance of 585.5 cM and had an average resolution of 4.6 cM between the markers. All 10 *B. rapa* linkage groups were identified and no extra linkage groups were found (Figure 6).
A number of F7 markers did segregate properly but were not used, including 5 SNP, 9 AFLP and 2 SSR markers. For the 5 SNP markers that were not used in mapping, 1 was excluded because its genetic distance to the linkage group was larger than 50 cM. The other 4 SNPs did not map to one of the linkage groups but mapped alone or in groups of 2. From the 9 AFLP markers that were not used, 2 were excluded because their genetic distance to the linkage group was larger than 50 cM. The other 7 AFLP markers were not used in mapping because they mapped alone, in groups of maximum 2 or with another marker that was not used. There were 2 SSR markers that did not map. Scoring proved to be difficult for 1 marker due to the faint result that was obtained with performing marker analysis. For the other marker genotyping went fine but the marker could still not be mapped.
To come to a genetic linkage map with 10 linkage groups, A7, A9 and A10 had to be forced in JoinMap. For A9 the 26 markers placed on this linkage group with non-forced mapping were taken and 3 extra SNP markers were added. The scoring for these 3 SNP markers was previously incorrect because of a mistake in assigning the parents genome, where ‘a’ (L58) and ‘b’ (R-o-18) were switched. After correcting the scoring, the SNP marker sequences were blasted against the *B. rapa* sequence by Keygene. Some hits were found with BAC contigs for which the linkage group they belong to known and can be found on the website of the Brassica Genome Project (www.brassica-rapa.org). So the 3 SNPs could be identified as

belonging to A9. Thus they were added to the other markers belonging to A9 (found by non-forced mapping) and put in a separate dataset to perform forced regression mapping. This revealed a linkage map highly similar to the previously acquired one.
Since markers from linkage groups A7 and A10 will not map separately from each other with normal linkage mapping, forced mapping was used. Also 5 extra markers were added for both A7 and A10, as they were also previously scored wrong by Keygene. The scoring was changed and the grouped markers were identified as belonging to either A7 or A10, as described above.
From all markers used for linkage mapping, 20.3% showed distortion (≥2:1) based on a $\chi 2$ test for goodness-of-fit to the expected 1:1 mendelian segregation ratio. 4.9% of them are highly distorted (≥3:1) ($P<0.001$). The genome of parent R-o-18 was preferred for 66% and 100% of the loci respectively. The loci for all 160 lines within the population are 1.3 % heterozygous, which is not much different from the expected 1.6% heterozygosity.
For seven different traits a total of 12 QTLs were detected, of which 5 major and 7 minor QTLs (Table 4). The highest QTL was found for seed coat colour and mapped to linkage group A9 where it co-located with a minor QTL for seed ripening time. No significant correlation between these traits was detected in correlation analysis (Table 2). The major QTL found for plant height mapped to linkage group A3 and co-located with a miner QTL for vivipary but also here there a significant correlation was not found. The second major QTLs found for seed coat colour and plant height and mapped to linkage group A3 and A10 respectively. The last major QTL was for number of seeds per silique and mapped to A3.
Quite noticeable were the high LOD threshold values calculated for the traits number of leaves in first 15 cm, seed ripening time, silique height at the top of the silique and vivipary. This resulted in one or multiple minor QTLs for the above mentioned traits.

Table 4. *Major and minor QTLs detected for seven traits as listed below. Major QTLs are reported in bold. LOD threshold value was determined with the permutation test (5000 repetitions and P<0.05) in MapQTL5. % Explained variance is the percentage of the phenotypic variance that can be explained by that particular QTL. The allelic effect per QTL is indicated (Effect) and was calculated with μA-μB (μ = mean) in which A and B are carrying L58 and R-o-18 genotypes at the positions of the QTLs, respectively. The effects are given in centimetre (plant height), millimetre (seed height top) days (seed ripening time), amount of viviparous seeds (vivipary) or without unit (seed coat colour, number of seeds, number of leaves in first 15 cm).*

Trait	Major and minor QTLs	Linkage group	LOD	LOD threshold value	Position of peak LOD cM	% Explained variance	Effect
Seed coat colour	**SCC1**	A9	33.3	3.1	75.8	56.2	+3.77
	SCC2	A3	11.8	3.1	46.5	113.6	+1.98
Plant height	**PH1**	A3	4.9	3.1	33.9	12.4	-12.94
	PH2	A10	3.1	3.1	41.4	7.4	-9.43
Number of seeds per silique	**NrSeeds**	A3	3.1	3.1	93.7	8.8	-3.09
Silique height top	SHT	A7	2.5	6.1	22.8	7.0	-0.44
Number of leaves in first 15 cm	NrL15	A5	3.4	16.2	14.1	9.6	-2.33
Seed ripening time	SRT1	A9	6.3	9.4	77.8	14.2	-6.00
	SRT2	A5	4.4	9.4	45.3	9.8	+5.32
	SRT3	A7	2.8	9.4	15.3	6.1	+4.19
	SRT4	A3	2.7	9.4	18.2	5.7	+3.89
Vivipary	Viv	A3	2.6	4.9	33.88	7.3	+0.46


A1
A2
A3
A4
0.0 P2147M286
4.8 BRMS-031-A1*
6.0 P2147M210*
0.0 chsssr13-A2*
2.5 P2348M272
20.1 P2348M130
27.9 P2348M229
32.7 904058|9927956
36.8 899118|9904922
37.4 904480|9925227
42.4 906075|9891907
47.4 904707|9914500
0.0 P2147M255
14.7 P2147M276
18.2 P2147M285
23.3 904814|9903023
33.9 900988|9961556*
38.1 899062|9911548*
41.7 905044|9901291*
42.3 47BRMS-042-A3*
43.7 899229|9960877*
46.5 902992|9923244*
56.5 P2348M323 P2348M322
84.8 905243|9909130
85.0 P2348M160
88.2 904438|9906337
93.7 BRMS-043-A3
106.3 902392|9921642
124.6 903854|9907725
126.6 899764|9949553
126.8 904284|9906100
126.9 901318|9960606
133.4 P2348M294
134.3 P2348M291
SRT4
PH1
Viv
SCC2
NrSeeds
0.0 Na10D09-A4
0.7 899241|9960908
4.1 902855|9917260
7.0 900794|9939865
10.7 902710|9929275
12.8 901680|9939680
17.9 898788|9937676
20.8 901002|9923573
36.9 905496|9898945


A5
A6
A7
0.0 BRMS-034-A5
3.2 905015|9915881
8.3 904315|9905134
15.3 905458|9910062
22.5 P2348M220*
24.9 905615|9893484**
25.0 903986|9914577**
29.4 904307|9918173* 904714|9903794
31.1 902084|9925496
41.2 181/1ENA10
45.3 899253|9960699
SRT2
NrL15
0.0 900744|9938076
6.9 900052|9939634
10.7 898943|9924960
12.7 P2348M102
14.5 902280|9923784
14.6 905950|9891649
21.4 904780|9915179
23.1 Na12H07-A6
30.2 904231|9917340
32.0 902942|9916766
34.1 899816|9963721
39.4 899015|9918455
45.1 899208|9961195
55.0 899601|9950692*
57.4 905326|9911105**
0.0 904702|9902415
15.3 900811|9943154
22.8 903607|9917837
33.8 173BRMS-040-A7*
36.8 901667|9958612*
38.2 BRMS-018-A7*
39.2 P2348M222*
62.4 Ra2A01-A7
104.5 901866|9957577
105.3 905815|9891453
115.3 901024|9929984
120.2 904704|9916258
134.3 906125|9896917
SRT3
SHT

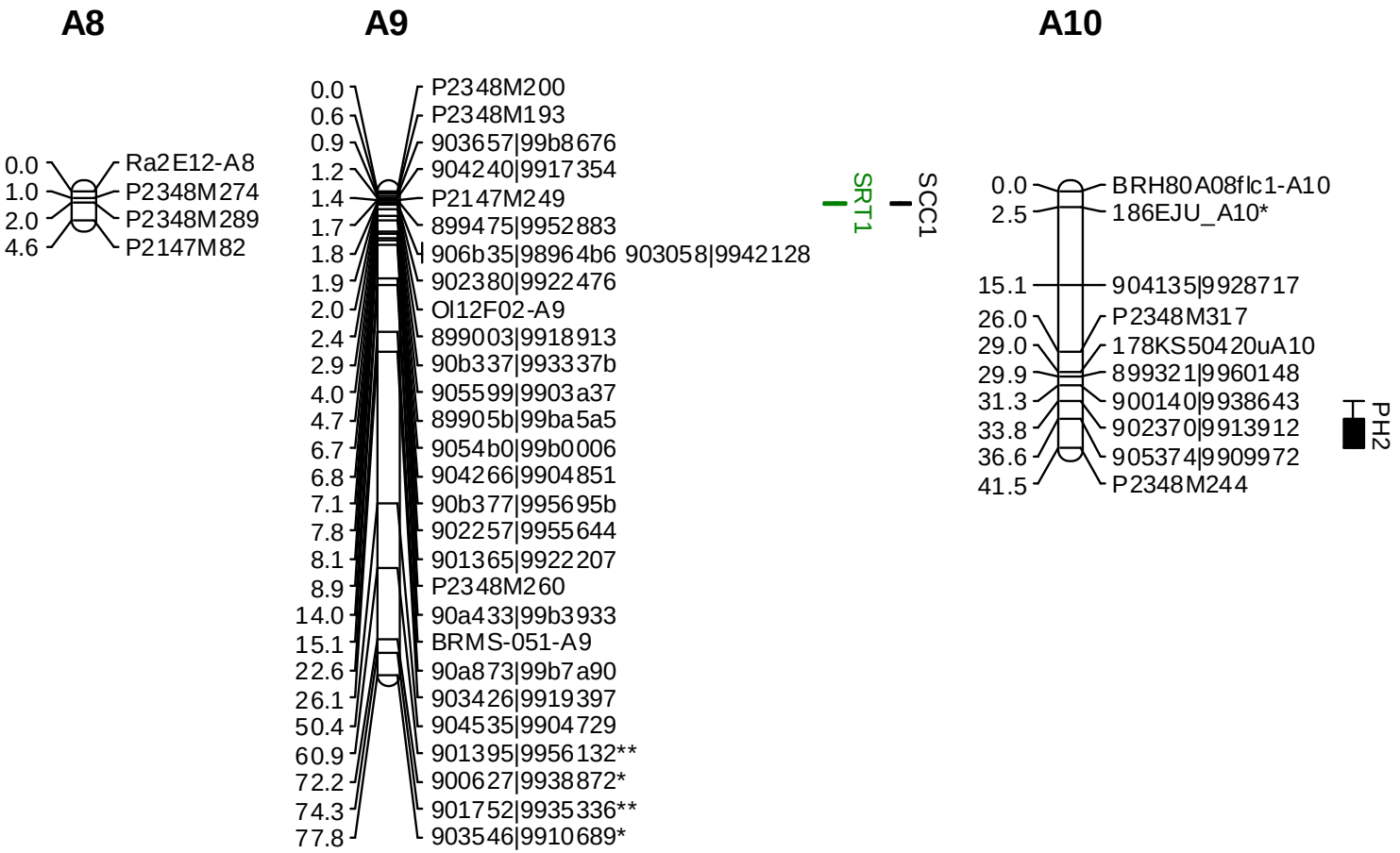


Figure 6. *Genetic linkage map of B. rapa based on the F7 (L58 × R-o-18) population, indicating how the 85 SNP, 25 AFLP and 17 SSR markers are distributed over the 10 linkage groups. All linkage groups correspond to the 10 chromosomes of the B. rapa reference linkage map (Choi et al., 2007). Mapped QTLs for the examined traits are represented by black boxes representing the 1-LOD and whiskers representing the 2-LOD confidence interval (95%) for the found QTLs. Major and minor QTLs are indicated in black and green respectively. SRT = seed ripening time; PH = plant height; Viv = vivipary; SCC = seed coat colour, NrSeeds = number of seeds per silique; NrL15 = number of leaves in the first 15 cm and SHT = silique height at the top of the silique. Skewed markers are indicated with an * (≥ 2:1) and ** (≥ 3:1).*

4. Discussion and conclusion

It was possible to identify all 10 linkage groups without unknown linkage groups being present. To come to this result, linkage groups A7, A9 and A10 were forced in JoinMap for different reasons. For A9, the 3 extra SNP markers that were identified as belonging to this linkage group could be forced. The fact that regression mapping could still be used, and that the forced map looks highly similar to the non-forced one, indicates the correctness of the A9 linkage map.
In case of A7 and A10, there was another reason for forced mapping. Since *B. rapa* has a common ancestor with its relative *A. thaliana*, parts of their genome are highly similar. Mun *et al.* (2009) recently revealed that *B. rapa* linkage groups A7 and A10 have a highly similar region in their genome. The resulting high degree of sequence identity could explain the tight linkage between these linkage groups. The resulting maps could be obtained with maximum likelihood and regression mapping for A7 and A10 respectively, indicating that still a minor problem exists. Adding more markers specific for these linkage groups should help with solving this problem.
In comparison to the reference map made by Choi *et al.* (2007) most linkage groups are significantly shorter. Improvements could still be made for the total linkage map by adding more markers, in order to find the true length for all linkage groups. Especially linkage groups A1 and A8 are very short and only contain 3 and 4 markers in total, respectively. Adding more genetic markers to the linkage map will result in linkage groups with a length closer to the one from the reference linkage map. A number of markers used could not be placed on the linkage map (see below).
In contrary to A1 and A8, A7 is longer than its reference linkage group. The reason might be the gap of 42.1 cM, making it difficult to estimate the true genetic distance. Adding more markers specific for the A7 linkage group will help in giving the real map distance.
The map distance covered by the final genetic linkage map is 585.5 cM, which is half the length of the reference linkage map made by Choi *et al.* (2007). This indicates that still a significant part of the genome is left unmapped, and logically more markers should be added to improve the map. This would also result in a denser map, with a resolution higher than the current resolution of 4.6 cM between markers. Overall, the markers are distributed quite equally over the genome with not too many large gaps. Only linkage groups A3 and A7 show a clustering of their markers, leading to relatively large gaps. To get a more even distribution of markers over the genome, markers specific for the non-covered mapping regions should be added.
From the SNP, AFLP and SSR markers that were used in the F7, an average of 88% mapped to the linkage map. There are a number of different reasons other markers did not map. A first could be that the distance between markers is too large so they will not map together, allowing one of the markers to be mapped all by itself. This was the case for example with the SSR marker for linkage group A8. When a map was made with only SSR and SNP markers, Ra2E12-A8 mapped alone. Though when the AFLP markers were later on added, linkage group A8 emerged. Another important reason some markers did not map could be incorrect scoring of the loci. This was prevented as much as possible by only taking the most reliable scored markers into account, but difficult scoring can partially explain that some markers could not be mapped. The AFLP markers had the highest percentage of non-mapped markers but they were also the most difficult to score due to smiling bands and faint gel pictures.
In this study, markers run in the F2 could also be used for linkage mapping of the F7. Because

the F2 lines are heterozygous for half of their loci and AFLPs are scored dominantly, replaced scores with 'u' (unknown) in case of heterozygous loci led to many unknown loci. There proved to be too many unknown loci and the markers could not be mapped. In case of the F2 SSR markers 62% mapped in the F7. As I ran some of the F2 markers separately in the F7, these were mapped closely to them, proving their correctness. 5 of the F2 SSR markers were finally added to linkage map.
After 7 generations of selfing all lines, the average heterozygosity of their loci is 1.3%, which is close to the 1.6% heterozygosity expected. Markers used for linkage mapping showed a skewedness of 20.3% (≥ 2:1) and a severe skewedness of 4.9 % (≥ 3:1). Here the genome of parent R-o-18 was represented in 66% and 100% of all cases respectively. Skewed markers map to all linkage groups but A4 and A8, and are mostly clustered and flanked by the less distorted markers. The phenomenon of segregation distortion is quite common in mapping populations, especially with outcrossing species such as *B. rapa*. There can be different causes, such as the presence of recessive lethal genes that become homozygous due to inbreeding and are expressed. Another reason could be artificial selection for one of the parents during the growing process. Errors could also have been made in the process of scoring the markers, or somewhere else in genetic analysis, resulting in a segregation ratio deviant from its expected value but this is not expected to be true for many cases (Xian-Lian *et al.*, 2006).
The skewed markers used are mostly in favour of R-o-18 alleles. This could be caused by unintentional selection for both the cultivated phenotype and viability during the single-seed-descent process. According to Xian-Liang *et al.* (2006), viability selection is an important cause of segregation distortion. It is especially more frequent in outcrossing crops such as *B. rapa*, where inbreeding will often happen. Not only does R-o-18 seem to be more viable, it also has significantly bigger seeds than L58. So during development of the recombinant inbred lines, the recombinants with the genome fraction from R-o-18 for seed size and viability may have had bigger seeds and higher viability than recombinants with the L58 genome fraction for these traits. So there could have been unintended selection for both these traits, resulting in skewedness towards the R-o-18 genotype.

Phenotyping was performed for 12 segregating traits. For most traits there is quite a strong transgression of the phenotypes beyond both the parental values, especially for parent L58. This suggests that many traits are influenced by multiple loci that have contrasting allele effects. Therefore, this population has the possibility to be very useful for studying genetic variation in *B. rapa* to use in breeding.
In total, 12 QTLs including 5 major and 7 minor QTLs were detected for 7 different traits, some of which have not been reported in *B. rapa* before. Minor QTLs also have a high chance of indicating the true positions of the genes underlying the traits (Price, 2006). Because these minor QTLs have a fair chance of being true, they were still reported in this thesis report.
One of the most important traits in oil types of *B. rapa* and the species *B. napus*, is seed coat colour, which can range from yellow to black. Yellow seeds are favoured since they produce more oil and protein, that both have a better quality than the black seeds (Tang *et al.*, 1997; Zhang *et al.*, 2009). One major QTL was found on A9 explaining 56.2% of the variation in seed coat colour. This looks to the same QTL as formerly mapped to A9 by Lou *et al.* and Bagheri (Lou *et al.*, 2009; Bagheri, 2009). This QTL even mapped to the same SSR marker in 2 of the cases (Ol12F02). Aditionally, a relatively new major QTL for seed coat colour was discovered which mapped to linkage group A3. This QTL had only been found before on the same marker in the F2 of the same population, but was previously mapped on an unidentified

group (Bagheri, 2009). Together, these QTLs have the possibility to be of great use in the breeding of *Brassica* oil types.
For plant height two major QTLs were found, among which a new QTL located on linkage group A10. The other plant height QTL mapped to the top of A3 and was probably the same one as mapped before by Lou *et al.* and Bagheri (Lou *et al.*, 2009; Bagheri, 2009). A minor QTL for vivipary mapped to the same locus on A3, but no significant correlation between plant height and vivipary was detected. So possibly there are two separate genes influencing plant height and vivipary that are located close to one another on the same chromosome. But measuring the amount of vivipary, which was crudely estimated by eye, could be done more accurately. Also taking this trait into account was only decided later on during the research, and for some lines the viviparous seeds had already been discarded because of their low germination rate. Taking both these factors into account will greatly improve the chance of finding a QTL in following studies.
The last major QTL found is for number of seeds per silique and was located on A3. No QTL was not found on the same linkage group before, so this should be a new QTL. Since the number of seeds per silique is one of the factors determining yield in oil types of *B. rapa*, finding the genes responsible for this trait will be very useful in breeding.
A minor QTL for silique height at the top of the silique was found on linkage group A7. A QTL found for exactly the same trait has not been reported before, but a QTL for seed pod width was mapped to approximately the same region in the genome by Lou *et al.* (2009). Most likely this is the same QTL, as measuring the trait was also done in a quite similar way. Since these two QTLs coincide and no QTL was found for seed height at the bottom of the silique, this indicates that the height at the top part plays a larger role in determining the whole seed pod width. This is also indicated by the much higher correlation that the traits silique height top and silique height bottom have with silique ratio, 0.369 and -0.191 respectively.
For both the number of leaves in the first 15 cm and seed ripening time very high threshold LOD values were determined by the permutation test, 16.2 and 9.4 respectively. The most probable reason for this seems to be their non-normal distribution of trait values. The histograms for these traits show a binomial distribution (Figure 5). To get a normal data distribution and thus a normal LOD threshold value, the differences in the trait values should be decreased with a logarithmic function.
For the number of leaves in the first 15 cm, a minor QTL was detected and mapped to linkage group A5. This QTL co-locates with the one found for total leaf number by Lou *et al.* (2009) which indicates that even though the LOD threshold value is high, the minor QTL might still be true.
4 minor QTLs were found for seed ripening time, on linkage groups A9, A5, A7 and A3. This trait was not studied before, in contrary to the flowering time or days to flowering, both of which make up most of the time needed for seed ripening. Both QTLs detected on A3 and A7 co-locate with the QTLs found for flowering time before by Lou *et al.* (2009) are likely to be controlled by the same locus.

It can be concluded that a very decent mapping population of recombinant inbred lines was developed that segregates for many different traits. A 585.5 cM long genetic linkage map including 127 markers was made, identifying all 10 *B. rapa* linkage groups. Distortion and heterozygosity were both relatively low. The finding of 5 major and 7 minor QTLs proves the usefulness of the *B. rapa* RIL population and corresponding linkage map for future studies on the genetics of breeding related traits.

References:

Front page:
http://hi.xihalife.com/wallpapers/images/399_640.jpg

Preface:
www.agri.tohoku.ac.jp/pbreed/Brassicae.htm

http://www.brassica.info/info/crop-enduses/nutritional-benefits.php
(May 26, 2009)

http://www.ncbi.nlm.nih.gov/projects/genome/probe/doc/TechCAPS.shtml
(September 3, 2009)

http://www.brassica-rapa.org/BRGP/chromosomeSequence.jsp
(October 15, 2009)

Bagheri, H., 2009, Genetic analysis of breeding-related traits in *Brassica rapa*, PhD thesis, Wageningen University.

Canadian Food Inspection Agency (CFIA), 1999, The Biology of *Brassica rapa* L., Plant Health and Production Division, Canadian Food Inspection, Government of Canada.
http://www.marz-kreations.com/WildPlants/CRUC/Docs/BRSRA/BrassicaRapa.pdf

Choi, S. R., G. R. Teakle, Plaha, P., Kim, J.H., Allender, C.J., Beynon, E., Piao, Z.Y., Soengas, P., Han, T.H., King, G.J., Barker, G.C., Hand, P., Lydiate, D.J., Batley, J., Edwards, D., Koo, D.H., Bang, J.W., Park, B.-S. and Lim, Y.P., 2007, The reference genetic linkage map for the multinational Brassica rapa genome sequencing project, *Theoretical and Applied Genetics* 115(6) 777-92.

Collard, B.C.Y., Jahufer, M.Z.Z., Brouwer, J.B. and Pang, E.C.K., 2005, An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts, *Euphytica* 142: 169-196.

Gomez Campo C., 1999, Development in plant genetics and breeding, vol 4., *Biology of Brassica coenospecies*, Elsevier, Amsterdam Lausanne New York Oxford Shannon Singapore Tokyo.

Hayashi, K., Hashimoto, N., Daigen, M. and Ashikawa, I., 2004, Development of PCR-based SNP markers for rice blast resistance genes at the *Piz* locus, *Theoretical and Applied Genetics* 108: 1212–1220.

Kole, C., Kole, P., Vogelzang, R. and Osborn, T.C., 1997, Genetic linkage map of a *Brassica rapa* recombinant inbred population, *The Journal of Heredity* 88: 553–557.

Lou, P., Zhao, J., Kim, J.S., Shen, S., Del Carpio, D.P., Song, X., Jin, M., Vreugdenhil, D., Wang, X., Koornneef, M. and Bonnema, G., 2007, Quantitative trait loci for flowering time and morphological traits in multiple populations of Brassica rapa, *Journal of Experimental Botany* 58: 4005-4016.

Mohan, M., S. Nair, A. Bhagwat, T.G. Krishna, M.Yano, C.R. Bhatia and T. Sasaki, 1997, Genome mapping, molecular markers and marker-assisted selection in crop plants, *Molecular Breeding* 3: 87–103.

Mun, J.H., Kwon, S.J., Yang, T.J., Seol, Y.J., Jin, M., Kim, J.A., Lim, M.H., Kim, J.S., Baek, S., Choi, B.S., Yu, H.J., Kim, D.S., Kim, N., Lim, K.B., Lee, S.I., Hahn, J.H., Lim, Y.P., Bancroft, I. and Park, B.S., 2009, Genome-wide comparative analysis of the Brassica rapa gene space reveals genome shrinkage and differential loss of duplicated genes after whole genome triplication, *Genome biology* 10: R111: 1-18.

Murashige, T. and Skoog, F., 1962, A revised medium for rapid growth and bioassays with tobacco tissue cultures, *Physiologia Plantarum* 15 (3): 472-497.

Myburg, A.A., Remington, D.L., O'Malley, D.M., Sederoff, R.R. and Whetten, R.W., 2001, Protocol for high-throughput AFLP analysis using LI-COR IR2 automated sequencers, *Biotechniques* 30: 348-352, 354, 356-357.

Price, A.H., 2006, Believe it or not, QTLs are accurate!, *Trends in Plant Science* 11: 213-216.

Quijada, P., Xiaowu Wang, J. C., Hirai, M. and Kole, C., 2007, Genome Mapping and Molecular Breeding in Plants, *Springer Berlin Heidelberg* 2: 211-263.

Song, K., Slocum, M.K. and Osborn, T.C., 1995, Molecular marker analysis of genes controlling morphological variation in Brassica rapa (syn. campestris), *Theoretical and Applied Genetics* 90: 1-10.

Suwabe, K., Iketani, H., Nunome, T., Kage, T., and Hirai, M., 2002, Isolation and characterization of microsatellites in *Brassica rapa L.*, *Theoretical and Applied Genetics* 104: 1092-1098.

Teunonico, R.A., and Osborn, T.C., 1994, Mapping of RFLP and quantitative trait loci in Brassica rapa and comparison to the linkage maps of *B. napus*, B. oleracea and *Arabidopsis thaliana*, *Theoretical and Applied Genetics* 89: 885-894.

Van Ooijen, J.W., 2006, JoinMap 4, Software for the calculation of genetic linkage maps in experimental populations. Kyazma B.V., Wageningen, Netherlands.

Van Ooijen, J. W., 2006, MapQTL 5, Software for the mapping of quantitative trait loci in experimental populations. Kyazma B.V., Wageningen, Netherlands.

Van Poppel, G., Verhoeven, D.T., Verhagen, H. and Goldbohm, R.A., 1999, *Brassica* vegetables and cancer prevention. Epidemiology and mechanisms, *Advances in Experimental Medicine and Biology* 472:159-68.

Voorrips, R.E., 2002, MapChart: Software for the graphical presentation of linkage maps and QTLs, *The Journal of Heredity* 93 (1): 77-78.

Xian-Liang, S., Xue-Zhen, S. and Tian-Zhen, Z., 2006, Segregation distortion and its effect on genetic mapping in plants, Chinese Journal of Agricultural Biotechnology 3: 163-169.

Zang, J., Lu, Y., Yuan, Y., Zhang, X., Geng, J., Chen, Y., Cloutier, S., McVetty, P.B.E., and Li, G., 2009, Map-based cloning and characterization of a gene controlling hairiness and seed coat color traits in Brassica rapa, *Plant Molecular Biology* 69: 553-563.

Zhao, J., Wang, X., Deng, B., Lou, P., Wu, J., Sun, R., Xu, Z., Vromans, J., Koornneef, M., and Bonnema, G., 2005, Genetic relationships within Brassica rapa as inferred from AFLP fingerprints, *Theoretical and Applied Genetics* 110: 1301-1314

Zhao, J., 2007, The genetics of phytate content and morphological traits in Brassica rapa, PhD Thesis, Wageningen University.