WAGENINGEN UNIVERSITY

Construction of a Linkage map of a *Brassica rapa* RIL population from a cross of a vegetable turnip and a Wutacai for marker turnip tuber trait association

Master Thesis
(PBR-80436)

Roman Gäbelein Reg. Nr. 890208 291 050

15.04.2014

Wageningen University – Department of Plant Science
Brassica research Group
Supervision: Dr. Guusje Bonnema
Dr. Ningwen Zhang

Acknowledgement

I want to thank Prof. Guusje Bonnema for the opportunity to do my thesis in the Brassica research group of Wageningen University, as well as for her guidance during the research group meetings and personal counselling. Her advice on structure and content and reviewing helped me to improve my final report a great deal.
Secondly I want to thank Dr. Ningwen Zhang for all the in dept information on the C29 population and her instruction on how to deal with data analysis to create linkage maps. She was a great support in times when results were rather discouraging and helped me to stay focused in pursuing my goal.
Further more I want to thank Johann Bucher for all the advice and hands on practical instructions in the lab.
Lastly I want to express my gratitude towards Wageningen University in general, for the opportunity to use lab and greenhouse facilities. In the first place the University made it possible for me to go on a Master program in Plant Biotechnology after my Bachelor in Horticulture, a change other Universities did not permit.

Abstract

The Recombinant Inbred Line (RIL) population C 29 results from a cross between a Japanese turnip and a Chinese Pak Choi.
It is widely segregating for turnip size and shape. To find genes underlying these traits a linkage map was to be created by analyzing SNP markers using high resolution melting curve analysis. A high proportion of linkage distortion in the population proved to be a challenging issue in the process. Marker trait association of turnip formation related traits was analyzed based on phenotypic data from a greenhouse and field experiment.
The linkage map I created consists of 21 Linkage groups (LG). The map remained fragmented, compared to the reference genome, marker order and position of some markers contradicted the physical position.
Twenty-two markers were found to be associated with turnip formation related traits. Three were involved in turnip weight, two in turnip length and 18 in turnip width. One of the markers was associated to both turnip weight and length.

Table of content

Figures

Tables

Abbreviations

ANOVA	Analysis of Variance
B.	*Brassica*
bp	basepairs
cM	centi Morgan
DH	Double Haploid
dNTP	Deoxyribonucleotide-tri-phosphate
exp.	example
Fig.	Figure
LG	Linkgage group
Mbp	mega base pairs
MQ	Milli-Q (ultrapure water)
PCR	Polymerase Chain Reaction
QQ plot	Quantile-Quantile plot
QTL	Quantitative trait locus
RIL	Recombinant Inbred Line
rpm	revolutions per minute
SD	Segregation Distortion
SDL	Segregation Distortion Locus
SNP	Single Nucleotide Polymorphism
Tab.	Table

1 Introduction

Several crop-types of *Brassica rapa* L. are used worldwide for human consumption and as animal fodder. The main crops are oil seed types, turnip types and diverse leafy vegetable types like Chinese cabbage and Pak Choi.

The wild *B. rapa* originates from mountain ranges near the coastal regions of Europe. (Tsunoda 1980 as in Hirai & Matsumoto, 2007) It thrives in cold climates which makes domesticated forms especially valuable as vegetables to grow in temperate regions which can be harvested late in the season. Tsunoda 1980 states (as in Hirai & Matsumoto, 2007) that wild *Brassica rapa* seeds were brought to Asia Minor along the Silk Road. Cultivation history may have started later in two distinct centres, one in Europe and one in India. Most turnip types originate from Europe, whilst oil type *B. rapa* was selected in India. (Golméz - Campo, 1999)
Oil type *B. rapa* was grown in India at least since 1500 (Jonathan D. Sauer, 1993) but today no wild types grow on the Indian subcontinent. (Rakow, 2004)
Linguistic facts reveal that *Brassica rapa* was already domesticated as a turnip type in Mesopotamia 1800 BC. Greek and Roman authors also mention turnips as crop species. The oldest excavated turnip was found in Sparta - Greece and dates back to times of the Byzantine Empire (512 AD). (Reiner, Holzner, & Ebermann, 1995) Again linguistics and historic documents indicate that oil type *B. rapa* were present in Europe for a long time too. Old Dutch documents testify the presence of rape seed for oil production in the 14th century. Nowadays vegetable turnips are used mostly in Europe but also in Japan and China.
Leafy *B. rapa* types form the most diverse group, they diversified mostly in China where a vast range of heading and loose leafy cabbage varieties were bred over the centuries. Chinese white cabbage Pak-Choi and celery cabbage Pai-ts'ai date back to the 5th and 6th century AD. Today heading and non heading cabbages are consumed all over the world and can be seen as one of the most important vegetables for human consumption. (Zhao et al. 2004)

Interestingly a study by Zhao et al. (2005) showed that the different crop types of *Brassica rapa* are often more related to different crop types from the same region, as to cultivars of the same crop type, but from regions far away. This suggests that only a few genes are involved in the expression of the different crop types and that these genes/alleles are present in *B. rapa* accessions in many regions in the world. The Japanese vegetable turnips, which represent one of the parents of the screening population, show more relation to winter oil type cultivars from Pakistan then to turnip types from Europe. (Zhao et al., 2005)
Dixon (2007) provides a second possible theory about evolutionary history of *B. rapa*. He states that *B. rapa* originates from the fertile crescent - todays high plateau of Iran and Iraq.

1.1 Genetics

The evolutionary history of the main *Brassica* crop species is well described in the triangle of U scheme (Nagahura 1935 as in Zhao et al., 2005). The diploid species *Brassica rapa* (n=10), *B. nigra* (n=8) and *B. oleracea* (n=9) gave rise to three amphidiploid species; *B. juncea*, *B. napus* and *B. carinata*.
B. napus is today the most important oil seed *Brassica*. *B. juncea* and *B. carinata* are also used for oil production; secondarily *B. juncea* is used as a green vegetable and tuber crop too.

Genome comparison revealed that a common ancestor of *B. rapa* and the close related model species *Arabidopsis thaliana* went through genome multiplication. Today we find most of the genome of *B. rapa* triplicated relatively to *A. thaliana*. Similar gene blocks of *A. thaliana* and *B. rapa* are synthenic. Genes similar in *A. thaliana* and *B. rapa* are described as orthologs in literature. The three copies of the *A. thaliana* genes in *B. rapa* are then called paralogues within *B. rapa.* (X. Wang et al., 2011) The first complete genome sequence of *B. rapa* was published in 2011 by Wang et al. They sequenced the Chinese cabbage cultivar Chiifu 401-42. The fact that a Chinese cabbage was sequenced first, points out that Chinese cabbage is supposedly the most important crop in *B. rapa*. The de novo assembled sequence spans 283.8 Mbp and covers more than 98% of the euchromatin genome.

Brassica rapa is naturally an outcrossing species, cultivars are mostly self incompatible. Inbreeding of *B. rapa* plants is possible to some extent, but inbreeding depression is likely to occur. (Waller, Dole, & Bersch, 2008) Most oil type cultivars and many leafy vegetable *B. rapa* are annual plants whereas especially European turnips are biannual.

1.2 SNP marker use in modern genetic studies

Single nucleotide polymorphisms (SNP) are an abundant source of molecular markers in all living organisms. SNP are used to distinguish alleles and genes between related organisms. Since next generation sequencing methods make it easy and cheaper to sequence organisms, SNPs can be identified whenever sequence information of two individuals of the same species is available. SNPs are present throughout genomes and with high density. Since a single base is used for allele distinction it is simple to design specific primers, even in genome regions of high similarity. The overall abundance of SNPs make them a suitable tool for linkage mapping and QTL analysis.

1.3 Linkage mapping in *B. rapa*

Linkage maps are usually inferred from populations segregating for at least one trait. Mostly wide crosses between cultivars, or inbred lines, of distinct phenotypic characteristics are made. The parents can be of the three different types. Li, Kitashiba, Inaba, & Nishio (2009) crossed Yellow Sarson an oil seed cultivar with Osome a Japanese leafy vegetable. The F2 progeny was segregating for flowering time and temperature induction of flowering. Yu et al. (2012) created a linkage map from a cross of a Chinese cabbage and a European turnip. Yu et al. (2009) crossed two Chinese cabbage cultivars to create a linkage map. In this case the progeny was segregating for resistance to downy mildew. The Brassica research group of Wageningen University used a double haploid population created from a cross between Pak Choi and Yellow Sarson to create a linkage map. (Xiao et al., 2014)
Approximately 30 linkage maps have been created for *Brassica rapa* in the past. (Yu et al. 2012)
Linkage mapping is based on recombination frequencies between markers from each chromosome. From these frequencies a data based genetic distance in cM is inferred. Only markers on the same chromosome can be linked and therefore group together. The display of linked markers ordered by distance, based on recombination frequency, results in a genetic linkage map. Commonly linkage maps are constructed from populations comprised of double haploid lines, F2 plants or more advanced recombinant inbred populations.
Comparison of different cultivars of the same species can reveal breeding history and relatedness. The biggest advantage of a linkage map is the possible use for QTL identification. Linkage maps contain valuable information for advanced breeding strategies. Recombination frequencies between markers are a great tool to estimate the number of progeny to be screened in breeding programs for

instance for introgression breeding. Introgression breeding can be applied for a number of purposes of which improving resistance to biotic an abiotic stress is one of the most important. Through several rounds of marker assisted breeding, pyramiding of genes is also possible.
Previous linkage maps of *B. rapa* depicting all 10 chromosomes exhibit a total distance between 1000 and 1400cM. (F. Li et al. (2009), Yu et al. (2012), X. Li et al. (2010), Iniguez-Luy et al. (2009), Yu et al. (2009), Yang et al. (2007)
Increasing difficulties in advanced recombinant inbred lines due to inbreeding depression was not reported in literature.

1.4 Association studies

The underlying principle of marker trait association is Linkage Disequilibrium between genes involved in expressing the relevant phenotype and a molecular marker in the vicinity of these genes. Turnip formation is a quantitative trait with several genes involved in its expression; therefore we expected to find several associated markers (QTL markers). In marker trait association statistical analysis is done per marker with no connection between markers and their position. The only possibility to locate QTL markers is to locate markers on the physical genetic map of *Brassica rapa*. This information could be contributed by a Linkage map of high density. The combined information of marker trait association and recombination frequencies between markers could be used in more precise interval mapping procedure for QTL detection.

1.5 Aim of the thesis work

All analysis and experiments of the thesis work were conducted to create a new Linkage map for the RIL population C29. The aim was to saturate the map with an average marker density of one marker every 5cM. Secondly Marker trait association was to be conducted to find molecular markers associated with turnip formation related traits such as turnip weight, width and length. These traits have shown clear segregation among the C29 progeny, observed in a field experiment and in a greenhouse experiment in autumn 2012.

2 Material and Methods

2.1 Plant material

The parental lines VT 115, a Japanese turnip, and PC 105, a Chinese Pak Choi with high similarity to Chinese Wutacai cabbage, were crossed. The F1 population was planted out in the greenhouse. A single F1 Plant was selfed. F2 plants were selfed and grown on as separate inbreeding lines through single seed descent. Two more rounds of selfing led to F4 recombinant inbred line plants.
In the course of the breeding program several lines were lost due to reduced vigour and or sterility. From an initial 222 F2 lines 35 were lost or failed to set fertile seeds. From the 187 F3 lines only 125 produced seed for F4.

To create double haploid plants, F1 plants were used. Pollen was taken and used for microspore culture. In this process haploid pollen mother cells are cultured in-vitro. In the medium they redirect their developmental pathway and become omnipotent. Many cells double their set of chromosomes and eventually regain the stable diploid state, now completely homozygous since the homologous chromosomes are perfect copies. Embryos are derived from the double haploid cells which can differentiate and form new plants.

Compared to the completely heterozygous F1 plant the F4 RIL plants are expected to show heterozygosity on 12.5% of all SNP loci. (Tab. 1)

Tab. 1: Genotype frequencies following Mendelian segregation; a for homozygous a alleles; b for homozygous b alleles and h for heterozygous ab alleles

	a	h	B
F1	0%	100%	0%
F2	25%	50%	25%
F3	37.5%	25.0%	37.5%
F4	43.75%	12.50%	43.75%

The parents are landraces and not homozygous for all loci since the cultivars are self incompatible they need to be propagated by crosspollination of plants from the same landrace.

Purified DNA of two plants from the parental accessions, 108 Recombinant inbred lines (RIL) in F4 and 9 double haploid lines (DH lines) was provided by the thesis supervisors. DNA samples were diluted to a concentration of 10-20ng/µl. The samples were arranged in three 96 well plates later referred to as pre screening, main screening and `rest` set. 200µl per sample were pipetted in each well.
The pre screening set was comprised of DNA of the two parental cultivars, and 10 F4 RILs.
The main set was comprised of DNA samples of 85 F4 RIL and 9 DH lines. The `rest` set was comprised of the remaining 13 F4 RIL DNA samples.
Each set was split in 4 equal copies with 50µl solution per sample. This ensures that samples show equal concentrations and qualities throughout the experiments. In case of spoiling or spilling of DNA samples backups are provided this way.
The described handling of the DNA samples was meant to make later preparation on PCR reactions more efficient, multichannel pipettes could be used to set up 8 or 12 reactions at once.

2.2 SNP data

A table of 1028 SNP markers polymorphic in comparison of 4 different cultivars (Turnip, L144 (rapid cycling), L143 (yellow sarson), Wutacai) was kindly provided by the science group of Prof. Wang Xiaowu of The Institute of Vegetables and Flowers - Chinese Academy of Agricultural Sciences. The SNP markers are evenly distributed over the genome of *Brassica rapa*. Flanking DNA sequence of 100bp left and right from the SNP position was also included in the dataset. The dataset was used to design additional primers for screening markers using Lightscanner® technology.

My thesis supervisor Dr. Ningwen Zhang selected a subset of SNPs, evenly distributed over the genome. K-Bioscience was ordered to screen these over 250 *B. rapa* accessions from a core collection, plus the parents of the different populations and 24 of their progeny. 170 SNP markers were polymorphic in C29. The 117 DNA samples and primer information (170x) were send to K-Bioscience for SNP profiling. The data from K-Bioscience was used for mapping to form the backbone of a linkage map later to be enriched by additional markers.

2.3 Primer design

To enrich the first draft linkage map created from the SNP profiling data additional primers were designed for more markers.
For each marker forward and reverse primer needed to be designed for amplification of the genome sequence flanking a SNP. For primer design, the free Primer3 online program was used with the following parameters:

- SNPs were selected from the dataset which don't show additional polymorphisms 100bp up- and downstream of the SNP location.
- The optimal length of the amplification product is 90 to 120bp. Amplification of the flanking 20bp left and right of the SNP was secured, primers were supposed to be found outside these borders.
- Optimal primer size was set to 18-23bp.
- Melting temperature was set on 59-61°C with optional changes to 57-63°C if no fitting primer was found.
- The maximum GC content was set to 50%.
- Primers with a high tendency to form dimers because of self complementarity were not used.

If no primer was found with these custom settings, the frame of secured amplification around the SNP marker was reduced to 10bp up and downstream of the marker. If this didn't result in an applicable primer pair the melting temperature was altered as described above. To increase the primer size to 25bp was considered as last option to find a useful primer combination. Primer combinations based on SNP data from the Chinese Academy of Agricultural Science and the resulting markers will later be referred to as IVF (Institute of vegetables and flowers) marker set.
A list of all primers for the IVF marker set is attached in Appendix C.

2.4 Lightscanner® reaction and analysis

To analyze the segregation of marker alleles in the screening set of F4 and DH lines the Lightscanner® Technology was used. Lightscanner® is a high resolution melting curve analysis device. The principle of this technique is to identify polymorphisms between PCR amplification products of equal length by difference in the melting temperature. The melting temperature is greatly defined by the GC content

of an amplification product. Change of a single base in a PCR product of optimal length results in a detectable change of the melting temperature. Melting temperature is detected by fading of fluorescence. A fluorescent dye is incorporated in the DNA double helix, when DNA is denatured the fluorescence fades. Temperature in the scanner is increased subsequently from 70°C to 95°C fluorescence of the probe is detected optically and depicted in a graphic. The machine is constructed to analyze special 96 well plates. The plastic framework is made of black plastic to reduce reflection of light on these parts of the plate; wells are made of white plastic to facilitate reflection of fluorescence light.

2.5 Marker screening over C29

A number of SNP markers was used to screen the C29 population. Additionally some markers designed by Xiao Dong were used. These were designed to amplify regions of genes, homologous to *A. thaliana* genes associated with leaf developmental traits. The connection between gene and leaf development in *A. thaliana* was mostly established through knock out mutagenesis. Several markers were designed to amplify different paralogous regions of these leaf trait genes present in *B. rapa*.(Xiao et al., 2014)

After primer design each primer combination was tested with the pre screening set of 10 RILs and DNA from both quasi parental cultivars.
The Lightscanner® PCR reaction is set up in 96 well PCR plates. 8 markers can be analysed with the pre screening set per PCR reaction. If polymorphisms were found among the individuals of the pre screening set for a marker, the primer combination was also used to screen the main screening set and the 'rest' set. The PCR reaction mastermix was prepared according to the Lightscanner® protocol (Appendix A). The mix was then pipetted into the Lightscanner® 96 well plates, 1µl of template DNA solution was added to the reaction mix. Finally it was covered with 20µl of mineral oil. The purpose of the mineral oil is to prevent evaporation during the Lightscanning procedure as during scanning the samples will be heated to 95°C. After pipetting everything together the 96 well plates are sealed with aluminium film and the PCR is run according to the protocol. (Appendix B)
To analyze marker segregation of the lines the plate is inserted into the Lightscanner® machine after PCR. Fluorescence of the probes is temperature dependent, to auto adjust exposure time for the camera the machine heats quickly to 68°C and produces a quick shot of the plate, optimal exposure will be depicted by a red dot in the middle of each well on the PCR plate. After this process described as 'ramping to exposure temperature' the temperature is risen to the lower level of the analysis temperature range. The lower temperature was set to 72°C, at this temperature the Lightscanner® machine starts recording fluorescence, the temperature is then gradually raised to the upper analysis temperature, which was set to 95°C, recording fluorescence in the process.
After several scanning procedures an acceptable range of fluorescence between 800 and 2000 was defined. All samples with lower fluorescence were not scored later on. Acceptable analysis results look like the output below. (Fig. 1)

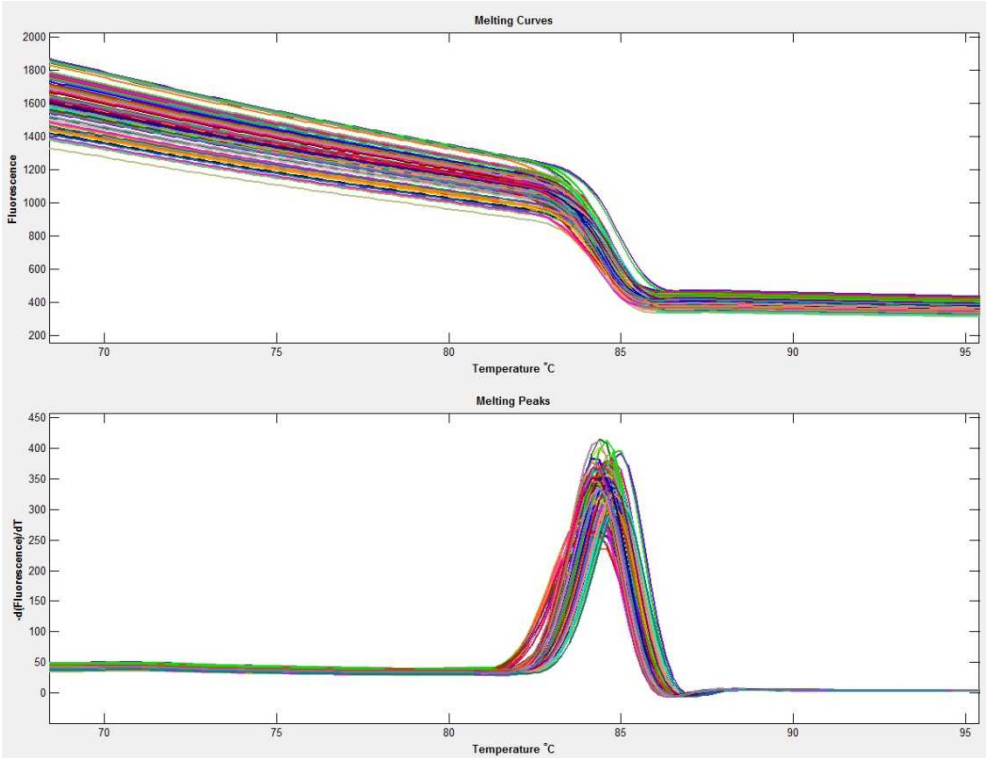

Fig. 1: Raw Lightscanner® output SNP_0857

The output graphics are normalized for each plate by selecting a region of parallel fluorescence graphs, the intensity of fluorescence is then normalized for all probes on the plate. Probes showing equal gradients of the fluorescence curves are grouped automatically by the Lightscanner® software. (Fig.2)

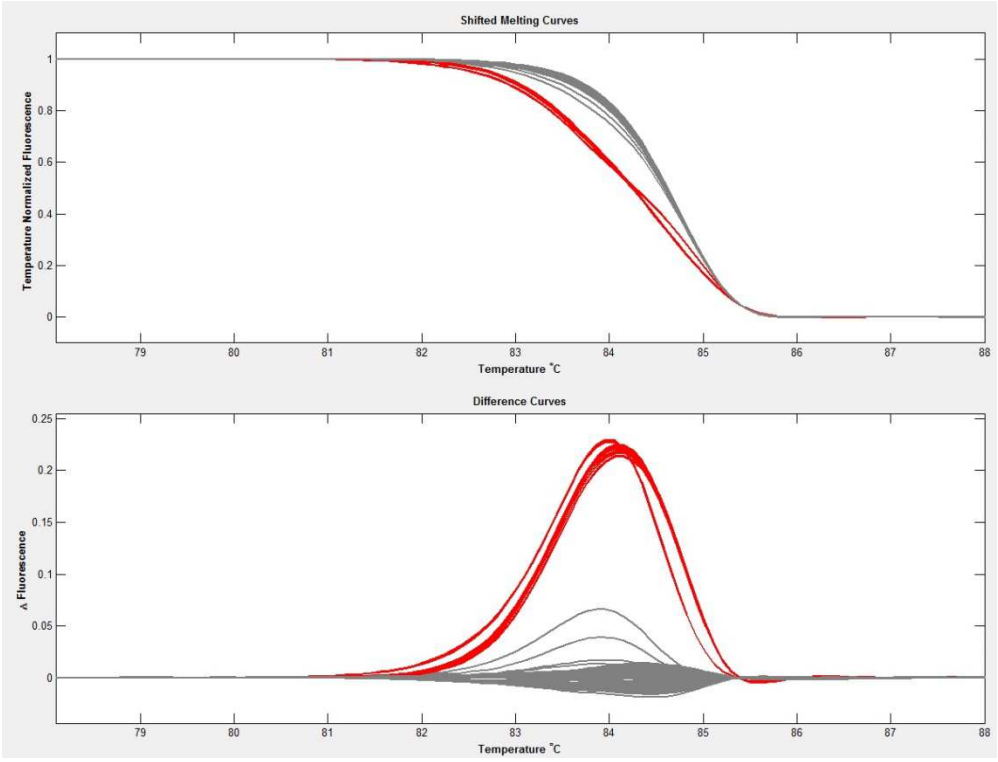

Fig. 2: Normalized Lightscanner® output SNP_0857, two distinct groups

As each marker is screened over DNA samples from 3 plates, the plates used to analyse the same marker are normalized equally to secure consistent grouping. The grouping has to be equalized manually, since software assisted grouping cannot be applied for several plates at the same time.

2.6 Denomination of alleles

To create a linkage map from marker information of a population the exact denomination of the parental alleles is crucial. Unfortunately no DNA from the parents of the C29 population was sampled. Therefore the denomination needed to be inferred from the reads of the quasi parental cultivar accessions; since these cultivars are heterogeneous landraces the information can be regarded as a hint only. Graphical genotyping and marker order in agreement with the physical position of the marker contributed the rest of the information for a consistent denomination. In the graphical genotyping table (Tab. 2) markers are ordered by their physical position based on the reference genome. The allele denomination is most probable in the constellation with least recombination events between neighbouring markers.

Tab. 2: SNP table to determine parental alleles; green cells VT115 allele, red cells PC105 allele

marker	VT-115	PC-105	PC-105	VT 105 derived	PC 105 derived	F4-53	F4-54	F4-55	F4-56	F4-57	F4-58	F4-59
SNP_0002	A:A	G:A	G:G	A:A	G:G	A:A	G:G	A:A	A:A	G:G	G:A	A:A
SNP_0003		C:T					C:C	T:T	T:T	C:C	C:T	T:T
SNP_0005	G:G	A:G		G:G	A:A	G:G	A:A	G:G	G:G	A:A	A:G	G:G
SNP_0014	G:G	G:A	G:A	G:G	A:A	A:A	A:A	G:G	A:A	A:A	A:A	A:A
SNP_0015	T:T	T:T				C:C	T:T	T:T	T:T	T:T	C:T	C:C
SNP_0017		G:G		A:A	G:G	G:G	G:G		G:G	G:G	A:A	
SNP_0029		T:C				T:T	T:T			T:T	C:C	C:C
SNP_0033	C:C	C:C	C:C			T:T	C:C		T:T	C:C	C:C	T:T
SNP_0037	A:A	G:A		A:A	G:G	A:A	A:A	A:A	A:A	A:A	A:A	A:A
SNP_0055	G:G	A:A	A:A	G:G	A:A	G:G	G:G	G:G	G:G	G:G	G:G	G:G
SNP_0058	C:T	C:C		T:T	C:C	T:T	T:T	C:C	T:T	T:T	T:T	T:T
SNP_0059	C:T	C:C	C:C	T:T	C:C	T:T	T:T	C:C	T:T	T:T	T:T	T:T
SNP_0060	C:C	T:T		C:C	T:T	C:C	C:C	T:T	C:C	C:C	C:C	C:C
SNP_0086	A:G	G:G		A:A	G:G	G:G	A:A	G:G	A:A	A:A	G:G	G:G
SNP_0089	T:T	G:G		T:T	G:G	T:T		T:T			G:G	T:T
SNP_0098	T:T	A:A	A:T	T:T	A:A	T:T	T:T	A:A	T:T	T:T	T:T	T:T
SNP_0101	G:G	A:A		G:G	A:A	G:G	G:G	G:G	G:G	G:G	G:G	A:A

The table above serves as an example, columns 2-4 show base pair identity of the quasi parental accession, on the SNP position. This information is used to reconstruct the alleles of the actual parents, which are named derived parents (columns 5 and 6).
However the information from these quasi parental genotypes is not always clear without ambiguity because of possible heterozygosity as for SNPs SNP_0002, _0003 for example. In case of heterozygosity of only one quasi parental accession inheritance can still be determined. To make the inferred information clear two more columns were added to the table, named as derived VT115 and derived PC105. In some cases quasi parental accessions could not be analysed for all markers or

VT115 and PC105 gave identical reads. Since the parental accessions were landraces as mentioned in chapter “Plant material” this case is not improbable, the real parents may have had different alleles resulting in the polymorphism in the progeny. However in some cases (as in SNP_0015, SNP_0017, SNP_0029 and SNP_0033) no clear denomination based on quasi parental accession is possible, therefore allele denomination needs to be inferred from flanking markers.

Tab. 3: Graphical genotyping table - SNPs ordered by physical position, allele denomination in accordance with Joinmap requirements

marker	position	F4-140	F4-142	F4-143	F4-144	F4-147	F4-152	F4-153	F4-154
SNP_0253A03	10871276	b		a	a	a	a	a	a
SNP_0255A03	11342547	h	b	a	a	b	b	a	a
SNP_0264A03	14770684	b	b	b	b	b	b	b	b
SNP_0267A03	15435114	h	a	a	b	a	a	b	a
SNP_1050A03	18284950	a	a	a	a	a	a	a	a
SNP_0284A03	19709518	a	a	b	a	b	a	b	h
SNP_0286A03	20138490	b	b	a	b	a	b	a	h
SNP_0297A03	22926755	b	b	a	b	a	b	a	h
SNP_0298A03	23143245	a	b	a	b	a	b	a	a
SNP_0309A03	26161200	a	b	a	b	a	b	a	a

Because of linkage disequilibrium it is likely that flanking markers show alleles of the same parent in each line. Recombination can of course occur but recombination on both sides at once is unlikely. (as in SNP_0284A03) Alleles of flanking markers are monitored over all lines and used to estimate recombination events by eye. If denomination of a marker compared to his immediate neighbours shows unusually many recombination events and the reverse denomination would show less recombination than the case with less recombination is always more probable. This is most obvious in cases were denomination of one marker appears to be the exact mirror image of the flanking ones. (exp. SNP_0284) The larger the distance between two markers the more unclear this method becomes because true recombination impairs the picture.
As in the table above, it would make more sense if the allele denomination would be switched in the marker SNP_0284A03, maybe even SNP_0267A03, SNP_1050A03 and SNP_0284A03 all together.

Grouping of markers with Joinmap statistical software (described in chapter 2.7 Mapping) is not affected by false denomination among markers. Changes as described above were only done in cases were markers belonging to the same linkage group have shown unusually large marker distance. False denominated markers appear as if recombination has taken place on most of the markers which is improbable. Marker distances in cases like this could even exceed 10000cM.

2.7 Mapping

To create the linkage map the program Joinmap 4.1® (Van Ooijen, J. W., 2011) was used. The first step towards the map is to create a table for data input in which all RIL lines and marker information is converted to a, b and h letters. The”a” (VT115) and “b“(PC105) stand for the parental alleles, h indicates heterozygosity, these lines carry both alleles, one on each homologous chromosome. (Tab. 3) Based on the input data independence LOD was calculated for each marker. A threshold of 3.0 was set. The independence LOD test is not affected by segregation distortion (Van Ooijen Join Map manual), false denomination of parental genotypes also doesn`t affect grouping. Since the

independence LOD can be compared to a chi² test for independence which tests proportions of alleles per marker, markers with no significant difference in their allele proportions are grouped together. The test doesn't depend on the denomination of the allele (a, b or h) markers will be grouped together if the proportions are not significantly different.

Segregation distortion was tested with Chi²-test for goodness of fit. Marker with significantly different segregation from Mendelian principles will have a p-value bigger than 0.05, we will reject the null hypothesis. Null hypothesis in this case assumes resemblance of observed genotype frequency and genotype frequency according to Mendelian segregation.
After grouping, group nodes of 3 or more grouped markers were selected for map calculation. The maximum likelihood mapping algorithm was used to calculate recombination frequencies. Kosambis mapping function was then applied to translate recombination frequencies into marker distance.

2.8 Collection of phenotypic data

A greenhouse and field experiment was conducted to access phenotypic difference of F4/F5 C29 lines. In autumn 2012 nine DH lines and 96 F4/F5 lines were planted in three blocks.
In the greenhouse experiment one plant of each line was planted per block. Phenotypic data of turnip length, width and weight was accessed.
In the field experiment 15 plants of each line were planted 5 plants per line were scores. Turnip length and width were measured, an average value of the 5 plants was kindly provided by Dr. Ningwen Zhang from the Brassica research group.

2.9 Statistical analysis of phenotypic data

Analysis of Variance (ANOVA) was used to analyse phenotypic data from field and greenhouse experiment. ANOVA was also applied to access difference between blocks of the experiments.
For some lines data is missing since some plants died. To ensure that these lines could still be included in the analysis an unbalanced model was chosen for the ANOVA. In unbalanced designs regression is used to predict a means value per line, the means value is then used in the analysis of variance.
ANOVA has to be applied under the assumption of normal distribution of the data and homogeneity of variance among individuals per line. To check these assumptions QQ plots and residuals plots were checked. If data was not normal distributed $\log_{10}$ transformation of the data was applied and ANOVA was run again. All analysis on phenotypic data was conducted with Genstat 15th edition statistical software.
For marker trait association regarding turnip formation ANOVA was used in R. The script used for the analysis was provided by Ram Kumar Basnet and Dr. Ningwen Zhang. In this ANOVA average phenotypic data for turnip width, length and weight (only in the greenhouse experiment) was analysed per marker over the whole population. Per marker means values from individuals with the same homozygous allele (a) were compared to the individuals homozygous for the other allele (b) of the population. Significant difference in this comparison indicates association of a single marker to the trait. Model assumptions were tested in the same way as described above, additionally Box-Plots were created for the phenotypic data across all lines. In case of insufficient fit to normal distribution the data was $\log_{10}$ transformed and the analysis run again.

F-distribution was used for test statistics of all analysis of variance. A significance level of 0.05 was set to the statistical analysis of the phenotypic data. To identify loci associated with turnip formation traits turnip length, width and weight, a $-\log_{10}$ (p-value) threshold of 3 was defined. A significance level of 0.05 allows 5% false positive associations in 246 independent tests (one per marker) this

would mean up to 12 markers could be found to be significantly linked to a trait, but they are false positives. A $-\log_{10}$(p-value) threshold correlates with a p-value of 0.001, this significance level only allows one false positive result in 1000 tests. This stringent criterion helps to identify only the most significant true positive marker trait associations revealing QTLs.

3 Results

3.1 Marker screening

The dataset from K-Bioscience contained 170 polymorphic markers, another 214 (IVF SNP+Leafmarkers) primer combinations were tested with the Lightscanner® in the prescreening. Of the 214 markers 76 proved to be polymorphic in the prescreening and were then used for main screening. In total 246 markers were polymorphic and could be used for mapping. The detailed composition is depicted in the pie charts below (Fig. 3).

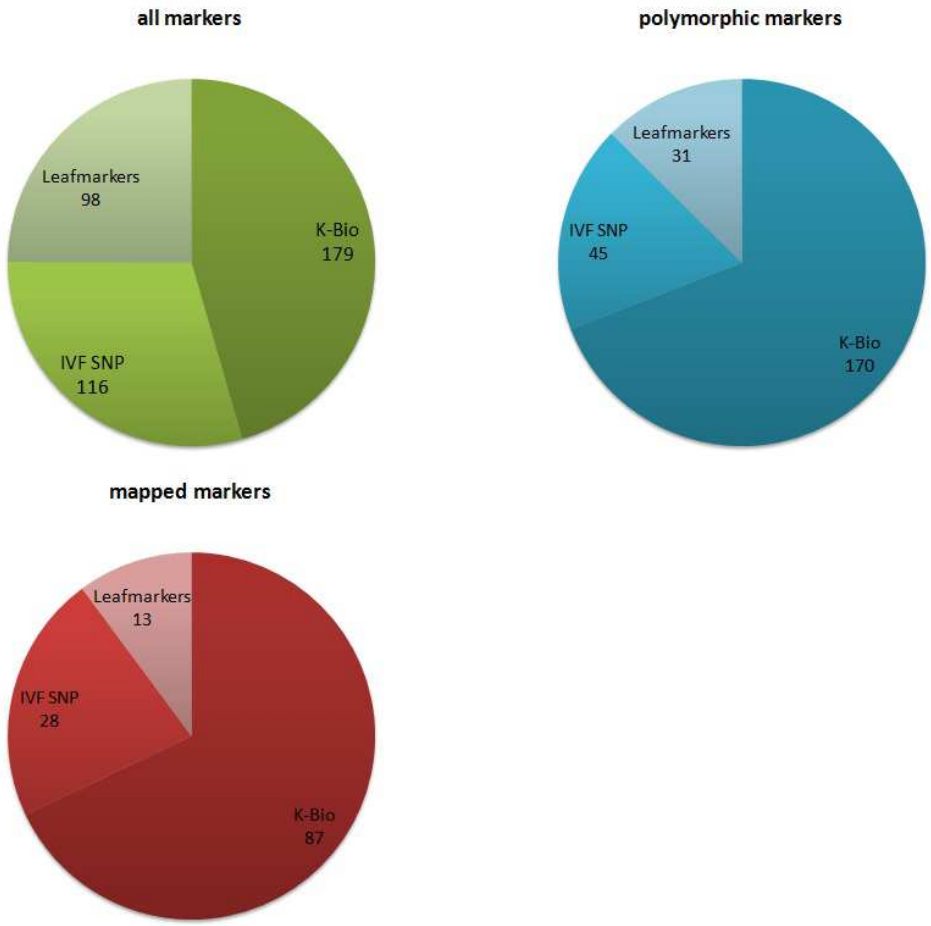

Fig. 3: Numbers of markers grand total (left side), polymorphic markers (right side), and mapped markers (lower left side)

Five different categories of Lightscanner® output can be distinguished in the prescreening. The first category (Fig. 4) is an optimal output showing up to 3 distinguishable groups among the 12 samples. Two of which are characterized by a single steep peak typically indicating a homozygous allele, the third group shows a shouldered peak indicating a heterozygous group.
No more than 3 groups can be present in the screening output, since only 2 alleles can be retained in the population. Each parent is able to produce gametes with up to two alleles per loci if the parent is heterozygous for the loci. Still only one allele can be inherited per parent in the F1. Only one F1 plant was used to produce seeds for the F2 population.

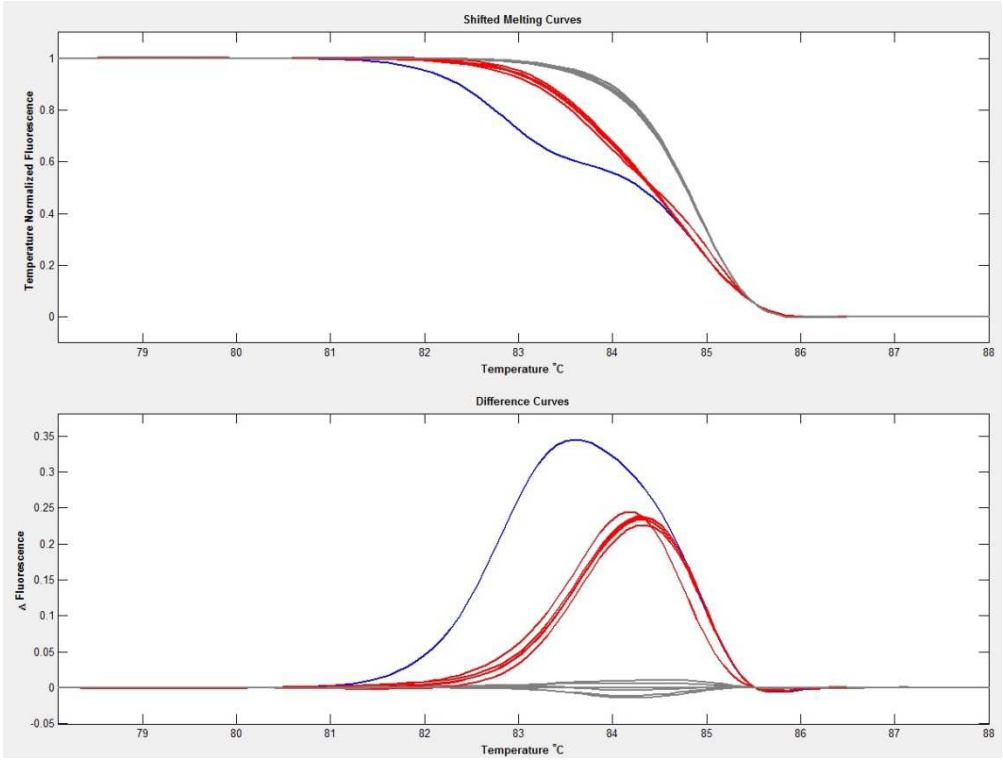

Fig. 4: Normalized Lightscanner® output SNP_0857, red and grey line homozygous groups, blue line heterozygous groups

The second category (Fig.5) shows one peak for all samples in the prescreening set, they all belong to the same group, only one allele is present for this marker among the individuals in the prescreening set. After data normalization the peak is reduced to a more or less straight line. Segregation among the rest of the progeny from C29 is unlikely, if no polymorphisms are found in 10 progeny, Mendelian segregation is unlikely for this marker.

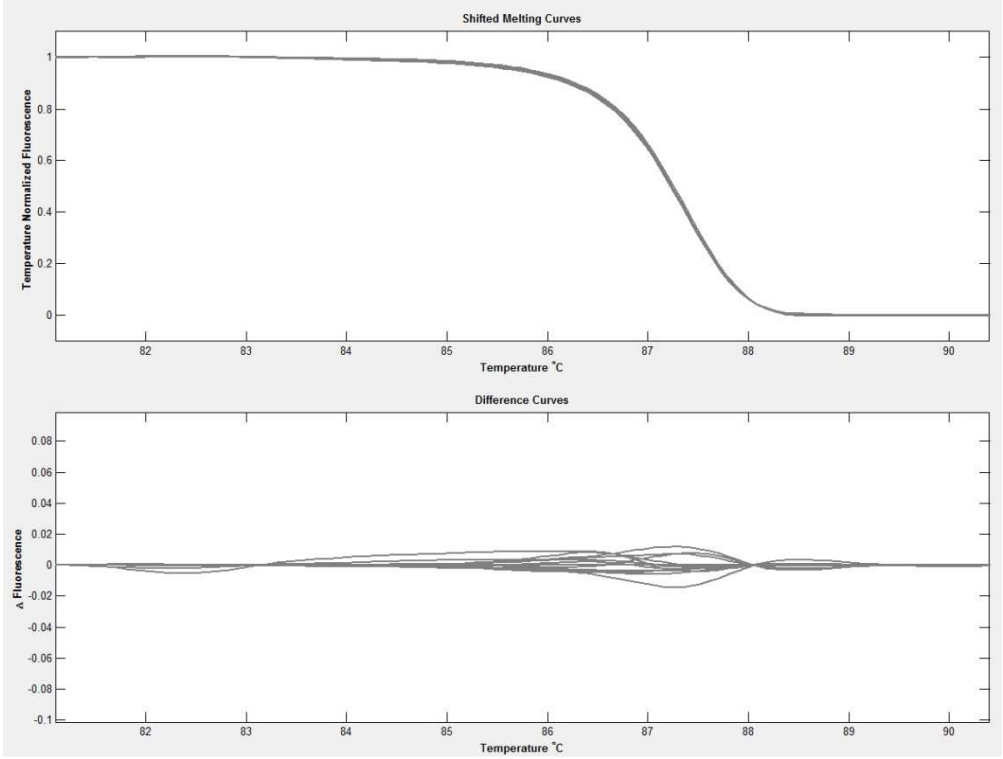

Fig. 5: Normalized Lightscanner® output SNP_0870

The third category (Fig. 6) shows too many but clearly distinguishable groups among the prescreening individuals. An interfering second SNP on the amplification product could be the cause.

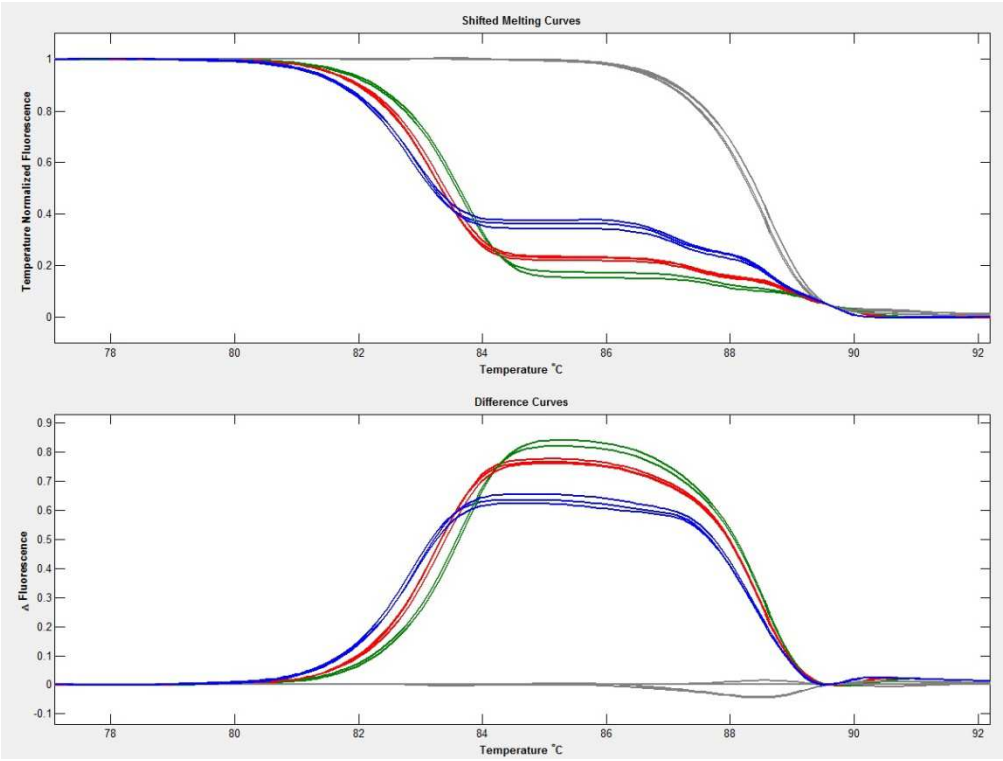

Fig. 6: Normalized Lightscanner® output SNP_0875

The fourth category (Fig. 7) is characterized by a smear like, computational undistinguishable, distribution of fluorescence curves. Possible reasons for this would be highly speculative.

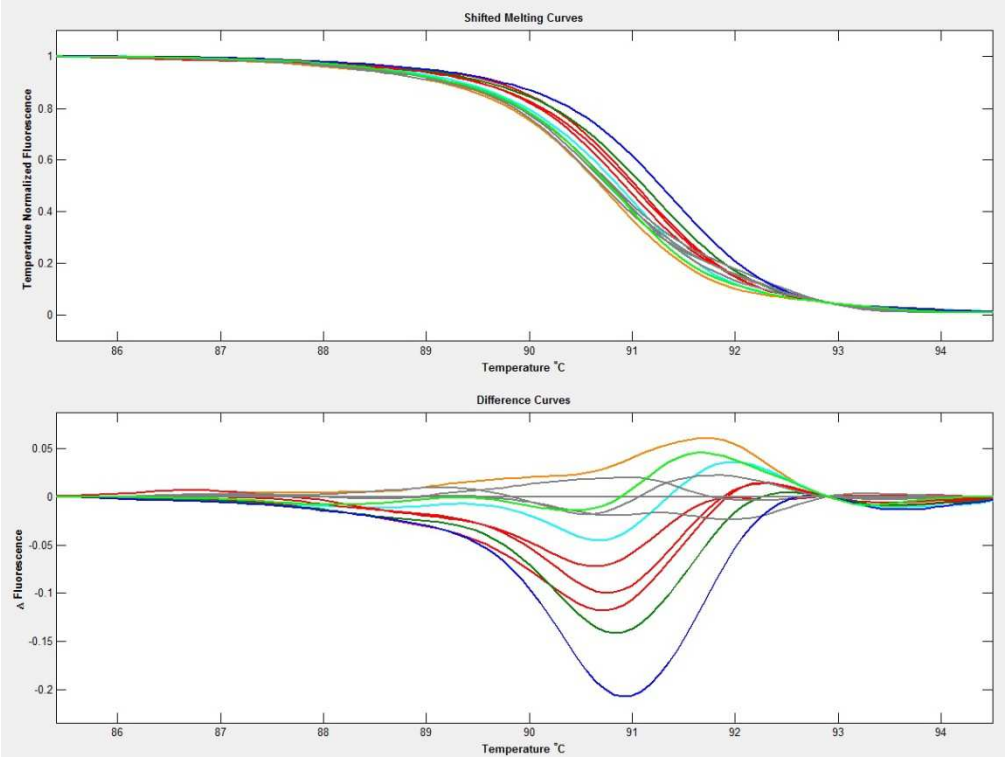

Fig. 7: Normalized Lightscanner® output L_0181

The last category (Fig. 8) is characterized by double peaks, in the fluorescence difference curves, with wide distance between the two tops. Unspecific binding of the primers could be the reason for this output. A second amplification product with higher melting temperature causes the second peak. Normalization of only one peak can result in reasonable grouping however attempts to map markers based on this kind of output are mostly impossible or faulty.

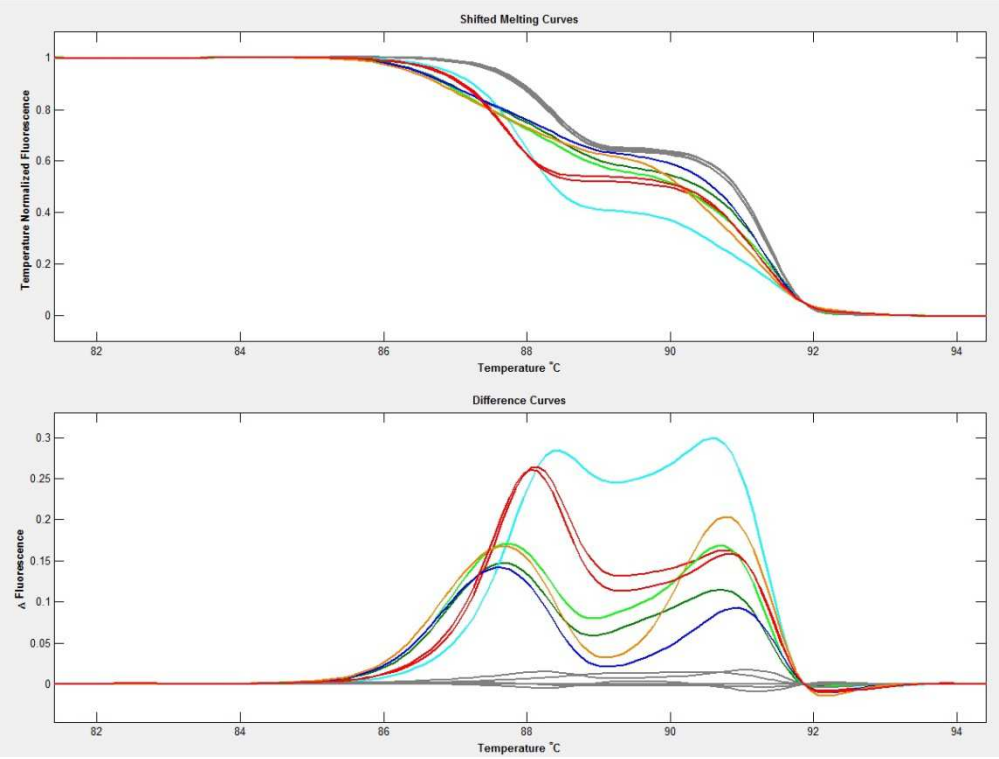

Fig. 8: Normalized Lightscanner® output SNP_0818

A high proportion of the polymorphic markers from both the dataset of K-Bioscience and the markers screened with Lightscanner® show diverging segregation compared to the expected Mendelian pattern (Tab. 1). These markers are addressed as distorted markers in literature. To test for marker distortion a chi² test was applied; in the 0 hypothesis of this test, allele a and b are distributed in a 1:1 ration for each marker. The test was applied at a confidence level of 0.05. (Appendix D)
The test result shows that 184 of the 246 polymorphic markers are distorted. Accordingly 75% of the polymorphic markers are distorted. Distortion was found in SNP data from K-Bioscience (69%) as well as the markers screened with Lightscanner® technique (85%). (Tab. 4)

Tab. 4: Marker distortion per chromosome and in total of the different marker sets

chromosome	All markers	distorted	distorted (%)
A01	28	22	79
A02	25	17	68
A03	22	13	59
A04	18	13	72
A05	18	12	67
A06	16	16	100
A07	27	21	78
A08	27	21	78
A09	39	29	74
A10	26	20	77
K-Bioscience	170	118	69
Lightscanner®	76	55	85

On all chromosomes more than 50% of the markers are distorted, on chromosome A06 all markers are distorted; at the same time the least number of polymorphic markers is assigned to chromosome A06. (Tab. 4) No primers for additional markers were designed for chromosome A06 to get more marker density, hence the few number of markers.

3.2 Statistical analysis of phenotypic data

In a QQ plot two datasets are plotted against each other to compare data distribution.
In this research the trait values (cross points) on the vertical axis are plotted against values of a random normal distributed dataset. If the plot shows a strait diagonal (green line) the experimental data can be considered normal distributed. A perfect fit is very unusual to help in the reasoning. The borders of the 95% confidence limit (red lines) are used to support the decision. These confidence limits are following a Kolmogorov-Smirnov test for normality. The advantage of the QQ plot is that the confidence limits make it easy to evaluate outliers.
Residual plots can be found in Appendix D, the fitted values plot depicts residuals plotted against data values fitted to a normal distribution. If residuals are scattered more or less evenly around the fitted values we can conclude the dataset is normal distributed.

3.2.1 Greenhouse trial

Based on the Normal QQ plot of turnip length, weight and width (Fig. 9) it can be concluded that only turnip length can be considered to be normal distributed. Turnip weight differs the most from a normal distribution. To correct for this and make ANOVA applicable the data was $\log_{10}$ transformed. Turnip width was $\log_{10}$ transformed too although necessity for this is questionable, over all $\log_{10}$ transformed data fits the normal distribution better.

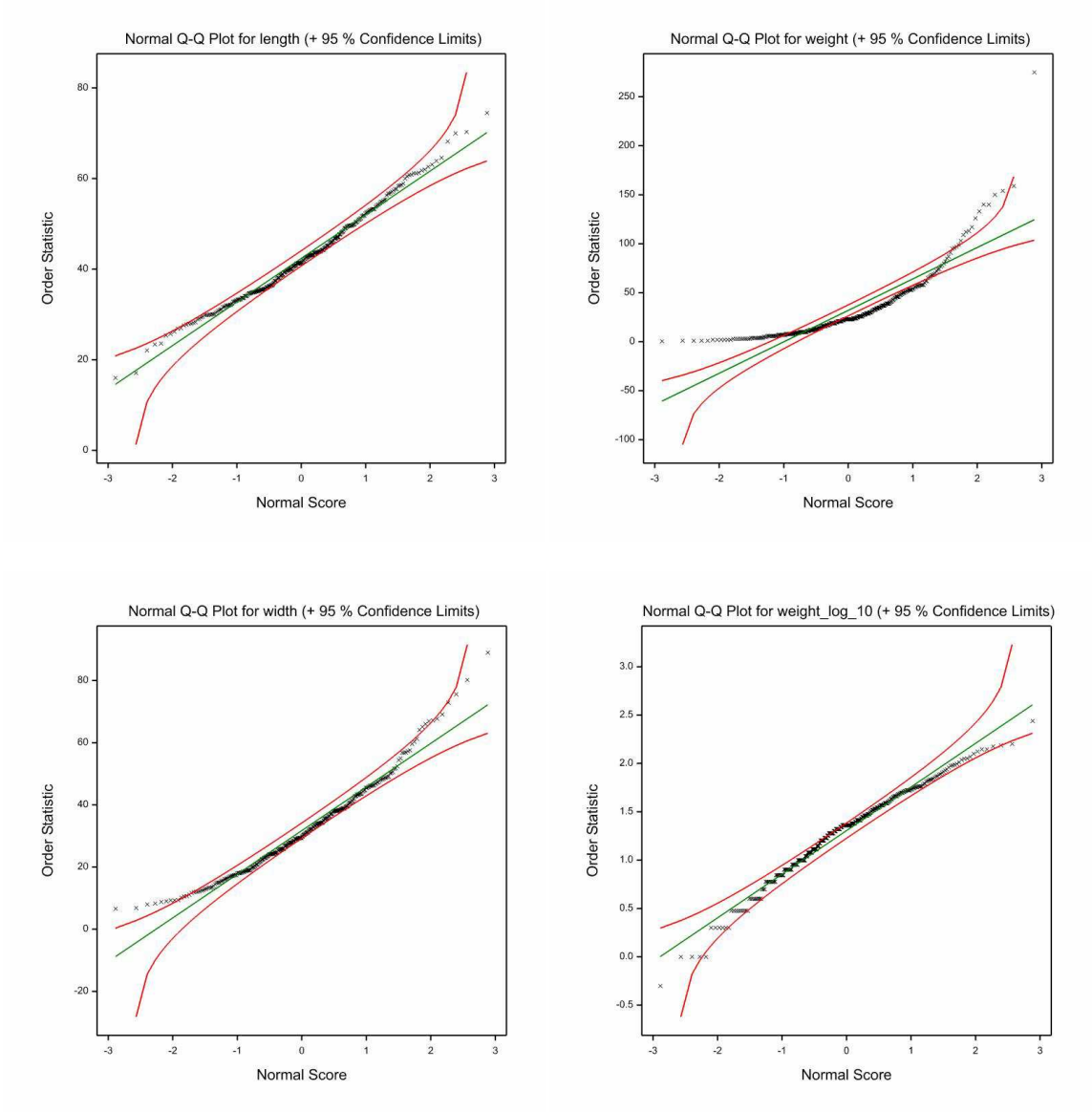

Fig. 9: Greenhouse experiment autumn 2012 - upper left QQ plot of turnip length; upper right QQ plot turnip weight; lower left QQ plot of turnip width; lower right log_{10} transformed turnip weight

ANOVA test shows that lines are significantly different for all 3 phenotypic traits; this indicates a significant genotype effect. (ANOVA tables in Appendix G)

To capture the block effects, ANOVA for length, width and weight over all lines were performed. Blocks are significantly different for all trait values. (Appendix G). To illustrate distribution of data per block additional boxplots were created. The boxplots (Fig. 10) illustrate the difference between blocks remarkably. Turnip weight and width are smallest in Block 2 whereas turnip length is smallest in block 3. Statistical analysis of differential interaction between lines and blocks could not be analysed since only the average value per line per block was available. A variance per line per block can not be deduced from one data point even though it was built from pooled data. Mean trait

values per block in the greenhouse experiment are indicated in table 5 below.

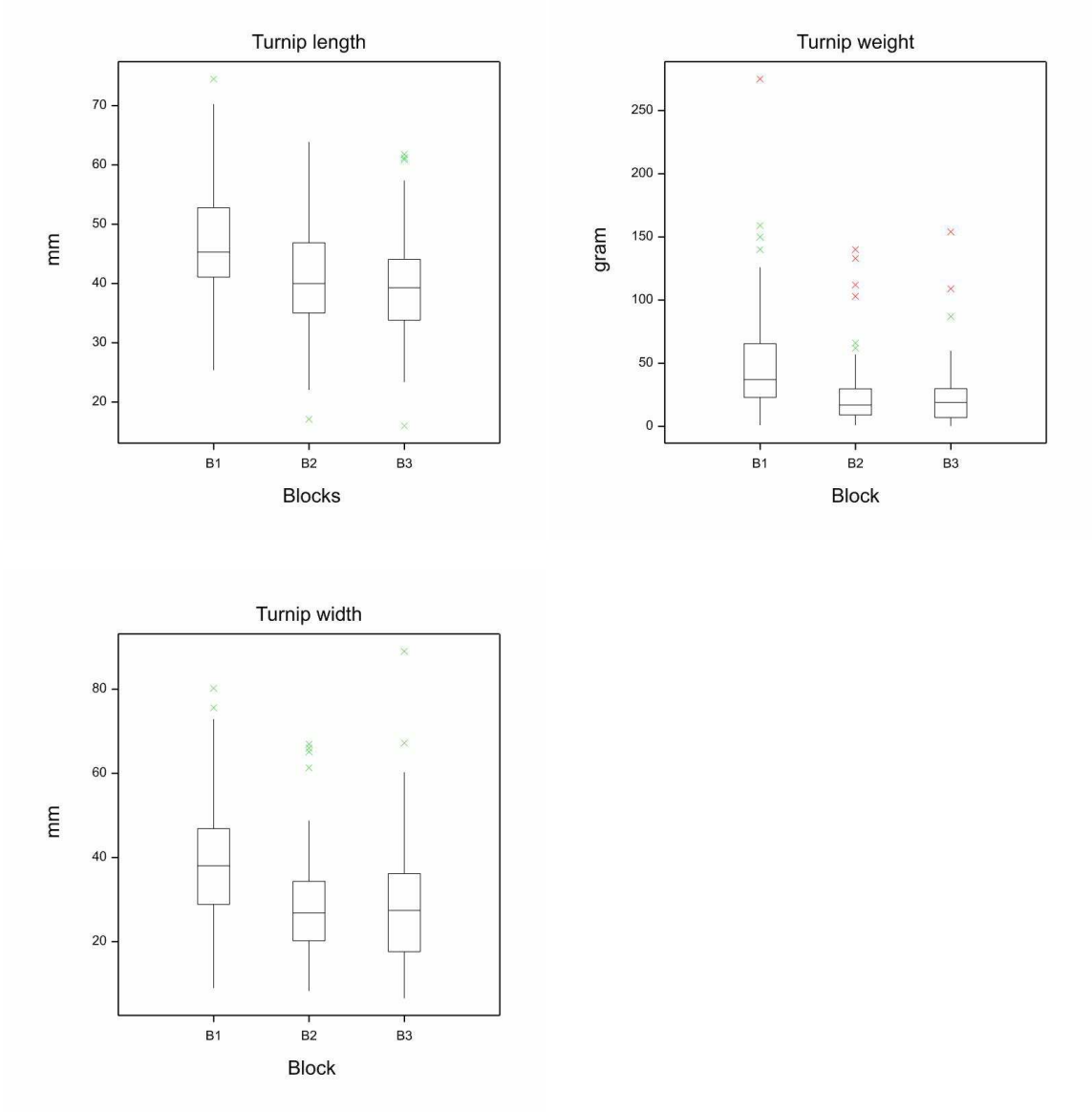

Fig. 10: Boxplot of turnip length (upper left); turnip weight (upper right); turnip width (lower left) measured in the greenhouse in 2012

Tab. 5: Mean values of traits per block in the greenhouse experiment

Block	means value		
	Length (mm)	**Weight (gram)**	**Width (mm)**
I	46.3	48.4	38.5
II	40.9	24.8	28.8
III	40.0	22.8	27.9

3.2.2 Field trial

In the field trial both, QQ plots both point towards a non normal distribution. (Fig. 11)

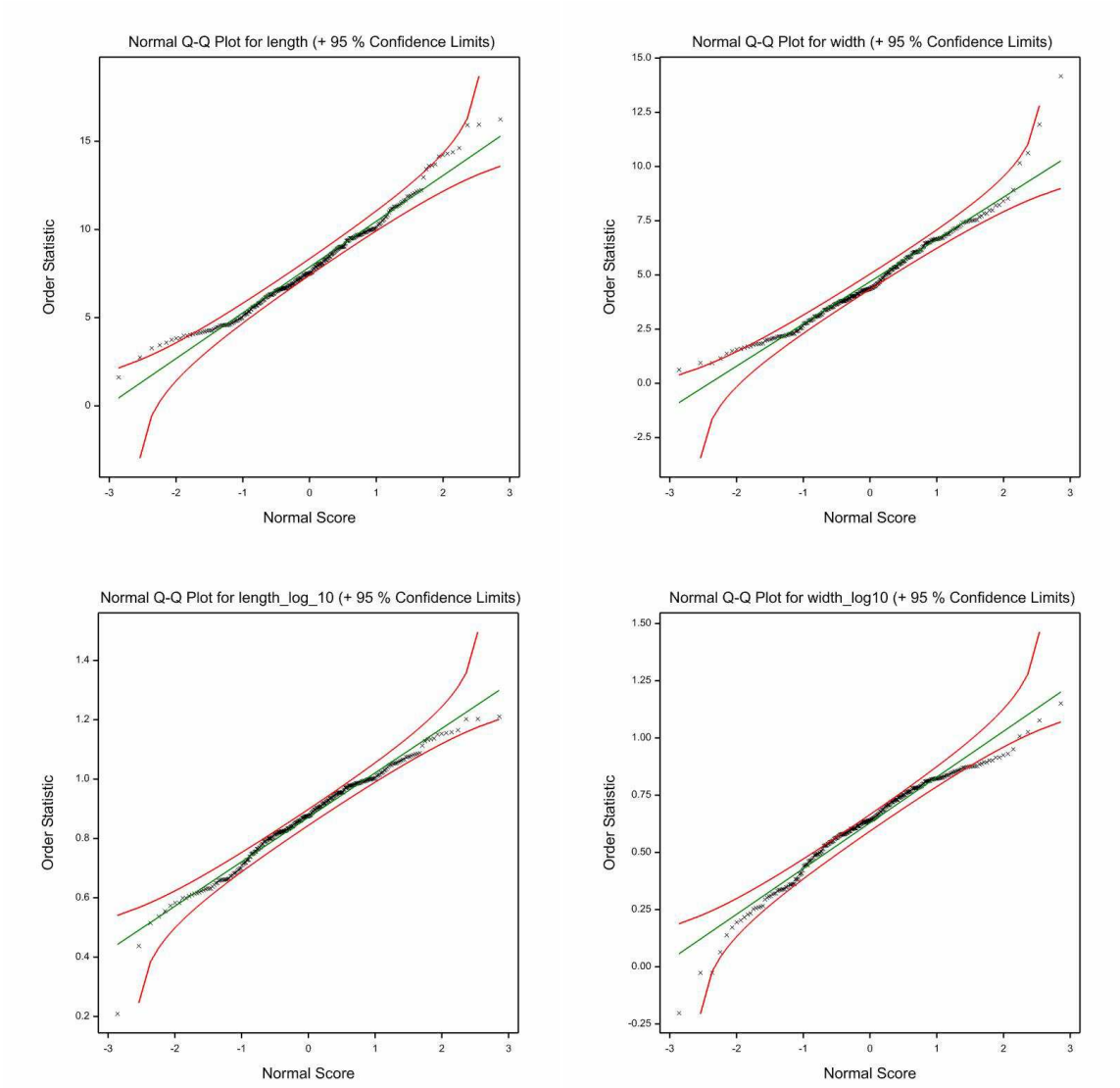

Fig. 11: QQ plot of turnip length (upper left) and width (upper right) measured in the field experiment. Log_{10} transformed data below, improvement of data distribution is best displayed by spread of residuals in the fitted values plot (Appendix F)

Again difference between lines of turnip length and width is significant, indicating a significant genotypic effect. (ANOVA table in Appendix G)
Blocks compared by the average over all lines show significant difference. (ANOVA table Appendix G)
Interaction of line and block could not be accessed for the same reason as mentioned before. The boxplots show very clearly that plants were growing strongest in block 3 and less vigorous in block 2 and least strong in block 1. (Fig. 12)

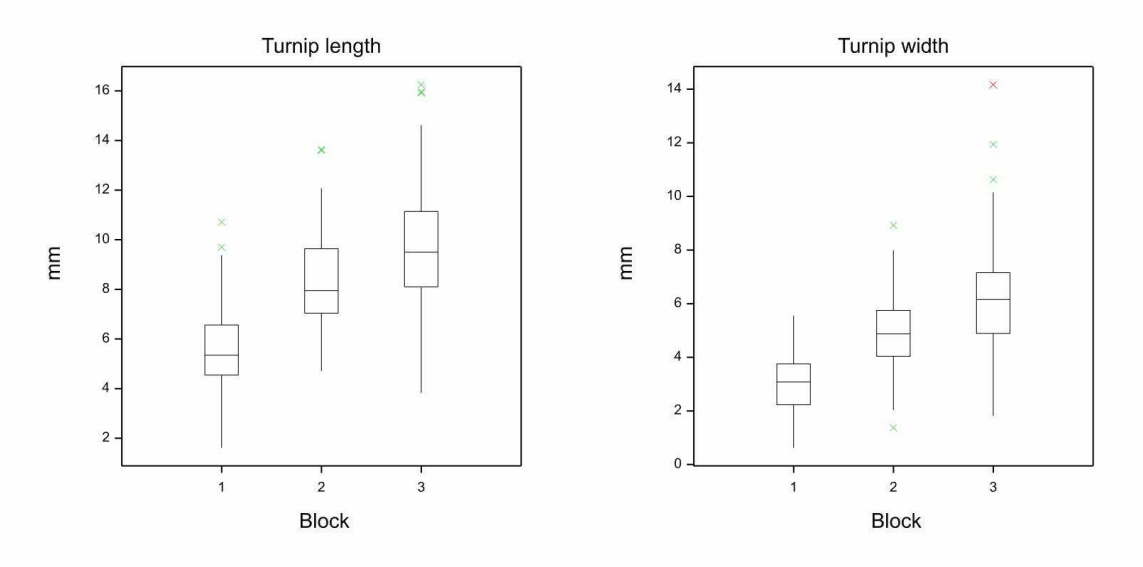

Fig. 12: Boxplot of turnip length per block (left) and turnip width per block (right) measured in the field experiment autumn 2012

3.3 Construction of a genetic map

Out of 248 polymorphic markers, 134 were assembled in 21 linkage groups. The linkage maps created all together span a distance of 1610.2cM. On average markers are spaced 12cM apart, 63% of the mapped markers are distorted. The mapped markers are from the K-Bioscience dataset, the IVF set and the Leafmarkers. (Fig. 3) To evaluate order, distribution and marker distance, maps were created from the 21 linkage groups using MapChart 2.2 (Voorrips 1999-2006). The map files created with Joinmap software were used as input data for that. Three of the 21 Linkage maps contain markers from more than one chromosome, based on the reference genome.
To compare the linkage maps to the physical genetic map of *B. rapa*, the physical positions of all markers were gathered and a physical map was created in MapChart 2.2.This map depicts marker order per chromosome based on the Chiifu reference genome.
The marker position in bp was divided by 100.000 (x 0.1Mbp) to make linkage maps and physical distance better comparable. Linkage maps per chromosome and the physical map were then aligned by drawing lines between complementary markers.

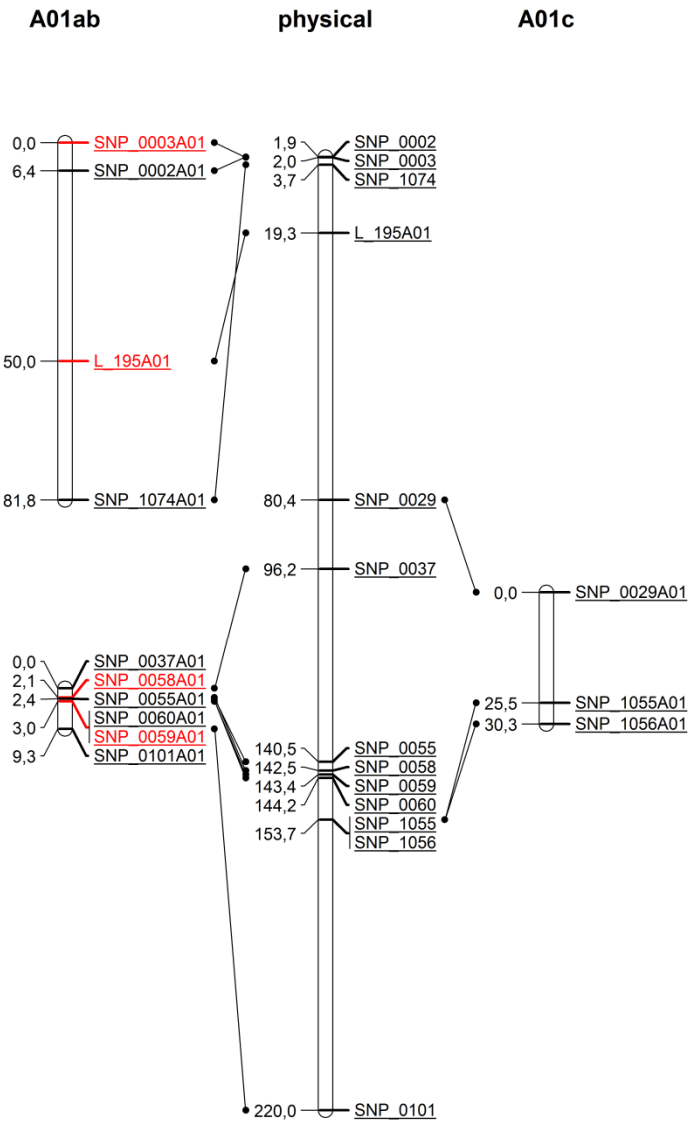

Fig. 13: Linkage maps for chromosome A01 aligned to physical map; red markers show contradicting order to the physical map

Chromosome A01 is represented by three linkage groups comprised of markers from chromosome A01. Linkage groups A01a covers the upper part of chromosome A01, two of four markers show confounded marker order compared to the physical map. The linkage map spans 81.8cM.The second linkage group (A01b) was built from 6 markers, two of which show confounded order. This linkage map is relatively short; it spans 9.3cM. The third linkage group A01c comprises three markers with reasonable marker order, its map spans 30.3cM. The second and third linkage group of chromosome A01 are nested within each other if compared to the physical positions of the same markers. (Fig. 13) To depict marker distortion, allele frequency and goodness of fit regarding the allele denomination a table was created for each mapped marker per chromosome. (Tab. 6)

Tab. 6: Chromosome A01 allele denomination and marker distortion; markers in green denominated in accordance to parental alleles, red inverted denomination

LG	Locus	a	h	b	X2	Signif.
A01a	SNP_0003A01	56	17	40	2.67	-
	SNP_0002A01	70	14	33	13.29	******
	L_195A01	83	11	12	53.06	*******
	SNP_1074A01	20	46	38	5.59	**
A01b	SNP_0037A01	101	6	9	76.95	*******
	SNP_0058A01	99	6	11	70.4	*******
	SNP_0060A01	96	7	12	65.33	*******
	SNP_0059A01	97	6	12	66.28	*******
	SNP_0101A01	88	7	19	44.5	*******
A01c	SNP_0029A01	33	10	64	9.91	****
	SNP_1055A01	52	16	49	0.09	-
	SNP_1056A01	37	18	45	0.78	-

In these tables marker distortion is depicted as significance levels of the chi² test as described in chapter 2.7. P-values of the significance level symbols are explained in table 7. All marker with a p-value below the threshold of 0.05 are considered to be distorted.
Allele frequencies are represented as true allele numbers per marker screened over the whole population. The goodness of fit of the allele denomination based on the quasi parental accession analysis results is depicted as colour code per marker. Green colour indicates that the denomination based on the quasi parental accessions and the denomination in mapping for accurate marker order is in accordance. This means allele a in the progenies denomination refers to the parent VT115 and allele b to the parent PC105.
Red colour indicates that previous allele denomination and denomination for mapping contradict each other. In these cases allele denomination is inverted with regard to the quasi parental accessions. If the marker name is not underlain with colour, denomination based on quasi parental accessions was not possible.

Tab. 7: Significance level and p-values of the chi² test for marker distortion

significance level	p-value
-	>0.1
*	<0.1
**	<0.05
***	<0.005
****	<0.0005
*****	<0.00005
******	<0.000005
*******	<0.0000005

In A01a and A01c distorted and non distorted markers are mapped, A01b is formed by distorted markers only. (Tab. 6) The three linkage groups used for mapping are denominated to form a map based on the physical mapping order and shortest recombination based distance of neighbouring markers. To equalize allele denomination between linkage groups the graphical genotyping table was adjusted. The seven markers fitting the reads of parental accession individuals prove that the adjusted denomination depicts true inheritance patterns. As a result it can be deduced that distorted markers mapped to chromosome A01 are skewed towards parent VT115.

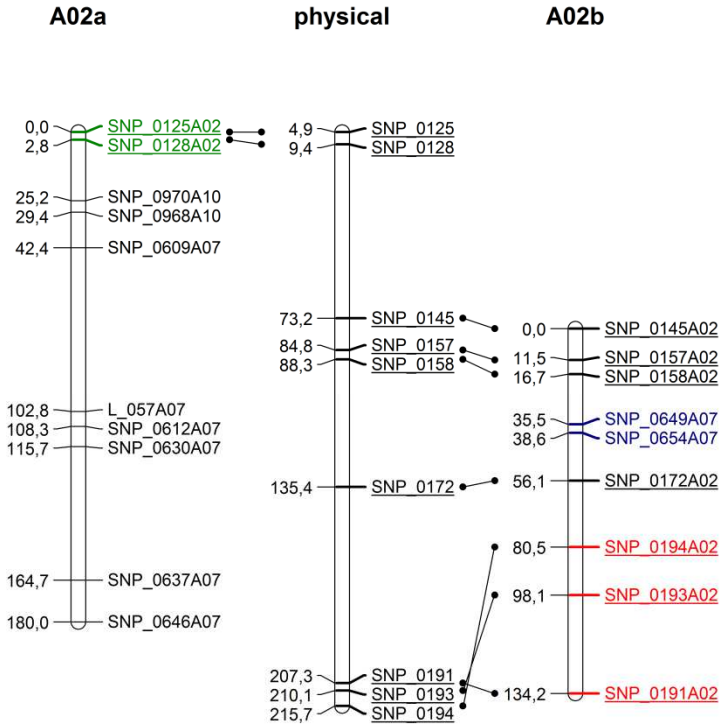

Fig. 14: Linkage maps for chromosome A02 aligned to physical map, red markers show contradicting order to the physical map, blue and green markers are from different chromosomes

Two linkage groups contain markers from chromosome A02. The first linkage group (A02a) depicts two markers from the upper part of chromosome A02 in conserved order. The markers from chromosome A02 are grouped together with markers from chromosome A07 and A10. The second linkage group shows 7 markers from chromosome A02 the last three markers depicting the lower part of chromosome A07 show confounded order. Among the seven markers from chromosome A02 two markers from chromosome A07 were grouped. The second map spans 134.2cM. (Fig. 14)

Tab. 8: Chromosome A02 allele denomination and marker distortion, markers in green denominated in accordance to parental alleles

LG	Locus	a	h	b	X2	Signif.
A02a	SNP_0125A02	41	17	59	3.24	*
	SNP_0128A02	43	18	55	1.47	-
A02b	SNP_0145A02	43	20	52	0.85	-
	SNP_0157A02	52	18	42	1.06	-
	SNP_0158A02	63	16	37	6.76	***
	SNP_0649A07	57	1	54	0.08	-
	SNP_0654A07	61	1	49	1.31	-
	SNP_0172A02	58	13	46	1.38	-
	SNP_0194A02	53	13	49	0.16	-
	SNP_0193A02	68	14	33	12.13	******
	SNP_0191A02	100	0	13	66.98	*******

Both maps A02a and A02b are formed by distorted and non distorted markers. Seven markers mapped to chromosome A02 show allele denomination in accordance to parental allele denomination. As for chromosome A01 it can be deduced that chromosome A02 is skewed towards parent VT115 for the three distorted markers. (Tab. 8)

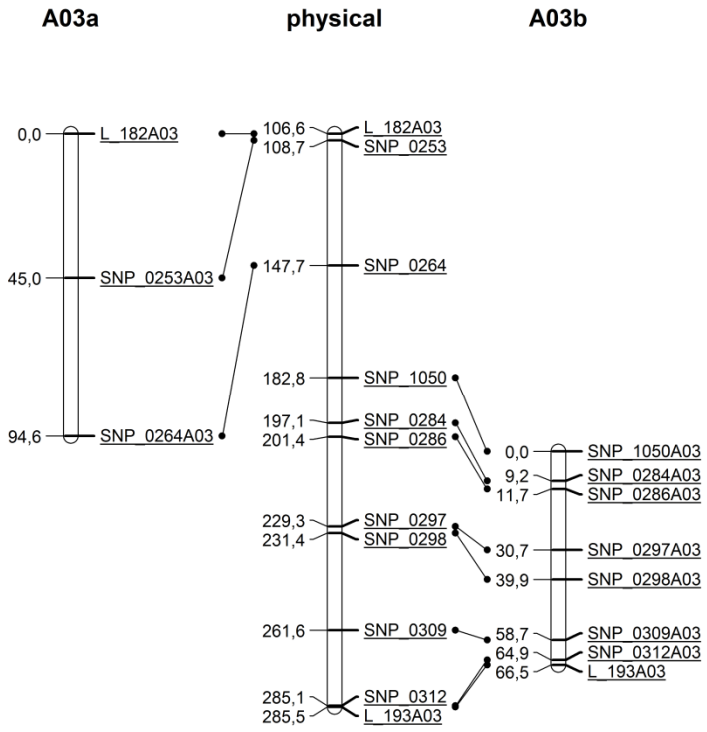

Fig. 15: Linkage maps for chromosome A03 aligned to physical map

Two linkage groups align to chromosome A03, the first is comprised of three markers the second of eight markers. Mapping order in both linkage groups was conserved relative to the physical map. The first map spans 94.6cM the second spans 66.5cM.(Fig. 15)

Tab. 9: Chromosome A03 allele denomination and marker distortion, markers in red show inverted denomination of parental alleles

LG	Locus	a	h	b	X2	Signif.
A03a	L_182A03	6	10	91	74.48	*******
	SNP_0253A03	37	15	65	7.69	***
	SNP_0264A03	31	18	67	13.22	******
A03b	SNP_1050A03	45	15	57	1.41	-
	SNP_0284A03	50	15	52	0.04	-
	SNP_0286A03	54	14	45	0.82	-
	SNP_0297A03	52	19	45	0.51	-
	SNP_0298A03	38	1	72	10.51	****
	SNP_0309A03	45	16	54	0.82	-
	SNP_0312A03	48	14	51	0.09	-
	L_193A03	40	8	46	0.42	-

Map A03a was calculated from distorted markers only, the map of A03b is composed of all but one non distorted markers. Eight markers show inverted denomination of parental alleles. Since all markers are inversely denominated I can simply swap allele a for b in the denomination of the progeny. This way the a and b allele becomes comparable to the parental inheritance again. Then denomination of parental accessions reveals that the four distorted markers on chromosome A03 are skewed towards parent PC105. (Tab. 9)

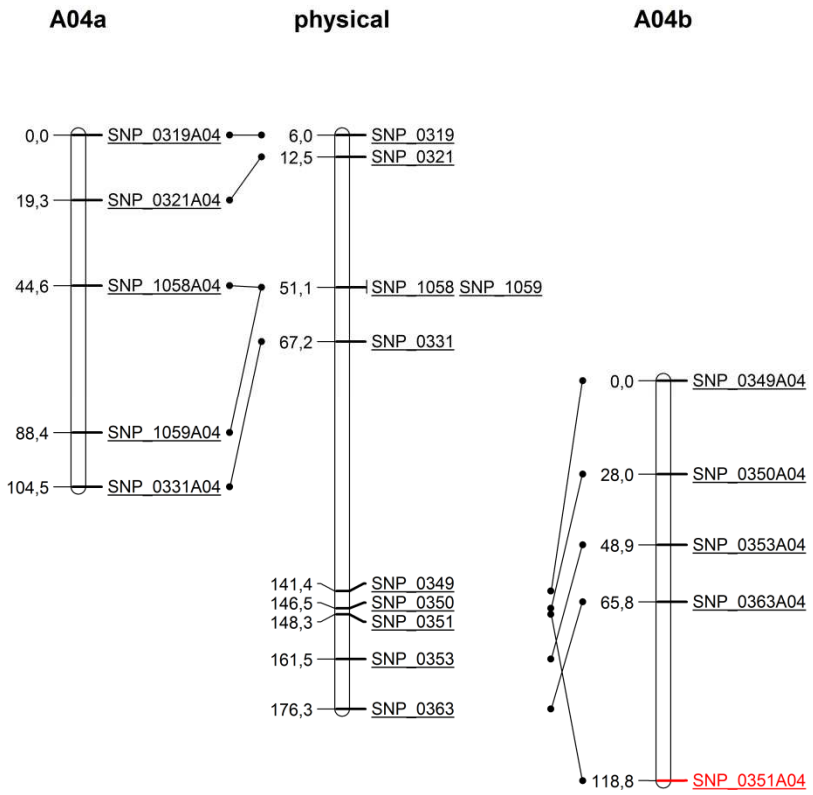

Fig. 16: Linkage maps for chromosome A04 aligned to physical map, red marker shows contradicting order to the physical map

Two linkage groups align to chromosome A04. A04a is composed of five markers. The order was conserved relative to the physical map. The markers SNP_1058 and SNP_1059 display a mapping

distance of 44.2cM although just 9bp apart in the physical map. The second linkage group is comprised of five markers, one shows confounded marker position compared to the physical map. Map A04a spans 104.5cM, map A04b spans 118.8cM. (Fig. 16)

Tab. 10: Chromosome A04 allele denomination and marker distortion, markers in red show inverted denomination of parental alleles

LG	Locus	a	h	b	X2	Signif.
A04a	SNP_0319A04	36	25	56	4.35	**
	SNP_0321A04	53	27	36	3.25	*
	SNP_1058A04	46	16	38	0.76	-
	SNP_1059A04	59	18	23	15.8	*******
	SNP_0331A04	26	11	80	27.51	*******
A04b	SNP_0349A04	78	14	25	27.27	*******
	SNP_0350A04	47	17	52	0.25	-
	SNP_0353A04	44	16	38	0.44	-
	SNP_0363A04	65	11	32	11.23	*****
	SNP_0351A04	93	3	14	58.33	*******

Both maps consist of distorted and non distorted markers. For the majority of markers the alleles can not clearly be assigned to a parent. Tendency to one parent in the distortion can not be deduced. (Tab.10)

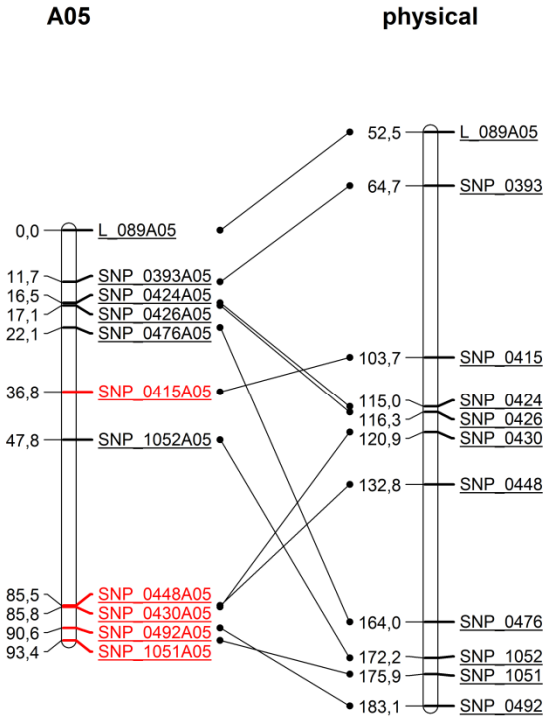

Fig. 17: Linkage map for chromosome A05 aligned to physical map, red markers show contradicting order to the physical map

One linkage group depicts chromosome A05. Five of the 11 mapped markers show confounded positioning compared to the physical map. The map spans 93.4cM. (Fig. 17)

Tab. 11: Chromosome A05 allele denomination and marker distortion

LG	Locus	a	h	b	X2	Signif.
A05a	L_089A05	53	15	40	1.82	-
	SNP_0393A05	49	12	53	0.16	-
	SNP_0424A05	54	3	56	0.04	-
	SNP_0426A05	51	7	56	0.23	-
	SNP_0476A05	43	14	56	1.71	-
	SNP_0415A05	70	12	33	13.29	******
	SNP_1052A05	71	12	7	48.05	*******
	SNP_0448A05	91	4	20	45.41	*******
	SNP_0430A05	93	5	17	52.51	*******
	SNP_0492A05	89	9	16	50.75	*******
	SNP_1051A05	84	13	19	41.02	*******

The map for chromosome A05 contains distorted and non distorted markers. The vast majority of markers can not be characterized by clear denomination of parental alleles. Tendency in distortion towards one parent can not be deduced. (Tab. 11)

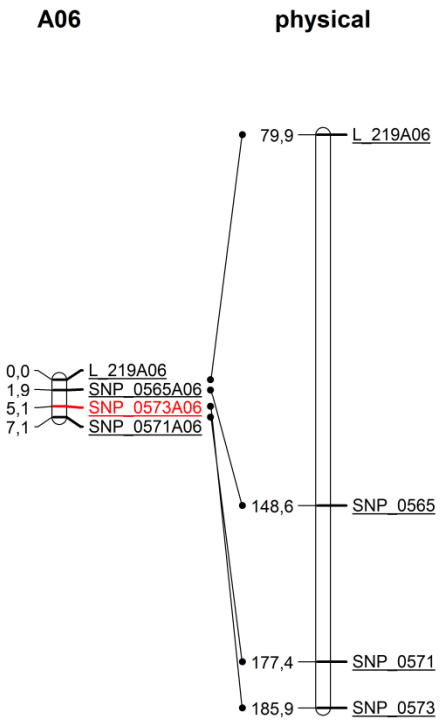

Fig. 18: Linkage map for chromosome A06 aligned to physical map, red markers show contradicting order to the physical map

The one linkage group comprises of four markers from chromosome A06. It forms a map of 7.1cM length. One marker shows a confounded position. (Fig. 18)

Tab. 12: Chromosome A06 allele denomination and marker distortion

LG	Locus	a	h	b	X2	Signif.
A06a	L_219A06	94	8	8	72.51	*******
	SNP_0565A06	97	8	8	75.44	*******
	SNP_0573A06	100	0	11	71.36	*******
	SNP_0571A06	101	3	11	72.32	*******

All markers mapped from chromosome A06 are distorted. Denomination of parental alleles per marker is not possible. (Tab. 12)

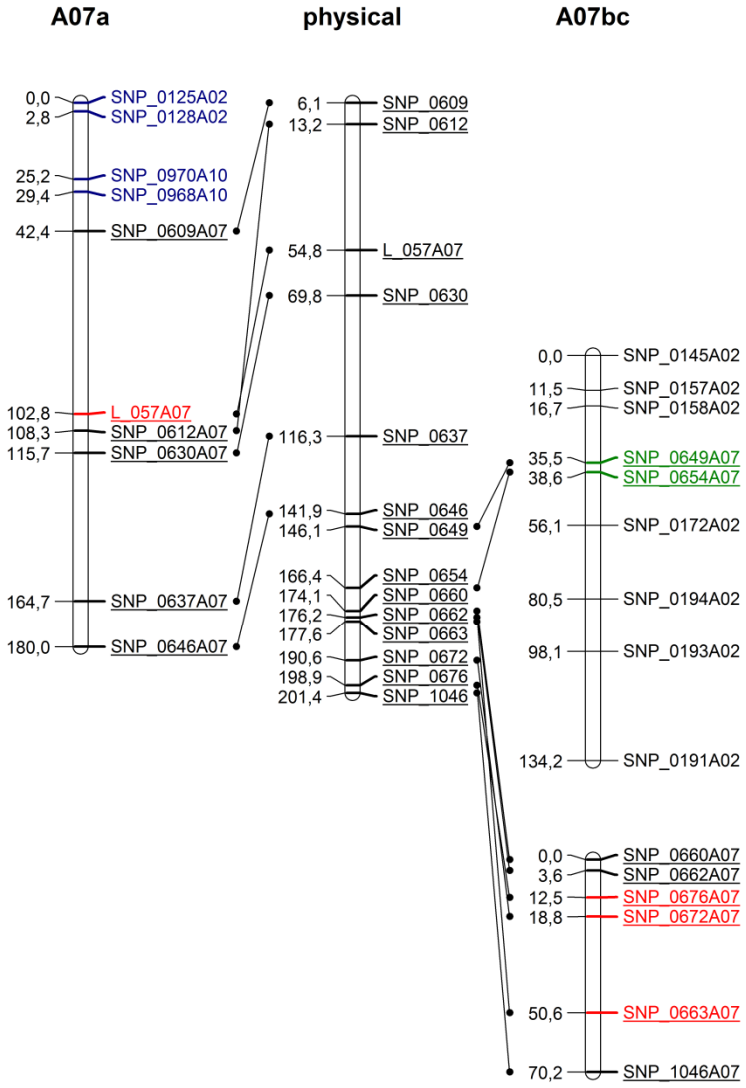

Fig. 19: Linkage maps for chromosome A07 aligned to physical map, red markers show contradicting order to the physical map, blue and green markers from maps containing markers of different chromosomes

Three linkage groups containing markers from chromosome A07 were mapped. A07a contains six markers from chromosome A07. The other four markers are two from chromosome A10 and two from chromosome A02. Except for marker L_057 all markers show conserved order compared to the physical map. This map depicting the upper part of chromosome A07 spans 180cM. Linkage group A07b contains two markers from chromosome A07, the other seven are from chromosome A02. The lower part of chromosome A07 is depicted by A07c six markers aligned to the physical map. Three for the markers show confounded order. (Fig. 19)

Tab. 13: Chromosome A07 allele denomination and marker distortion

LG	Locus	a	h	b	X2	Signif.
A07a	SNP_0125A02	41	17	59	3.24	*
	SNP_0128A02	43	18	55	1.47	-
	SNP_0970A10	46	16	55	0.8	-
	SNP_0968A10	53	13	49	0.16	-
	SNP_0609A07	57	16	41	2.61	-
	L_057A07	91	6	13	58.5	*******
	SNP_0612A07	90	5	19	46.25	*******
	SNP_0630A07	92	3	19	48.01	*******
	SNP_0637A07	53	12	37	2.84	*
	SNP_0646A07	40	19	58	3.31	*
A07b	SNP_0649A07	57	1	54	0.08	-
	SNP_0654A07	61	1	49	1.31	-
A07c	SNP_0660A07	89	3	21	42.04	*******
	SNP_0662A07	89	7	13	56.63	*******
	SNP_0676A07	91	3	19	47.13	*******
	SNP_0672A07	98	2	12	67.24	*******
	SNP_0663A07	71	15	29	17.64	*******
	SNP_1046A07	50	19	48	0.04	-

Map A07a and A07c contain distorted and non distorted markers. Two markers of chromosome A02 and two of chromosome A10 are grouped in the map A07a, all 4 are non distorted and with comparable allele proportions. Three of the markers from chromosome A07 mapped on in A07a are non distorted, the other 3 are strongly distorted. The small map A07b contains two non distorted markers. Denomination of parental alleles per marker is ambiguous. (Tab. 13)

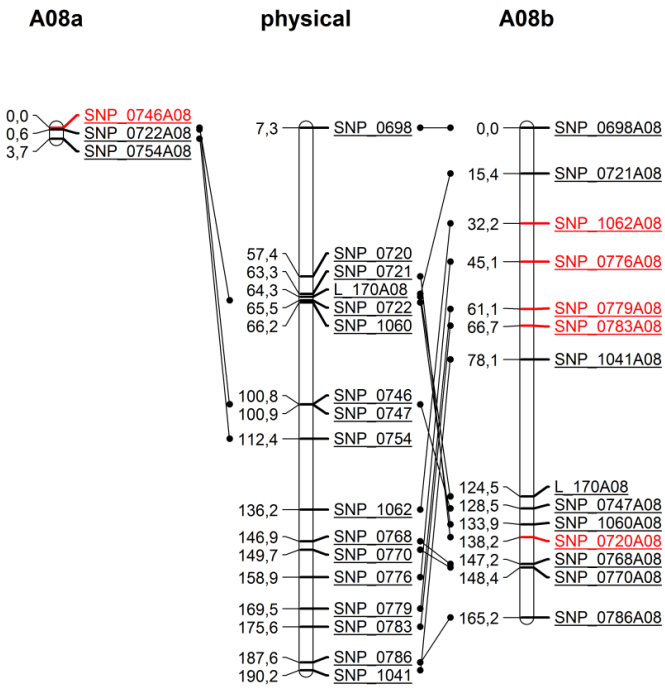

Fig. 20: Linkage maps for chromosome A08 aligned to physical map, red markers show contradicting order to the physical map

Two linkage groups were formed from markers for chromosome A08. A08a comprises of three markers, one shows confounded positioning and the map spans 3.7cM. The linkage map A08b was formed from 14 markers. Five markers show confounded positions compared to the physical map. The linkage group spans 165.2cM. (Fig. 20)

Tab. 14: Chromosome A08 allele denomination and marker distortion

LG	Locus	a	h	b	X2	Signif.
A08a	SNP_0746A08	100	8	8	78.37	*******
	SNP_0722A08	97	8	9	73.06	*******
	SNP_0754A08	109	5	3	100.32	*******
A08b	SNP_0698A08	49	15	53	0.16	-
	SNP_0721A08	52	14	50	0.04	-
	SNP_1062A08	37	18	61	5.88	**
	SNP_0776A08	42	23	51	0.87	-
	SNP_0779A08	58	19	40	3.31	*
	SNP_0783A08	62	18	37	6.31	**
	SNP_1041A08	41	9	66	5.84	**
	L_170A08	93	6	14	58.33	*******
	SNP_0747A08	97	0	17	56.14	*******
	SNP_1060A08	91	0	12	60.59	*******
	SNP_0720A08	96	12	5	81.99	*******
	SNP_0768A08	98	0	16	58.98	*******
	SNP_0770A08	97	0	17	56.14	*******
	SNP_0786A08	89	2	23	38.89	*******

The small linkage map A08a is formed by three distorted markers. The second map contains distorted and non distorted markers. No skewness towards one of the parents for the whole chromosome can be deduced. Denomination of the alleles is not homogenous throughout the chromosome. (Tab. 14)

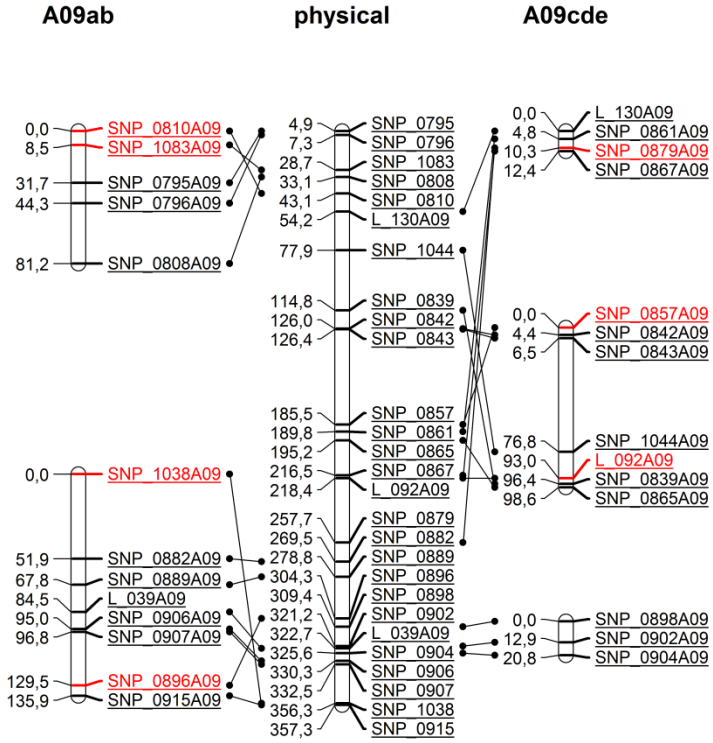

Fig. 21: Linkage maps for chromosome A09 aligned to physical map, red markers show contradicting order to the physical map

Five linkage groups were formed from markers from chromosome A09. The first group (A09a) consists of five markers, the first two show confounded order to the physical mapping position. The map inferred from the linkage group spans 81.2cM. The second group A09b comprises eight markers two with confounded position. The inferred map spans 135.9cM. Group A09c comprises of only four markers one with confounded position, the adjacent map spans 12.4cM. Map A09d was build from seven markers, two with confounded position, the map spans 98.5cM. The last group A09e consists of only three markers with conserved order. The relatively short map spans 20.8cM. Compared to the physical map A09d should be nested within A09c, as well as A09e should be nested in A09b. (Fig. 21)

Tab. 15: Chromosome A09 allele denomination and marker distortion

LG	Locus	a	h	b	X2	Signif.
A09a	SNP_0810A09	52	15	50	0.04	-
	SNP_1083A09	55	14	47	0.63	-
	SNP_0795A09	44	18	53	0.84	-
	SNP_0796A09	76	15	26	24.51	*******
	SNP_0808A09	97	0	15	60.04	*******
A09b	SNP_1038A09	47	13	56	0.79	-
	SNP_0882A09	44	24	49	0.27	-
	SNP_0889A09	66	14	35	9.51	****

	L_039A09	33	12	51	3.86	**
	SNP_0906A09	71	14	32	14.77	******
	SNP_0907A09	74	13	29	19.66	*******
	SNP_0896A09	94	0	18	51.57	*******
	SNP_0915A09	90	9	5	76.05	*******
A09c	L_130A09	95	2	14	60.19	*******
	SNP_0861A09	109	0	5	94.88	*******
	SNP_0879A09	105	6	6	88.3	*******
	SNP_0867A09	87	6	6	77.92	*******
A09d	SNP_0857A09	92	2	19	48.01	*******
	SNP_0842A09	90	1	23	39.73	*******
	SNP_0843A09	91	0	21	43.75	*******
	SNP_1044A09	39	21	54	2.42	-
	L_092A09	54	18	40	2.09	-
	SNP_0839A09	51	20	40	1.33	-
	SNP_0865A09	48	23	46	0.04	-
A09e	SNP_0898A09	69	7	25	20.6	*******
	SNP_0902A09	36	15	65	8.33	****
	SNP_0904A09	64	17	35	8.49	****

Linkage maps A09a, A09b and A09d consist of distorted and non distorted markers. The linkage map A09c and A09e consist of distorted markers only. Denomination of parental alleles is irregular throughout the chromosome. (Tab. 15)

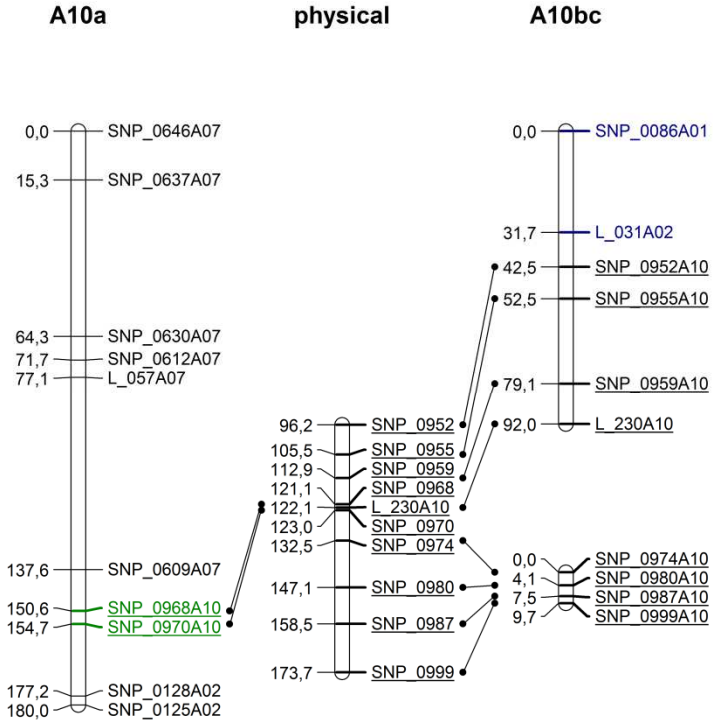

Fig. 22: Linkage map for chromosome A10 aligned to physical map, green markers from maps containing markers of different chromosomes

Three linkage groups were formed from markers of chromosome A10, the group A10a consists of two markers from chromosome A10 and another six were based on chromosome A07 sequence. Two more were designed based on chromosomes A02 sequence. The order of markers in the adjacent map is conserved relatively to the physical mapping position of A10. Map A10a spans 180cM. Map A10b contains two markers out of place, one from chromosome A01 and one from chromosome A02, the other four markers are from chromosome A10 and they align to the physical marker order with conserved order. The map spans 92cM. The third map A10c was formed from four markers of chromosome A10, the map spans 9.7cM. (Fig. 22)

Tab. 16: Chromosome A10 allele denomination and marker distortion, markers in green denominated in accordance to parental alleles

LG	Locus	a	h	b	X2	Signif.
A10a	SNP_0968A10	53	13	49	0.16	-
	SNP_0970A10	46	16	55	0.8	-
A10b	SNP_0086A01	57	7	35	5.26	**
	L_031A02	69	1	28	17.33	*******
	SNP_0952A10	67	3	29	15.04	******
	SNP_0955A10	58	11	30	8.91	****
	SNP_0959A10	50	8	41	0.89	-
	L_230A10	39	10	38	0.01	-
A10c	SNP_0974A10	106	5	4	94.58	*******
	SNP_0980A10	97	9	8	75.44	*******
	SNP_0987A10	101	7	9	76.95	*******
	SNP_0999A10	101	6	8	79.35	*******

The two markers of chromosome A10 in A10a are not distorted. A full description of map A10a is given in the description of chromosome A07 linkage maps. The two markers of chromosome A10 in A10a map next to two markers of chromosome A02, all four are not distorted and allele proportions are comparable. In the map A10b two markers from other chromosomes were mapped among the ones of chromosome A10. Three of the markers in A10b are distorted and three are non distorted. The four markers in A10c are all distorted. Allele denomination of all markers is in accordance to the denomination of parental alleles, the distorted markers are exclusively skewed towards the parent VT115. (Tab. 16)

3.4 Association studies

To find markers associated with the measured traits single trait ANOVAs were conducted as described in chapter 2.9. The data for all traits was $\log_{10}$ transformed the average trait values were used in the ANOVAs per trait per experiment, instead of conducting ANOVAs for all blocks separately.

3.4.1 Greenhouse experiment

Three markers were found to be significantly associated with turnip weight in the greenhouse experiment. (Fig. 23) The markers SNP_1056 and SNP_1074 are both located on chromosome A01 and were also mapped on this chromosome but on different sub maps of chromosome A01. Both markers segregate in Mendelian fashion.
The marker SNP_0898 is a significantly distorted marker, based on the physical reference map it is located on chromosome A09 and was also mapped accordingly. (Tab. 17)

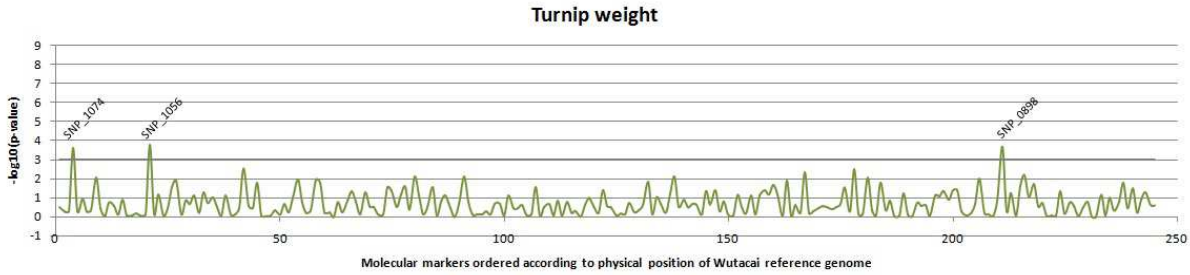

Fig. 23: -$\log_{10}$(p-values) association study for turnip weight in greenhouse experiment

Tab. 17: Significant QTL markers of turnip weight in greenhouse experiment - marker position, test for marker distortion and allele distribution

Marker	Position in bp	-$\log_{10}$ (p-values)	a	h	b	X2	Signif.
weight							
SNP_1074A01	374356	3.62	20	46	38	5.59	**
SNP_1056A01	15365196	3.77	37	18	45	0.78	-
SNP_0898A09	30938434	3.65	68	5	22	23.51	*******
length							
SNP_0898A09	30938434	3.21	68	5	22	23.51	*******

One marker was found to be significantly associated with turnip length in the greenhouse experiment. It is the marker SNP_0898 again which was also identified to be significantly associated to turnip weight. (Fig. 24)

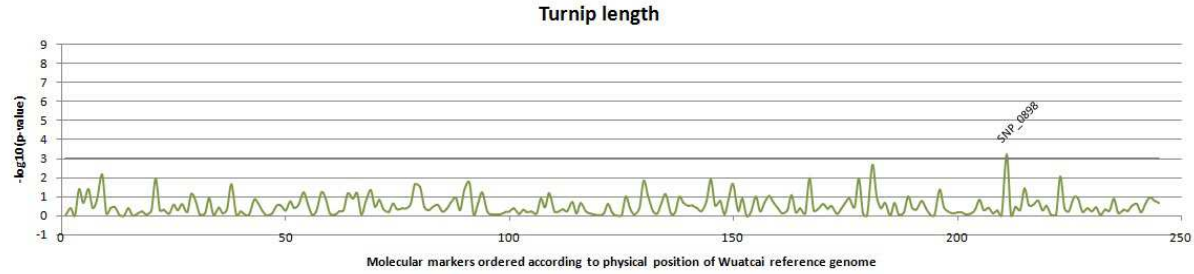

Fig. 24:-log$_{10}$(p-values) association study for turnip length in greenhouse experiment

No marker was found to be significantly associated with turnip width in the greenhouse experiment.

3.4.2 Field experiment

Eighteen markers were found to be significantly associated with turnip width in the field experiment. (Fig. 25) All markers associated with turnip width are strongly distorted, seven of the markers were included in the linkage maps, markers are clustering on chromosome A07 and A08. (Tab. 18)

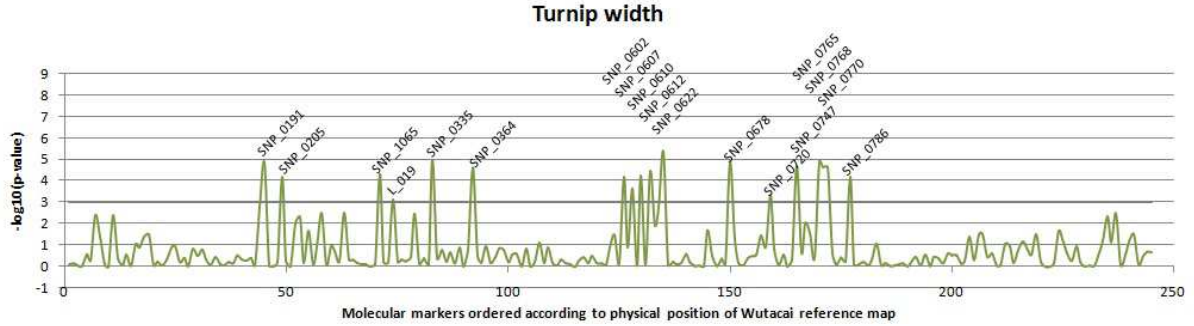

Fig. 25:-log$_{10}$ (p-values) association study for turnip width in field experiment

Tab. 18: Significant QTL markers of turnip width in field experiment - marker position, test for marker distortion and allele distribution, markers in blue included in linkage maps

Marker	Position in bp	-log$_{10}$ (p-values)	a	h	b	X2	Signif.
SNP_0191A02	20728211	4.85	100	0	13	66.98	*******
SNP_0205A02	23890520	4.16	88	14	12	57.76	*******
SNP_1065A03	27864125	4.32	106	7	2	100.15	*******
SNP_0335A04	7089683	4.95	98	0	16	58.98	*******
SNP_0364A04	17743027	4.60	103	0	8	81.31	*******
SNP_0602A06	24995754	4.20	104	1	1	101.04	*******
SNP_0607A07	286182	3.64	113	0	2	107.14	*******
SNP_0610A07	797437	4.22	113	1	1	110.04	*******
SNP_0612A07	1316612	4.44	90	5	19	46.25	*******
SNP_0622A07	5385095	5.36	110	0	1	107.04	*******
SNP_0678A07	20498940	4.86	108	1	1	105.04	*******
SNP_0720A08	5738432	3.35	96	12	5	81.99	*******
SNP_0747A08	10090362	4.69	97	0	17	56.14	*******
SNP_0765A08	14098893	4.88	96	5	13	63.2	*******
SNP_0768A08	14692670	4.61	98	0	16	58.98	*******
SNP_0770A08	14966189	4.63	97	0	17	56.14	*******
SNP_0786A08	18758475	4.16	89	2	23	38.89	*******

One marker was found to be associated with turnip length in the greenhouse experiment. (Fig. 26) The marker is located on chromosome A01 according to the physical reference map. The marker was not mapped in the linkage map, it is segregating in Mendelian fashion. (Tab. 19)

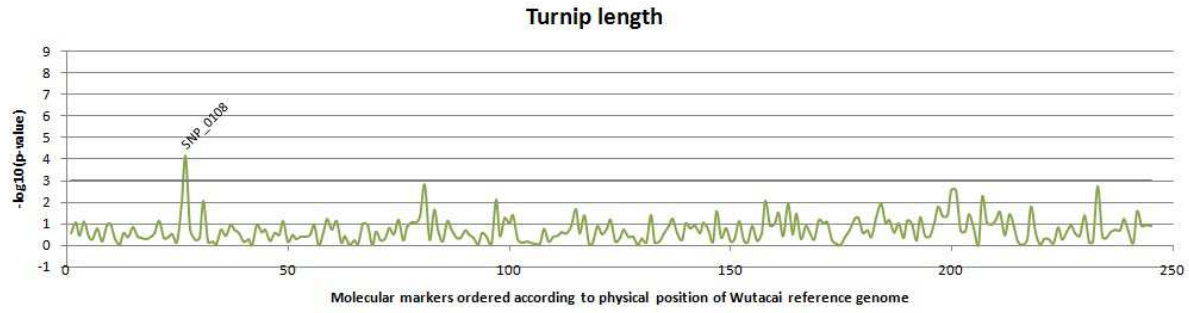

Fig. 26: -log$_{10}$ (p-values) association study for turnip length in field experiment

Tab. 19: Significant QTL markers of turnip length in field experiment - marker position, test for marker distortion and allele distribution, markers in blue included in linkage maps

Marker	Position in bp	-log$_{10}$ (p-values)	a	h	b	X2	Signif.
SNP_0108A01	25004591	4.14	44	16	56	1.44	-

4 Discussion

Strong marker distortion is a distinct feature of the C29 population in F4 generation. The strong distortion is present in the dataset from K-Bioscience as well as in the markers screened with Lightscanner® technology. This indicates that the distortion is not due to false data procession when setting up excel tables. The various causes and effects of the distortion are the main focus in the discussion of marker screening and mapping, followed by discussing the marker association studies.

4.1 Marker screening

Of 116 markers from the IVF set approximately 40% were polymorphic and about 25% of the total were mapped in the end. The overall yield of polymorphic markers in the IVF set is relatively low. It is likely that some potentially polymorphic markers were lost due to marker distortion. Extremely distorted markers like SNP_0915A09 (Tab. 15) show an allele ratio of 15:1. The prescreening set in my experiment was limited to 10 individuals. With a wide allele ration like this it becomes more unlikely to find polymorphisms in the first place if the prescreening set is small. One could suspect that markers as extremely distorted as that are not useful, because they may not contribute to a linkage map.In the end however the afore mentioned marker was mapped in linkage group A09b, in a context of more distorted markers every marker showing some polymorphisms may still contribute to the mapping.
In total 98 leafmarkers were screened. Approximately 30% of the markers have shown polymorphisms, but only 11 markers (10%) were mapped. The primers for the leafmarkers were not designed purely based on SNPs but on various genome differences in genes associated with leaf development analysed in *A. thaliana*. Many of the markers have shown Lightscanner® outputs of category 3 and 5 (Fig. 6 and 8) in the prescreening. In the main screening some more have shown outputs of category 4 (Fig. 7) which was not predictable based on the prescreening results. A likely cause of the strange marker screening results is unspecific binding of the primers. The primers may bind to several paralogs of the leaf development genes in *B. rapa*.
Based on this it is not recommendable to further analyse markers that show some odd segregation patterns in the prescreening.

4.2 Cause and effects of marker segregation distortion in literature

Taylor & Ingvarsson (2003) analysed marker distortion in *Silene latifolia*. They observed a common sex ration bias in some families in *S. latifolia*. They come to the conclusion that this bias is caused by a meiotic drive effecting pollen fertility. The effect is especially obvious in crosses of different families. Individuals carrying a deleterious allele produce pollen with reduced fertility but on the other hand when receiving pollen these individuals are not inhibited in seed production. They also sire a small number of offspring showing reduced pollen fertility. This suggests that deleterious alleles causing segregation distortion can be maintained in a population in a heterozygous form. Taylor and Ingvarsson state that forms of sex chromosome meiotic drive as the case of reduced pollen fertility are rather common in crop plant species.

Törjék et al. (2006) analysed segregation distortion in RILs of *A. thaliana* crosses (C24 x Col-0). They report that segregation distortion has been found in several other plant species such as maize, poplar, rice and mung bean. They claim that molecular mechanisms causing segregation distortion are not jet fully understood, but the causes can be various environmental, physiological and genetic factors. Segregation distortion can affect male and female germ lines. Negative selection of some individuals of a population can also occur after formation of the zygote. Observable indicators of segregation distortion can be pollen-pistil incompatibilities, gametic competition, negative epistatic interaction or gamete abortion. They found both environmental and genetic effects involved in reduced fertility of RILs propagated through single seed descent. They were able to pinpoint the genetic factors to epistatic interaction of genetic regions on chromosome IV and V. They found strong segregation distortion of markers on these two chromosomes and additionally on chromosome III. Pair wise interaction of the markers on chromosome IV and V revealed the epistatic effect. They identified markers from Col-0 on chromosome four and markers from C24 on the bottom of chromosome five and hypothesized that alleles in homozygous form from both parental lines would lead to the reduced fertility.
In a testcross approximately 23% of the offspring have shown reduced fertility, thus the evidence for the epistatic effect was hardened. The mentioned reduction in fertility was observed as significantly reduced number of pollen formed in the flowers or no pollen formed at all. The described system shows high similarity to common male sterility systems found in many species of crop plants which are widely implemented in breeding today. In the end Törjék et al.(2006) claim that segregation has commonly been found in intra- and interspecific crosses in the family *Brassicaceae*, the literature resources they cite in this context are all focused in crosses within A. thaliana and crosses between *A. thaliana* and *A. lyrata*.

J. Wang et al. (2011) analysed segregation distortion in DH lines of *Brassica napus* for map integration of *B. napus*, *B. rapa* and *A. thaliana*. They found 22% to 49% of the markers per DH line to be significantly distorted. They were able to depict a parental genotype that was consistently favoured for its alleles in the progeny per linkage group. Several linkage groups have shown strong segregation distortion for up to 95% of the markers of an entire linkage group. In the end Wang et al. were able to create maps for *B. napus* with good analogy to the physical map. They used Joinmap 4.0 and MergeMap software for creation and integration of linkage maps.

He et al. (2001) crossed two rice cultivars, one an *indica* type and one a *japonica* type to create RIL populations and DH populations later to be used to create linkage maps. The marker analysis of the RIL population was conducted on F9 plants.
At a 0.05 threshold of the chi² test for segregation distortion they found 48.7% of the markers to be distorted in the RIL population, whereas only 18.3% of the markers showed segregation distortion in the DH population derived from the same cross. In the RIL population 90% of the distorted markers favoured alleles from the *indica* parent. In the DH population no clear preference of the alleles of one parent or the other was found, 51% of the markers tended to the *indica* parent and 49% tended towards the *japonica* parent. In the RIL population He et al. (2001) were able to pinpoint the allele preference to different chromosomes so markers favouring *indica* were not found mixed with markers favouring *japonica* on the same chromosome. They furthermore found out that regions showing strong marker distortion were in close proximity of gametophytic gene loci or sterility loci. The comparison of the maps derived from DH and RIL populations revealed that the marker order

was identical but marker distance deviated. Difficulties in creation of a linkage map caused by the high level of segregation distortion were not reported.

Shanmugavadivel et al. (2013) also crossed two rice cultivars, an aromatic Basmati cultivar and a normal *indica* cultivar, with the intent to map QTLs for grain size. A RIL population in F7 was analysed and 54% of the markers were found to show significant segregation distortion. Interestingly they claim that a small number of segregation distortion causing loci may lead to a large number of other loci deviating from Mendelian segregation patterns. As additionally causes of SD they mention unconscious selection practices of the breeder while selecting plants for future generations or sampling of material. They indicate that SD may be especially common in wide crosses of plants because of differential competition of male gametes or post zygotic selection. According to their research SD in RIL populations is an accumulative effect of genetic and environmental factors. The SD increases with the progression of selfing over generations.

Zhu et al. (2007) developed a method to reconstruct linkage maps including strongly distorted markers. They found that map distance of distorted markers can be larger or smaller than the true distance. Mapping of two linked segregation distortion loci (SDL) causes overestimation of the mapping distance in most cases. Underestimation of mapping distances was also observed. When several SDL were linked, effects accumulate in unpredicted ways, leading to imprecise map construction. They reconstructed a linkage map from rice they found distortion in 69.5% of all markers. Boot strap sampling on markers of chromosome 12 was conducted to evaluate the new corrected marker distances. They were able to get a smaller standard deviation and confidence interval with the improved mapping which indicates the practicality of the new method.

Literature describing SD in *B. rapa* populations was not found. The issue of SD in creation of linkage maps in many plant species seems to be a topic that is rather ignored. In many cases markers displaying SD are excluded from linkage maps (Zhu et al., 2007). This is not applicable if the majority of markers analysed in a population show distortion.

4.3 Causes of the strong segregation distortion in the C29 population

A number of properties of the C29 population points towards a genetic cause of SD. Over the course of repeated selfing 43% of the RILs were lost because they failed to set seeds or plants died before they reached maturity. Dr. Ningwen Zhang reported that some plants were growing weakly upon germination with discolouring of leaves and stunted growth, eventually they died. (personal communication with Dr. Ningwen Zhang) Plants were self pollinated, under Johan Buchers supervision, some displayed vivipary and therefore failed to produce seeds, some plants produced empty seed pods. These pods may contain tiny but infertile seeds.
This may also be caused by unfavourable external conditions during pollination and lack of experience. Specific investigations on pollen quality were not conducted. The pollinations were done on plants from generations later than F4, if alleles that reduce pollen fertility were present in the population it is likely that these alleles are selected against and they won't show anymore.

Environmental effects on the populations after planting and during seed production are reduced to a minimum as the plants were propagated in the same greenhouse compartment. Temperature, watering, light and nutrients were tightly controlled to provide optimal conditions for plant growth and these conditions were equal to all plants. Therefore uneven growth conditions are unlikely a

reason for the loss of plants in the breeding program. Unconscious selection by the breeder is not a cause of SD in C29 no active selection was done, all lines that could be grown on the F4 were screened with the molecular markers.

Failed seed production indicated presence of SDL linked to production of gametophytes or sterility, or lethal embryo's. The empty seed pods may indicate post zygotic selection. The tiny seeds can result from embryo or seed abortion. The reduced vigour and stunting of lines indicates the presence of unfavourable allele combinations affecting various functions of plant health, growth or development. These effects can be the results of epistatic interactions with several alleles acting at the same time.
It is likely that a mixture of the presented possibilities causes the strong SD in C29. SDL don't accumulate per chromosome and to pinpoint genomic regions of increased SD the resolution of the linkage map is not sufficient. Overall the distribution of SDL appears to be scattered throughout all chromosomes with distorted and non distorted loci on all chromosomes except A06 were all loci are distorted. (Appendix D)
One of the main questions is when the selection process causing SD started. Since the parents are landraces propagated by cross pollination SDL may be retained within the parental lines.
The fact that these landraces suffer from inbreeding depression points out that this is likely the case. Another possibility is that SDL are caused by epistatic effects of both parents within the progeny. Possibly both cases are true, this may explain the severity of SD in C29.

The first strong signs of distorted segregation will show in F2 generation when alleles become homozygous after the first selfing. If for instance one recessive sterility locus is causing SD we would expect 25% of the plants to either fail to produce fertile seeds or pollen. If two recessive loci are acting epistatic that would be the case in 6.25% of the offspring. If many loci with indifferent interaction are acting at the same time the segregation patterns will become unclear, this way it will be unlikely that the number of loci involved can be deduced from the proportion of plants showing abnormal plant growth or ability to reproduce in F2.
A marker screening of F2 plants would be advisable to find regions containing SDL in F4 more recombination events may have diffused the picture in graphical genotyping tables making it more complicated to allocate possible SDL.

4.4 The impact of segregation distortion on grouping and map construction

In the ideal case, markers located on one chromosome compared to the reference genome, would group together depicting the whole chromosome. In case of the presented research up to five linkage groups (A09) were formed from markers of one chromosome, resulting in 5 linkage maps aligning to the physical map of the chromosome. The reason is to be seen in the high marker distortion and overall unequal segregation among the markers located on the chromosome. The chi^2 test for SD is rather stringent. Markers equally distorted according to the test can have big differences in their proportions of alleles.
The grouping is applied by a statistical test for independence of markers. Markers displaying significant analogy of proportions in their alleles are considered to be dependent, they will be grouped together. The independence LOD score grouping is the only applicable grouping method provided in Joinmap if markers are distorted. Distortion, as long as allele proportions don't deviate too much, does not impair the result of this grouping method. However in C29 population markers

from the same chromosome can deviate a lot in their allele proportions, which is the reason for the multiple linkage groups per chromosome.
To connect linkage groups more markers were designed, these markers were preferably located between linkage groups (physical map). It was hypothesized that markers between linkage groups could show transitional allele portions. In a new grouping procedure markers from previously separated LGs and the newly designed markers would then be grouped together. In practice marker could be added to present linkage groups, some formed new linkage groups, but linkage groups never connected.
It is likely that it is not possible to construct unambiguous grouping displaying all polymorphic markers per chromosome in one group with the presented methods. The attempt was made because I observed grouping of distorted and non distorted markers belonging to the same chromosome in one linkage group. Under the assumption of clearly defined regions of distortion and normal segregation on each chromosome it was assumed that additional markers, representing transitions from distorted to non distorted regions, on each chromosome would allow to infer clear and complete maps for all chromosomes. In most cases where distorted and non distorted markers were grouped together later mapping reveals that the order of distorted markers and non distorted marker is not random. Distorted markers will map together and non distorted markers map in another region of the map. A good example is linkage map A08b (Tab. 14), where non distorted markers map on the upper part of the map, distorted markers on the lower part. In most cases where distorted and non distorted markers are mapped on one linkage group, mapping distance between distorted and non distorted regions are extremely large sometimes exceeding 50cM. A mapping distance exceeding 50cM indicates that markers are likely not linked however with the described settings Joinmap infers the mentioned distance with the maximum likelyhood method. The result was accepted in cases where markers belong to the same chromosome based on the reference genome.

Overall the marker order differs from the order in the physical map of the reference genome. In other cases entire maps are consisting of only distorted or non distorted markers. The linkage map A10c (Tab. 16) consists entirely of extremely distorted markers, the linkage map A10a (Tab. 16) on the other hand consists of normally segregating markers. It is unlikely that increasing marker density can connect these two maps.

The maps created from the linkage groups show divergent marker order in comparison to the reference genome. This is the case in the maps aligning to chromosome A01, A02,A04, A05, A06, A07, A08 and A09; only groups aligning to chromosome A03 and A09 show conserved marker order in comparison to the reference genome.
A likely reason for the confounded marker order lies in the accumulated effects described by Zhu et al. (2007) Because of the overall abundance of markers with ambiguous mapping it is unlikely that true genome rearrangements break conserved marker order, in a cross of two cultivars from the same species. There is no literature indicating that any genome rearrangements have been observed in *B. rapa.* Re-sequencing of a number of *B. rapa* cultivars and aligning the sequence never showed genome rearrangements among cultivars. (personal communication Dr. Guusje Bonnema and Prof. Wang Xiaowu)
A succession of inversions within chromosomes would be needed to create a genotype as depicted by the linkage maps. If markers are analysed pair wise which combination causes overestimation of marker distance and which cause underestimation is an interesting question. (Zhu et al., 2007) Very

likely marker distance between strongly distorted markers of related allele proportions would be underestimated since a lot of recombination will go unnoticed. Similar alleles recombine and can't be tracked in every population but in a pair of markers of 1:1 allele proportion much more recombination will be seen.
If a normal segregating marker and a highly distorted marker are compared a lot of recombination events are inferred because the mapping program expects the unequal allele proportions to be result of recombination although it is actually selection that causes gradual decline of the allele frequency of one allele. The same will be the case if two distorted markers drifting to opposite directions are mapped. This is a theory which could be proven if the same markers would be analysed in another population of *B. rapa* without marker distortion, ideally the two populations should be closely related they could share one parent for instance.
The maps created all together span 1610.2cM maps created by other scientists published in literature rarely exceed a distance of 1400cM. The map I created is on the upper end of this scale which can indicate that marker distance was overestimated in most cases. This coincides with the conclusions of Zhu et al. (2007).
Another noticeable strange feature of the constructed linkage maps is the mapping of markers from A02 and A10 among markers of A07. Again genome rearrangements in the form of chromosome translocations are very unlikely the cause. These rearrangements are even more unlikely happening within a species than the inversions mentioned before. However Dr. Ningwen Zhang observed comparable grouping of markers from chromosome A07 and A02 in maps she created from RILs of a cross between rapid cycling (RC144) and the Japanese turnip (VT115) which was also a parent in the C29 population. (personal communication with Dr. Ningwen Zhang) If a genome rearrangement is present it is likely inherited from VT115.
Interestingly the four markers from chromosome A02 and A10 are all non distorted but map among distorted markers from chromosome A07. The grouping is very consistent, even raising the independence LOD threshold to five doesn't impair the grouping, raising the overall threshold is not sensible since other linkage groups depicting true linkage fall apart.
Another possible reason for the confounded grouping is possible unspecific binding of primers causing amplification of genomic regions other than the ones primer design was done for. Genomic regions of high similarity such as the paralogous regions from *A. thaliana* forming synthenic blocks are likely candidates for false amplification of genomic regions from two chromosomes. And indeed synthenic blocks (gene block E from A. thaliana) are present on chromosome A02 and A07 but not on A07 and A10. (Schranz, Lysak, & Mitchell-Olds, 2006) The mentioned markers are all from the K-Bioscience dataset, Lightscanner® data can't be consulted to identify possible unspecific binding. Blasting primer sequence against the reference genome was not possible because the data wasn't provided.
In linkage map A10b again two markers which should belong to other chromosomes were mapped. SNP_0086A01 and L_031A02 were most likely grouped together with markers from A10 because of high similarity of the allele proportions. The markers in the mentioned linkage group are distorted and non distorted ones. Could miss-priming be the cause in this case? At least for the leaf marker primer sequence could be blasted against the reference genome sequence. (Tab. 20) The reverse primer indeed shows sequence similarity to sequence of chromosome A10, however the scores reveal that only 18 of 20 bases align as a perfect fit. This can't be expected to result in strong amplification of chromosome A10 regions using this primer combination, since the forward primer doesn't align to chromosome A10 sequence. Both primers show a perfect fit to chromosome A02 sequence (score 40).

Tab. 20: Primer Blast results

Sequence aligning to	score (bits)	E-value
Forward primer		
A02 [12.17-2010]	40	0.001
A07 [12.17-2010]	34	0.072
Reverse primer		
A02 [12.17-2010]	40	0.001
A10 [12.17-2010]	36	0.018

Over all the maps created don't depict unimpaired information on marker distance and order. These maps are not useful for QTL mapping methods like interval mapping since these methods rely on exact recombination frequencies and marker order. Instead marker trait association was used to identify QTLs for turnip formation traits. This method performs independent marker and trait association tests, interaction of QTLs found with this method are difficult to infer.

4.5 The impact of unclear allele origin on mapping and association studies

Parental accessions are partially heterozygous and DNA of the true parents was never isolated and therefore couldn't be analysed. This represents one of the most difficult parts of the whole thesis research. If a conclusive connection between the alleles of all or most markers to the parents could have been established, a clear and undeniable true graphical genotyping table could have been constructed. This would result in more reliable information on marker distance and order of all linkage groups but of course it doesn't eliminate the issues connected to SD. Nevertheless this bottom to top inference of the genetics of the population is more reliable than the top to bottom approach that was applied. In the approach the most likely parental allele origin was inferred based on reduced recombination frequency between neighbouring markers and analysis of quasi parental accessions. The adjustments on the graphical genotype were done in an estimation by eye fashion they are prone to mistake.
Consistent connection between allele and parent is important to identify the parent that predominantly inherits its allele per marker. This can reveal which parent is actually the carrier of SDL leading to reduced fertility or plant vigour. If QTLs can be identified for turnip formation traits it is important to know which parent contributes the alleles, leading to the phenotypes that were analysed. If information on QTLs will be applied in commercial breeding it is crucial to know the effects of alleles on the phenotype of future generations.

In the tables containing information on allele denominations, allele distribution and marker distortion for each chromosome (Tab. 6-16), I attempt to draw conclusions on which parental alleles are preferentially inherited in the distorted markers. As can be seen in the tables alleles of many markers could not be assigned to the parents because information was not sufficient. Markers marked green show the following denomination, VT115 allele is denominated a and PC105 b. In markers marked red the opposite case is present. If we look selectively at these markers that were coloured and show significant distortion we can infer towards which parent the allele proportions are skewed. (Tab. 20) Overall distorted markers from chromosome A01, A02, A03, A06, A07 and A10 show a higher proportion of allele from parent VT115. Chromosome A08 shows a tendency towards parent PC105 in its allele proportions. The allele proportions of chromosome A09 are pointing towards both parents. In many linkage maps the tendency is ambiguous markers pointing to both

parents are present. Overall the F4 progeny of the C29 populations shows a higher proportion of alleles inherited from parent VT115, the Japanese turnip.

Tab. 21: Tendency in allele distribution among distorted markers of linkage maps

Map	skewed towards
A01a	VT115
A01b	VT115
A02b	VT115
A03a	VT115
A03b	VT115
A04a	ambiguous
A05a	ambiguous
A06a	VT115
A07a	VT115
A07c	ambiguous
A08a	PC105
A08b	PC105
A09a	PC105
A09b	ambiguous
A09c	PC105
A09d	VT115
A09e	ambiguous
A10c	VT115

4.5.1 Quality of the phenotypic data

Greenhouse experiment:
By observing the box plots of all traits the difference between block 2 and 3 appears to be very small. The means line in the boxplots indicates that there may be no significant difference for all 3 traits measured on the turnips. Another ANOVA was run using only the data of blocks 2 and 3 and the presumption was confirmed, indeed there is no significant difference for all traits between blocks 2 and 3. (Appendix G) This means overall the growing conditions in block 1 were the best, leading to significantly wider and longer turnips which are over all heavier than the turnips in block 2 and 3. The main effect leading to the difference between blocks is may be light intensity. Possibly block 2 and 3 received less light.

Field experiment:
Based on the observation of the boxplots for turnips length and width of the field experiment, all 3 blocks show significant difference in the turnip traits. There are no additional peculiarities in the dataset.

General remarks:
For both datasets only one datapoint per trait per block was available this is a very unfortunate situation. This prevents an analysis of block x genotype interactions in this kind of analysis means values per line per block are compared pair wise. If means values of the lines per block react differentially by means they don’t increase or decrease in the same way per single block, than

conditions within the block impair the benefit of the blocking structure. This is valuable information giving indication which data can safely be used in following analysis such as the marker trait association. It may be beneficial to exclude data from a certain block if there is a genotype x block interaction present. The one datapoint per block problem also prohibits a correlation analysis of the data from greenhouse and field trial. For analysis of sufficient statistical power an average per line build from at least 5 datapoints per trait is necessary. Only 3 datapoints were available, one per block. For the marker trait association study the fact that only one datapoint is available per block is no big problem, since the one datapoint already represents an average over 5 plants per line.
The other problematic condition in the dataset is that only for the greenhouse experiment data of turnip weight was available.
To get a higher certainty about present QTLs for turnip related traits comparison, of QTLs found for all three traits in both the greenhouse experiment and the field experiment, would have been worthwhile. The lack of data prevents this for turnip weight.
All issues of the datasets are very unfortunate because all the data was gathered in the first place but not accessible anymore.

4.5.2 Marker trait association study

All conditions on the dataset were met to conduct marker trait association studies by ANOVA of traits by marker. In all cases the phenotypic data was $\log_{10}$ transformed to fit a normal distribution. Another $-\log_{10}$ transformation of the p-values and setting the confidence level at 3 greatly controls the number of false positive associations.
Physical characteristics of the turnips of the C29 population and the segregation of these characteristics lead to the assumption, that alleles associated with turnip formation will be found among the normal or near normal segregating markers. (personal communication with Dr. Guusje Bonnema)
A number of markers were found to be associated with turnip formation traits. (Chapter 3.4) Markers found to be significantly associated to a trait are supposed to be linked to a gene or several genes involved in the expression of the associated trait. To find possible candidate genes, gene annotations in the proximity of the markers were checked in an interval 5000bp up and downstream of the markers associated to the turnip formation traits. The interval was arbitrarily chosen, genes under lying the QTLs could be located much further up or downstream the locus. Marker density is rather low checking all gene annotations in the interval between the bespoken QTL marker and its flanking markers is not feasible. Too many genes would have to be checked to give an overview. The Brassica database was used to check gene annotations. A list of all genes and their annotations can be found in Appendix H.

It can be expected that genes expressing a phenotype are similar even under different growing conditions. Therefore similar markers could be found to be associated to a trait in the greenhouse and the field experiment. The fact that no overlap was found however is not unlikely. The very different conditions in greenhouse and field can change the gene expression patterns and or gene expression levels of the genes underlying turnip formation.

Three markers were associated with turnip weight in the greenhouse experiment. The two markers found on chromosome A01 are segregating in 1:1 proportions both markers are from the IVF set. (see Tab.17) Interestingly the marker SNP_1074 shows a high proportion of heterozygous individuals

in the population. The third marker is from chromosome A09, it does not follow Mendelian segregation patterns.
One marker was found to be associated with turnip length in the greenhouse trial. This marker SNP_0898 from chromosome A09 is the same marker that was also found to be associated to turnip weight. The fact that it was found twice in the greenhouse experiment makes the marker SNP_0898 an even more valid true QTL for turnip formation.
In the field experiment 18 markers were found to be associated with turnip width. This unusually large number of markers seems suspicious to me, the fact that all these markers are highly distorted in segregation could point out that the result is flawed. In the very distorted markers the group of individuals homozygous for the allele of low proportion is small. In the ANOVA the precision of the average phenotypic value in this group will be relatively low. If individuals with unusually large phenotypic values are present in this group these outliers can be overrated in their importance. The small group is prone to imprecision because of high impact of outliers.
Overall I deem the markers found to be associated with turnip width in the field experiment the least credible among the markers found in all association studies.
One marker was found to be associated to turnip length in the field experiment. The marker SNP_0108 from chromosome A01 is segregating following Mendelian patterns and the relatively high $-\log_{10}$ (p-value) of 4.14 (Tab. 19) indicates a reliable QTL.
The fact that no data on turnip weight was available in the field experiment data is rather unfortunate, this way no comparison between greenhouse and field experiment can be done in that regard. The contribution of each QTL per trait on the realization of the phenotype is hard to determine in association studies compared to a classical QTL analysis, hence it is not mentioned here.

A number of genes were proposed to be involved in turnip formation before:
Lou et al. (2007) found a QTL (BrFLC2) for turnip formation on the upper part of chromosome A02. The QTL was found in separate analysis of weight, width and length data of turnips, similarly to the presented experiments. The FLC locus was identified in *A. thaliana* as a main QTL for flowering time. Paralogs of FLC were also found to influence flowering time in *B. rapa.* Lou et al. (2007) interpret the association of BrFLC2 with turnip formation as a possible epistatic effect. In this sense plants with later flowering time develop bigger turnips.
Lu et al. (2008) identified 18 QTLs for taproot formation traits the loci *qTRT4b* on A04 and *qTRW4* on A04 explained 55.7% of the variance. Other loci on A01, A02, A03, A04, A05, A06, A07 and A09, were also found to be associated to length, weight and thickness of the taproot.
Kubo et al. (2010) found two QTLs for turnip size and turnip weight on A01 and A05. Together they explain only 17% of the phenotypic variance. They also included the marker BrFLC2 on A02 that was proposed to be associated with turnip formation according to Lou et al. (2007). The same QTL was identified again, therefore hardening the evidence that BrFLC2 controlls turnip formation in *B. rapa.*
H. Habtemariam (2012) studied gene expression during turnip tuber formation in the Brassica research group in Wageningen UR. She found differential expression of genes Bra040904 and Bra035787 both are located on A05. The two genes express MATH protein domains, they have shown significantly different expression levels compared to the reference gene, expression was up to 5 fold higher.

Dr. Ningwen Zhang analysed a different population of *B. rapa* in the Brassica group of Wageningen UR. She found 15 QTLs for turnip weight on A01, A02, A06, A07, A08, A09 and A10. (unpublished data Dr. Ningwen Zhang)

She found three QTL markers associated with turnip formation A01, one was the marker SNP_1055. In the presented thesis the marker SNP_1056 was found to be associated with turnip weight. SNP_1055 and SNP_1056 are in very close proximity, they are only 52bp apart. This hardens the evidence for a true association to turnip formation for this marker. However it raised the question why SNP_1055 was not associated, both markers are segregating normally but the overall allele proportions are quite different. Both markers were mapped 4.8cM apart. Given the close physical distance between the markers the genotyping appears to be ambiguous.

The QTL BraVIM3P3b she found is located on chromosome A09 at 30.99Mbp. In the presented thesis the marker SNP_0898 on chromosome A09 is located at 30.93Mbp, both are in close proximity. It is likely that the same QTL was found. In the vicinity of the QTL, found by Dr. Ningwen Zhang, the genes Bra023122-Bra023124 annotated as, FAD-binding Baberine family protein were found to be likely candidate genes, behind the expression of the phenotype.
She identified the markers SNP_0609 and SNP_0616 as QTL. Both markers were not identified as QTLs in the presented thesis, however the markers SNP_0607 and SNP_0612 were found to be associated with turnip width, they are in close proximity.
The other QTLs found in the presented thesis are not comparable to the finds of Dr. Ningwen Zhang.

Apart from the markers described, a connection of the remaining QTL markers in the presented thesis, to markers in literature is not feasible, a co-location is unlikely. The gene annotations described in literature don't match with the annotations found in the QTL intervals listed in Appendix H. The interval was likely to small to find the genes expressing the phenotype.

BrFLC2 located at 2.9cM (Brassica database) was not identified as QTL for turnip formation in the presented thesis work, it does not co-locate with SNP_0191 and SNP_0205. Both markers were found to be associated with turnip width, they are located on A02.

5 Conclusion

The linkage map of RIL C29, created in this study is fractionated in comparison to the reference genome, and the marker order is confounded. In some cases markers from several chromosomes map together, mapping distances between markers are very large in some cases even bigger than 50cM.
Marker distortions lead to various problems in the determination of marker position, order and distance. Furthermore the loss of parental material for genotyping poses a range of problems in bringing markers in the right confirmation for map construction. The value of the map for breeding programs for instance is questionable. The allele with beneficial effects on a genotype found for QTLs may be hard to connect to a parent if parental allele denomination is not clear.

It is not expected that screening more markers or analysing more progeny for this particular population would give substantially better results in linkage mapping. Additional markers would increase the overall marker density, however getting to a clear grouping with one linkage group per chromosome is unlikely, due to strong distortion

In the marker trait association analysis 23 markers were found to be associated with a turnip formation related trait. Three markers on A01 and A09 were found to be associated with turnip weight. Two markers one on A01 and one on A09 were associated to turnip length. Eighteen QTL markers associated with turnip width that were found in a field experiment are extremely distorted the association is questionable. They are located in A02, A03, A04, A06, A07 and A08. Another QTL marker associated with turnip length and weight is strongly distorted too, the credibility of the association is ambiguous as well. Increasing marker density could improve the association study result if more undistorted markers associated with turnip formation can be found. Enriching the vicinity of proposed QTLs may give better indication on the true position of genes underlying the expression of the trait.

6 Future perspective

To improve mapping in the C29 population a computer based solution may be the best option. Zhu et al. (2007) developed a method to reconstruct linkage maps created from distorted markers in a rice backcross with 69.5% marker distortion. The same method may be applied on the C29 population.

Shanmugavadivel et al. (2013) write that segregation distortion increases through recurrent selfing. Therefore marker analysis of the F2 plants in C29 might be useful to create a better map and even to pinpoint the location of SDL. A previous master student of Dr. Ningwen Zhang has attempted to map the F2 population of C29, the process was problematic I suggest to check the data from this experiment this may help to understand the driving force of marker distortion better.

Mapping of *Brassica rapa* to find QTL for various QTLs is still an interesting topic, especially to find markers associated with turnip formation and to confirm the markers found in this study. I suggest to start a new population for mapping and QTL analysis.
This new population may be started with DH lines of *B. rapa* that are already available.
SD may not occur in the progeny at all or if it is still occurring the driving force would likely be epistatic effects resulting from unfavourable allele combinations of both parents. However epistatic interaction within one parent could be excluded. Additionally I would screen the DH lines for a number of markers to see if they are truly homozygous. Nine DH lines were included in my marker analysis experiments, all have shown heterozygosity to some extent, which is unexpected. This may be caused by imprecision of the genotyping method or a mix up of samples, but it is better to check before starting a new population.
Due to the homozygous nature of DH lines genotyping of parental lines can be done at any time, the parental lines are “immortal” new DNA can be isolated fresh for every experiment.

I suggest that the Brassica research group installs a central data storage where all raw data from experiments is gathered and backups can be made in frequent time intervals. This makes it easier for everyone to access information for all kinds of analysis. I furthermore suggest that other data such as primer sequences and a clear layout of the cold storage of DNA samples and primers should be saved in the central data storage as well.

7 References

Dixon, G. R. (2007). Vegetable Brassicas and Related Crucifers. *CAB International*.

Golméz - Campo, C. (1999). Biology of Brassica Coenospecies. *Elsevier science B. V.*

Habtemariam, H. G. (2012). Gene expression analysis during turnip tuber formation in Brassica rapa.

He, P., Li, J. Z., Zheng, X. W., Shen, L. S., Lu, C. F., Chen, Y., & Zhu, L. H. (2001). DH and RIL Populations Derived from the Same Rice Cross, *41*(AUGUST), 1240–1246.

Hirai, M., & Matsumoto, S. (2007). Brassica rapa. *GGenome Mapping and Molecular Breeding in Plants*, *5*(Li 1981), 1–6.

Iniguez-Luy, F. L., Lukens, L., Farnham, M. W., Amasino, R. M., & Osborn, T. C. (2009). Development of public immortal mapping populations, molecular markers and linkage maps for rapid cycling Brassica rapa and B. oleracea. *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik*, *120*(1), 31–43.

Jonathan D. Sauer. (1993). Historical Geography of Crop Plants. *CRC Press*.

Kubo, N., Saito, M., Tsukazaki, H., Kondo, T., Matsumoto, S., & Hirai, M. (2010). Detection of quantitative trait loci controlling morphological traits in Brassica rapa L. *Breeding Science*, *60*(2), 164–171.

Li, F., Kitashiba, H., Inaba, K., & Nishio, T. (2009). A Brassica rapa linkage map of EST-based SNP markers for identification of candidate genes controlling flowering time and leaf morphological traits. *DNA research : an international journal for rapid publication of reports on genes and genomes*, *16*(6), 311–23.

Li, X., Ramchiary, N., Choi, S. R., Van Nguyen, D., Hossain, M. J., Yang, H. K., & Lim, Y. P. (2010). Development of a high density integrated reference genetic linkage map for the multinational Brassica rapa Genome Sequencing Project. *Genome / National Research Council Canada = Génome / Conseil national de recherches Canada*, *53*(11), 939–47.

Lou, P., Zhao, J., Kim, J. S., Shen, S., Del Carpio, D. P., Song, X., Jin, M., et al. (2007). Quantitative trait loci for flowering time and morphological traits in multiple populations of Brassica rapa. *Journal of experimental botany*, *58*(14), 4005–16.

Lu, G., Cao, J., Yu, X., Xiang, X., & Chen, H. (2008). Mapping QTLs for root morphological traits in Brassica rapa L. based on AFLP and RAPD markers. *Journal of applied genetics*, *49*(1), 23–31.

Rakow, G. (2004). I . 1 Species Origin and Economic Importance of Brassica. *Biotechnology in Agriculture and Forestry*, *54*(Vol. 54).

Reiner, H., Holzner, W., & Ebermann, R. (1995). The development of turnip-type and oilseed-type Brassica rapa crops from the wild type in europe - an overview of botanical, historical and linguistical facts. *International Rapeseed Congress*, *4*(July), 1066–1069.

Schranz, M. E., Lysak, M. a, & Mitchell-Olds, T. (2006). The ABC's of comparative genomics in the Brassicaceae: building blocks of crucifer genomes. *Trends in plant science*, *11*(11), 535–42.

Shanmugavadivel, P. S., Mithra, S. V. A., Dokku, P., Kumar, K. A. R., Rao, G. J. N., Singh, V. P., Singh, a. K., et al. (2013). Mapping quantitative trait loci (QTL) for grain size in rice using a RIL population from Basmati × indica cross showing high segregation distortion. *Euphytica*, *194*(3), 401–416.

Taylor, D. R., & Ingvarsson, P. K. (2003). Common features of segregation distortion in plants and animals. *Genetica*, *117*(1), 27–35.

Törjék, O., Witucka-Wall, H., Meyer, R. C., Von Korff, M., Kusterer, B., Rautengarten, C., & Altmann, T. (2006). Segregation distortion in Arabidopsis C24/Col-0 and Col-0/C24 recombinant inbred line populations is due to reduced fertility caused by epistatic interaction of two loci. *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik*, *113*(8), 1551–61.

Waller, D. M., Dole, J., & Bersch, A. J. (2008). Effects of stress and phenotypic variation on inbreeding depression in Brassica rapa. *Evolution; international journal of organic evolution*, *62*(4), 917–31.

Wang, J., Lydiate, D. J., Parkin, I. a P., Falentin, C., Delourme, R., Carion, P. W. C., & King, G. J. (2011). Integration of linkage maps for the Amphidiploid Brassica napus and comparative mapping with Arabidopsis and Brassica rapa. *BMC genomics*, *12*(1), 101.

Wang, X., Wang, H., Wang, J., Sun, R., Wu, J., Liu, S., Bai, Y., et al. (2011). The genome of the mesopolyploid crop species Brassica rapa. *Nature genetics*, *43*(10), 1035–9.

Xiao, D., Wang, H., Basnet, R. K., Zhao, J., Lin, K., Hou, X., & Bonnema, G. (2014a). Genetic Dissection of Leaf Development in Brassica rapa Using a Genetical Genomics Approach. *Plant physiology*, *164*(3), 1309–25.

Xiao, D., Wang, H. G., Basnet, R. K., Zhao, J. J., Lin, K., Hou, X. L., & Bonnema, G. (2014b). Genetic dissection of leaf development in Brassica rapa using a “genetical genomics” approach. *Plant physiology*.

Yang, X., Yu, Y., Zhang, F., Zou, Z., Zhao, X., Zhang, D., & Xu, J. (2007). Linkage Map Construction and Quantitative Trait Loci Analysis for Bolting Based on a Double Haploid Population of Brassica rapa, *49*(5), 664–671.

Yu, S., Zhang, F., Yu, R., Zou, Y., Qi, J., Zhao, X., Yu, Y., et al. (2009). Genetic mapping and localization of a major QTL for seedling resistance to downy mildew in Chinese cabbage (Brassica rapa ssp. pekinensis). *Molecular Breeding*, *23*(4), 573–590.

Yu, S., Zhang, F., Zhao, X., Yu, Y., Zhang, D., Zhao, X., & Wang, W. (2012). An Improved Brassica rapa Genetic Linkage Map and Locus-specific Variations in a Doubled Haploid Population. *Plant Molecular Biology Reporter*, *31*(3), 558–568.

Zhao, J., Wang, X., Deng, B., Lou, P., Wu, J., Sun, R., Xu, Z., et al. (2005). Genetic relationships within Brassica rapa as inferred from AFLP fingerprints. *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik*, *110*(7), 1301–14.

Zhu, C., Wang, F., Wang, J., Li, G., Zhang, H., & Zhang, Y. (2007). Reconstruction of linkage maps in the distorted segregation populations of backcross, doubled haploid and recombinant inbred lines. *Chinese Science Bulletin*, *52*(12), 1648–1653.

Appendices

Appendix A

To prepare PCR mastermix MQ, reaction buffer, dNTPs and primer solution are pipette into a 2ml Eppendorf tube. The tube is then transferred to a box of ice. The reaction mix will store this way on ice for several hours without forming primer dimers, this also depends on the primer design. Right before the master mix will be pipetted into the 96 well PCR plates, LC green and phire enzyme will be added. The mix needs to be vortexed before that. The fluorescence dye LC green is light sensitive and will deteriorate in the mastermix if exposed to light for too long before starting the PCR reaction. After pipetting the mastermix into the PCR plates, which will be done also on ice, the template DNA will be added using an 8 or 12 channel pipette. Lastly the reaction mix in every well is covered with 20µl of mineral oil. Before starting the PCR reaction in the thermo cycling machine the PCR plates need to be centrifuged briefly. The centrifuge will be accelerated to 1000rpm and then stopped.

Tab. 1: Lightscanner PCR reaction mix

Master mix	µl	x 14	x100
Phire enzyme	0.1	1.4	10
Reaction Buffer for Phire (5x)	2	28	200
LC green	1	14	100
dNTP	0.4	5.6	40
forward Primer	0.5	7	50
reverse Primer	0.5	7	50
MQ	4.5	63	450
Template DNA	1	14	100
Total volume	10	140	1000
Mineral oil	20	280	2000

Appendix B

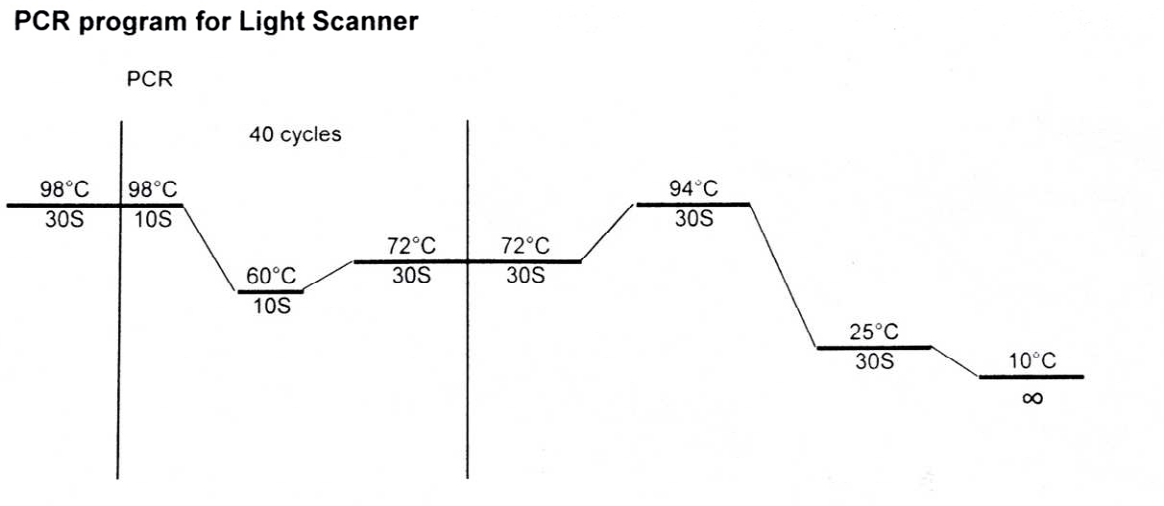

Fig. 1: Hot start PCR program for Lightscanner PCR reaction

Appendix C

name primer	**Chromosome**	**sequence** (in 5'----> 3' order)
SNP_1056_1F	A01	AAACCATTGGCAATTATGCTG
SNP_1056_1R	A01	TTCGCAACAGACTGTGCTTT
SNP_1056_2F	A01	AAACCATTGGCAATTATGCTG
SNP_1056_2R	A01	TCCTTTAGCTTTCGCAACAGA
SNP_1063_1F	A01	CCACCTCGTACTTTTCGGAAT
SNP_1063_1R	A01	TAAACCGGGAAGTTTCTTCG
SNP_1063_2F	A01	GGAGAACATCTCCACCTCGT
SNP_1063_2R	A01	TGAGATTCCGGATAGCATAGC
SNP_1073_1F	A01	AAACGTCTCCGTTCAGCAAC
SNP_1073_1R	A01	GCGGGAGAGACAATTTCTTTTT
SNP_1073_2F	A01	TCCACTACAACGCAAACGTC
SNP_1073_2R	A01	TTTCTTTTTATCCGTCGTGAAGA
SNP_1074_1F	A01	CTAAAGAGCCGTCCCGTTTT
SNP_1074_1R	A01	GCTATCTTCGGGCATGTGAC
SNP_1074_2F	A01	CTAAAGAGCCGTCCCGTTTT
SNP_1074_2R	A01	GCTATCTTCGGGCATGTGA
SNP_0184_F	A02	GGGCGATCTCATGACGTAGT
SNP_0184_R	A02	TACCAGTTCTGTCGCCATTG
SNP_0185_F	A02	CGCAATCACATTTCCTTCTG
SNP_0185_R	A02	CAAGATTGTGTGTGCAAGGAA
SNP_0188_F	A02	ACTCCGTCTCGGTCGAAATA
SNP_0188_R	A02	GTTCAACATGCCTGGAAACTC
SNP_0191_F	A02	TTTTCGCCATCAAAATGGTT
SNP_0191_R	A02	ACAACGGAGTTGCTTCTGCT
SNP_0198_F	A02	TGGACAGTCCTCCTCATCAA
SNP_0198_R	A02	CTTGGCCCATGGTTAAGAGA
SNP_0200_F	A02	TGCGTCCCTTCTTTCAAGTC
SNP_0200_R	A02	TGGGATATGTCCAAGTTCTGG
SNP_0201_F	A02	TCGATGCAATGTTGGATGTT
SNP_0201_R	A02	GAGTTCTTGATTTACCGCGAGA
SNP_0205_F	A02	GCTTTCATGAAGCTGTTGACAC
SNP_0205_R	A02	TGCCCATGCTTGTAAGAAAA
SNP_0214_F	A02	GTTAAGGAAGCTTGTGCTTGG
SNP_0214_R	A02	GACTACCCATGCAAAGAATAACG
SNP_0217_F	A02	GCTTGTCAGACACGTTACCATT
SNP_0217_R	A02	GCGAAAGGAAGCTCAGAGAA
SNP_0220_F	A02	GAAGCCACAAGCCTATCAGC
SNP_0220_R	A02	TTTGATTTGCTGCAGAGGAG
SNP_1075_1F	A02	GCTATTGATGGTTTCCGAATTG
SNP_1075_1R	A02	CGATTGCAACACCTCGTTT
SNP_1075_2F	A02	GCTATTGATGGTTTCCGAATTG
SNP_1075_2R	A02	CAAACCTTTGCATGCGATT
SNP_1076_1F	A02	CGTGAACCCTTGAATCAACC
SNP_1076_1R	A02	AATGTGGAGCTGAACGATCC
SNP_1076_2F	A02	GACCACGTGAACCCTTGAAT
SNP_1076_2R	A02	TGGAGCTGAACGATCCAGA
SNP_1077_1F	A03	ACATTTGCCCAAAAGGACAG
SNP_1077_1R	A03	CCGGTTCAGGGGTAGAGG
SNP_1077_2F	A03	ACATTTGCCCAAAAGGACAG
SNP_1077_2R	A03	GCCGGTTCAGGGGTAGAG

SNP_1078_1F	A03	CGATTTCTCGGAGTTTCTGG
SNP_1078_1R	A03	GACCAACCGGGTATGTTTGT
SNP_1078_2F	A03	CGATTTCTCGGAGTTTCTGG
SNP_1078_2R	A03	ACCAACCGGGTATGTTTGTC
SNP_1079_1F	A03	TACGGTTTTGGCAATGGTCT
SNP_1079_1R	A03	GCAGCACTTACTCCACCACA
SNP_1079_2F	A03	TACGGTTTTGGCAATGGTCT
SNP_1079_2R	A03	TGCAGCACTTACTCCACCAC
SNP_1080_1F	A03	CGGTTTCGAACTCTTCCTCA
SNP_1080_1R	A03	GCAGCACTTACTCCACCACA
SNP_1080_2F	A03	AGGGAAACCAACTTCCATGA
SNP_1080_2R	A03	TGCAGCACTTACTCCACCAC
SNP_0325_F	A04	GCACTACGTAGAGGAGCTTGA
SNP_0325_R	A04	GCCAACACATGCCCTTTCC
SNP_0326_F	A04	TAGGCCATTAGGCCCAAAAT
SNP_0326_R	A04	TCAAAGCGTCAGTCATCACC
SNP_0335_F	A04	GCAATCGTGGAAGACCAAGT
SNP_0335_R	A04	TTATGACCTTCGATCTCGACAC
SNP_0337_F	A04	GAAAACGGTGTGGCGATT
SNP_0337_R	A04	AGTGTTCTCCGGCTATGTCAG
SNP_0343_F	A04	GGCCGTGGAAATAAGACAGA
SNP_0343_R	A04	CAAGTGGAATTCAAGATCACACA
SNP_0344_F	A04	GCACACAGAAATACCCAGTCG
SNP_0344_R	A04	GTCCTTTTGCTGCTCATCGT
SNP_0351_F	A04	CTCGGAGTGAACCCAAGC
SNP_0351_R	A04	TCGATGTTTGCATCTTTCCA
SNP_0352_F	A04	AAACCGTGAAGCCCTAGGAT
SNP_0352_R	A04	GACACTCGCCTTTACGTGATT
SNP_0367_F	A04	GGCCCATGATTTGATCCTC
SNP_0367_R	A04	GGGAAAAGATGCAGTGGAGT
SNP_0370_F	A04	CTGCGGTGACACGAGAGTT
SNP_0370_R	A04	CAACATTTTTGCACGTCGTC
SNP_1058_1F	A04	CAAGTGATGAAGAAAGCTCAGG
SNP_1058_1R	A04	GGAAGGTTCTTCACGTCGTC
SNP_1058_2F	A04	TCCTCAAGTGATGAAGAAAGCTC
SNP_1058_2R	A04	GGAAGGTTCTTCACGTCGTC
SNP_1059_1F	A04	TGAAGAAAGCTCAGGCAGAAG
SNP_1059_1R	A04	GGAGAGGAATCACAGGTTCG
SNP_1059_2F	A04	GATGAAGAAAGCTCAGGCAGA
SNP_1059_2R	A04	GGAGAGGAATCACAGGTTCG
SNP_1070_1F	A06	AGGGTTAGACGTGGTGAAGG
SNP_1070_1R	A06	GACAAGGACGCTGACGCTAT
SNP_1070_2F	A06	AGGGTTAGACGTGGTGAAGG
SNP_1070_2R	A06	CAAGGACGCTGACGCTATAA
SNP_0608_F	A07	CTGTTTCCATGCAAGGCGAC
SNP_0608_R	A07	TCAGACAAATGAACCGAGAAGT
SNP_0612_F	A07	ACAAGCAAGCAACAGTTTCAGT
SNP_0612_R	A07	ACTCAGGGTTGGTTCTGCAC
SNP_0613_F	A07	TGCGTCAACTTCAGCATCCT
SNP_0613_R	A07	GTAAATCTCTCTTCGCCGCG
SNP_0621_F	A07	GGTAAGCAGTTCACAGTGGACT
SNP_0621_R	A07	CCTTACGCTTCTTCCCTTTCCT

SNP_0625_F	A07	TGCACCTGGAAGCTGTATCA
SNP_0625_R	A07	TCATCCACCGGTGAGTGAGA
SNP_0630_F	A07	TGCAAAGCCACTGATATTTCGG
SNP_0630_R	A07	GGACGGCCTTTTGTATGTGG
SNP_0632_F	A07	TCATTGCTCTGCCAGGGAAC
SNP_0632_R	A07	TCTTCATCTGGTTTCCACTGGT
SNP_0635_F	A07	AGATATCTCGGTGGAATGCCC
SNP_0635_R	A07	GGCCCTTGAGATGTTTGGGA
SNP_0636_F	A07	GATTCTCTTTCGGCGCGC
SNP_0636_R	A07	TGACACGGACGTTCAAGATGT
SNP_0640_F	A07	ACACACGAGGAATCTGAACAGT
SNP_0640_R	A07	AGCGTGGATCCGTTCTCAAT
SNP_0642_F	A07	GCACCGGAGATTCTGAAAGGA
SNP_0642_R	A07	CTGCTTCGTACCTGCTTACAATG
SNP_0643_F	A07	AAACTACTAACGGTGCTGAGGA
SNP_0643_R	A07	CGCGGAAGGAAACGATGAAG
SNP_0647_F	A07	CGGGGAGTAACTAGTATGGTTGA
SNP_0647_R	A07	GGGGAACAAGACCACGATTCA
SNP_0649_F	A07	GAAACTGTGTCTTCGCTGCC
SNP_0649_R	A07	GCAACCGCGTCTCTTAAAACT
SNP_0650_F	A07	ATTAGCCCCAGTCCAAACCC
SNP_0650R	A07	GAAGACGTTTGTTCCTGGACAG
SNP_0652_F	A07	GGTTCTGCGTATTCTCCGGT
SNP_0652_R	A07	CGAAGGTCAACTCTCCGTCC
SNP_0654_F	A07	TCTTTTGAGTTCCCGGTCGAC
SNP_0654_R	A07	CACCCTGAATCCGCCAAGAT
SNP_0658_F	A07	TGATTGCGTCGCTTCCCATA
SNP_0658_R	A07	AAATCCGCATGTTGTTCACGT
SNP_0659_F	A07	TTGTTCTGCATCCATGGTGC
SNP_0659_R	A07	ACTTCATCTCTAGCACAAAGTCG
SNP_0660_F	A07	TGGACACTACTATCCGCAGC
SNP_0660_R	A07	GGGACACCTTGAAAAGTTTGTAA
SNP_0662_F	A07	TTTCGTCTCACCGTCCACTG
SNP_0662_R	A07	ACCTCCCTGTCATAAGTCAGACT
SNP_0667_F	A07	CCTAACCTCCTCCTCGTCCT
SNP_0667_R	A07	TGTGGTGGTGAAGGAAGCAG
SNP_0669_F	A07	CATCAGCCGCCGGTTCAT
SNP_0669_R	A07	AGTCCCTGCGGTTTACTTGG
SNP_0672_F	A07	ACCAAGGAAAGAGTCAGTGTGA
SNP_0672_R	A07	CACCGCCTTCTTCATCAAGC
SNP_0674_F	A07	CGCAGTTATCGAAGGAGGCA
SNP_0674_R	A07	GTGCTCTGCTTCACCTGTGT
SNP_0675_F	A07	GCAAGAAAGGCCTCGATGAC
SNP_0675_R	A07	TGAGCGTTCACATTATTGCAAT
SNP_0676_F	A07	GAACAATGGACTTTCTCTTCCCA
SNP_0676_R	A07	GGATTCAAAACAGCTTGCACCT
SNP_0679_F	A07	TTTCTTTTCTTCGTTATGGTTGGT
SNP_0679_R	A07	TCACTGTGGTTGAGCGATGT
SNP_0684_F	A07	GGGTGAGAGGCTTAGGAAACC
SNP_0684_R	A07	GCACATGAGTCTTCCGCGA
SNP_0686_F	A07	CCGGAAACCACACTCAACCT
SNP_0686_R	A07	AGGGGAAGAAGAGCAACCAAA
SNP_0689_F	A07	TCCATGTCACGGTGCTATGT
SNP_0689_R	A07	TGGCTTTGAGTAGTGTAGCTCA
SNP_0690_F	A07	TGGGAAGCCTTGGAAAGCTG

SNP_0690_R	A07	GCTCCAATGGCGCGAATAAC
SNP_0720_F	A08	CCTTCGGATCCAGAAATCAA
SNP_0720_R	A08	CGGTTCCTGACTTTGATCGT
SNP_0731_F	A08	TGCTTACTTGAAGAAAACCCACT
SNP_0731_R	A08	AACCTCTCACAAGGGAGGAAT
SNP_0732_F	A08	CAAACTTTGACGCCATCTCC
SNP_0732_R	A08	ACGGCGAACTCGTGTATTCT
SNP_0739_F	A08	GACGTTGAGGTCGTGTCACTAA
SNP_0739_R	A08	CACTCAGATCGAAAAGCCAGA
SNP_0741_F	A08	GGGCCTGAATCTCGATGAC
SNP_0741_R	A08	AGCACGTGGAGTTTTGCATT
SNP_0747_F	A08	CGCTTATAGTCGTCCTGCAAA
SNP_0747_R	A08	CCGTTCCCAGGTAAATAGAAAA
SNP_0748_F	A08	GGCTTACACTCGGCAAATTC
SNP_0748_R	A08	CAGATACACCATCTTGTTTTGGA
SNP_0751_F	A08	GGAGAGTCTGCAGATGCTGTT
SNP_0751_R	A08	CGATTCCTCAATCTTCTCAACC
SNP_0753_F	A08	TTGCAACAAGCGCTTAGAAA
SNP_0753_R	A08	TTCCGGGTTTTGACTTTTTG
SNP_0763_F	A08	TGCAAATATGTGTAGTGGCACA
SNP_0763_R	A08	TAACTGAAGCGGATCTGTGTGT
SNP_0765_F	A08	TCTCCTTCCCAACTGTAATGAAT
SNP_0765_R	A08	GAAAGGGTTTCAGGTATTTCTTTTT
SNP_0768_F	A08	AACAGAGGAACCACCTGGAA
SNP_0768_R	A08	CTCTGAAGCAAACCCCAGTC
SNP_0770_F	A08	TCAGATGACTATTGAGTCCTGCAT
SNP_0770_R	A08	GCGTCCCTTGTACCTCTACCT
SNP_0786_F	A08	GGTGTGAAGAAACTCGAGCA
SNP_0786_R	A08	TCAACGTGACCTTCATGAGAA
SNP_1060_1F	A08	TCCAGCTTCTAGCAACTCCAA
SNP_1060_1R	A08	CGTTTGCGTTCTTGAGAAAA
SNP_1060_2F	A08	GCTCCAGCTTCTAGCAACTCC
SNP_1060_2R	A08	CGTTTGCGTTCTTGAGAAAA
SNP_0801_F	A09	TTCTCGAGCTTAGCCGCAAA
SNP_0801_R	A09	GCTGCAGGCGTCAGGAAC
SNP_0802_F	A09	TCTCTCCCATGATGCTAAAGAAGT
SNP_0802_R	A09	GTGTTCATCTACATTGGTGCTGG
SNP_0808_F	A09	TCCAAACTATGAATGATAATATCCAG
SNP_0808_R	A09	GCCTTTGGTGTCATATCTCCTT
SNP_0815_F	A09	GGCACATCTTACGTCCAGGT
SNP_0815_R	A09	CGATTAATTCTAGTTAAAGGACGAAGC
SNP_0817_F	A09	TTCCGCTATTATGGTTTTGACCG
SNP_0817_R	A09	GCCCTTCATCAGCTACGGAA
SNP_0818_F	A09	GAAGATGACGCAATCTGCAA
SNP_0818_R	A09	GTCGTCCTCGTCGTCAGTCT
SNP_0842_F	A09	CTGGTGTTGAAGCTGAGTTATATGT
SNP_0842_R	A09	GTTTGGTCCCATGCGAATCC
SNP_0843_F	A09	AGACCATCCGTATAATCAGGGC
SNP_0843_R	A09	GTTAAAGTCCCGACCGCAGA
SNP_0844_F	A09	TTGGATGAGAAAAGCTCTGACA
SNP_0844_R	A09	TGTGCGCTCTAAGTCTTTCG
SNP_0857_F	A09	TCTGGTCATGAGGAAAGACAAACA
SNP_0857_R	A09	TCTCCATGGATCTCAGGCTCA

SNP_0861_F	A09	TCTATCTTTGAATGTTAGGTAGGCC
SNP_0861_R	A09	GAGTTTCTTGGCGCCCTTAG
SNP_0862_F	A09	TCCTAACCGGTACCAAATCTTC
SNP_0862_R	A09	GCCAAAGAAGGCTTCAACAT
SNP_0870_F	A09	TCTAACCTCCTAACCTCCTTCACA
SNP_0870_R	A09	GTACTTTCCTCAGGCAGCGT
SNP_0875_F	A09	TGATTTTGGCATCTCTTTCAGTGG
SNP_0875_R	A09	TCCTGCTGCTTTCGTCACAA
SNP_0876_F	A09	ACGGTGACGAGGTAGCAGAT
SNP_0876_R	A09	TTTTGTCCCCAAAATCCATC
SNP_0891_F	A09	GGACGACGAGATTTGGACCA
SNP_0891_R	A09	CACGTTAAGCCGCAACCAAA
SNP_0896_F	A09	AGCGTTGACTCTCTTCACGT
SNP_0896_R	A09	TCACCAAGGCACCAGTGATC
SNP_0898_F	A09	CAATTTCGCATCACAATTCG
SNP_0898_R	A09	TCCATTGAATGTTTTGTTTTTGA
SNP_0912_F	A09	ACCGGGAGAGACTTGAGGTG
SNP_0912_R	A09	CTCCTCAGACTCCTCCAGCT
SNP_0915_F	A09	ATGCGCACAATCCATCTCATC
SNP_0915_R	A09	GAAGAAGTAAGCTAGTCTCTGCGA
SNP_0916_F	A09	CACGCTCATAACCTTCAGCA
SNP_0916_R	A09	CACAACAGATTCAAAGGCGTTA

Primer	**Chromosome**	**Primer sequence**(in 5'----> 3' order)	**paralog gene in A. thaliana**
L_005F	A08	AGTGCTTGTTGATGCCTTGTC	AGO1P2cF
L_005R	A08	ACTTGGTAACTTCAGGCCAGTC	AGO1P2cR
L_019F	A01	CATCATCATCTCCTTCCTCCAC	ATGRF2P2bF
L_019R	A01	GAGTTTCCCCATGAGATTGGT	ATGRF2P2bR
L_031F	A02	ACCCTCGAACATCTCCACAA	CUC2P1dF
L_031R	A02	CACAGGCGGTAGAAGATGGT	CUC2P1dR
L_033F	A10	GGTTTGACCATTTTTCTCGTTT	CUC2P2bF
L_033R	A10	CAGCCTTTGTCTCGTTCTCTTC	CUC2P2bR
L_039F	A09	CTGATAATGTCCAAGACCGTGA	ERP1bF
L_039R	A09	AAACATAAAGCTGCCGAGGA	ERP1bR
L_042F	A05	TTTCGCTCCCTTGTCTCACT	HST1P2aF
L_042R	A05	CTTGCGGGTAGACACGATCT	HST1P2aR
L_048F	A06	GTACACAGGAGCTGCAACAAAG	HYL1P2aF
L_048R	A06	CTTTCCTTTCCGAGCCTTCT	HYL1P2aR
L_050F	A10	TCGCATGTACCGTTTCTGTC	HYL1P3aF
L_050R	A10	TCTTGCAACGTGTTTTGTCCT	HYL1P3aR
L_057F	A07	TCGGAGCCAATAAATGTCCT	KAN1P2aF
L_057R	A07	AGTTCTCATCATCGTCGTCCTC	KAN1P2aR
L_065F	A09	GCTCGACTTAGGGTTTTGGAG	KAN2P3aF
L_065R	A09	TGTTGTGTCGGAAACTGGAA	KAN2P3aR
L_083F	A03	GAAGAAGCAAGAAGAGGTTGGA	LEPP1aF
L_083R	A03	TGGACAAGTAAGAGAAGCAGCA	LEPP1aR
L_089F	A05	GGGTAGTAATGACGGTGCAGA	PHBP2aF
L_089R	A05	AGGAATCTAACCAGCACAGCA	PHBP2aR
L_092F	A09	TATACCCACTGGAAATGGTGGT	PHVP2aF

L_092R	A09	CACAAAGCTCGAAGAAAGTGG	PHVP2aR
L_107F	A09	GTCACCGTTGCTGATTATGGT	RON1P3aF
L_107R	A09	CGACCGTTAGGACCACCTT	RON1P3aR
L_110F	A01	GCATACCCCAAATCAATCACT	ROT3P1aF
L_110R	A01	GAACCAAGGGAAGTTGAGACC	ROT3P1aR
L_114F	A03	ATGGCACCAGAGGAGAATAGCA	ROT4P1aF
L_114R	A03	GTCACGCCAGCAGACTAACA	ROT4P1aR
L_130F	A09	CCAAGAGAGCGGTAACAACAA	WUSP3aF
L_130R	A09	GGATGAAGTTGTTGGTGTGGT	WUSP3aR
L_139F	A08	AAAGCCCACCTACTCCCATT	YAB2P2aF
L_139R	A08	CGGACGAATAGGAGGCATT	YAB2P2aR
L_154F	A02	TGTCCCCTGAAATCGAGCTA	JAGP1cF
L_154R	A02	GGGCGTACGGATGTTAAAGAG	JAGP1cR
L_167F	A08	TTTGTTCTGGGAAGACGAAGA	CycD3;3P3bF
L_167R	A08	AAGCAACAGAAACGAGCTGAA	CycD3;3P3bR
L_168F	A01	GCAAGATCCTCGTTTCTCACC	PPD2P1aF
L_168R	A01	ATAGTGAGCAGAGCCACCAGA	PPD2P1aR
L_170F	A08	AACAAAAAGGCACCAGGAGTT	PPD2P2aF
L_170R	A08	GTGACTCGTGCTGTTCTCCTC	PPD2P2aR
L_181F	A02	ATGGACTTGTTATGCGGAGAG	CYCD1;1P1bF
L_181R	A02	AGCGTCCAGAGATTGAGAGTG	CYCD1;1P1bR
L_182F	A03	TCGAGCAACATCAAGAACAGA	ARLP1aF
L_182R	A03	GGAAGAATCAACGGAAGGAAC	ARLP1aR
L_187F	A05	CGAGCCCTGTTTCAGTTCTT	lng2P2aF
L_187R	A05	TGCGTTGGACCCTTATTCTC	lng2P2aR
L_193F	A03	TCTCAACCTTCCAGAACGAA	PKLP1aF
L_193R	A03	ACCACGCAGACCTCAAAAAG	PKLP1aR
L_195F	A08	GCGATACCAAAGGAGGAAGAA	CycD3;1P1aF
L_195R	A08	CTTCGTCTTCCCAGAACAAATC	CycD3;1P1aR
L_206F	A08	GCTAGAGGGTTACGTGAGGAAG	cow1P3aF
L_206R	A08	TCGACATTAGCCGGTTTCA	cow1P3aR
L_214F	A08	CAGAGTTGGAGGATTCGGAG	CNAP3bF
L_214R	A08	TAAGTTCAGCACACGTTCCAAG	CNAP3bR
L_219F	A06	ACACTTGATGGGTTTGAGGAAT	CycB2;3P1aF
L_219R	A06	TGTGATGGAAATCAACCATCTT	CycB2;3P1aF
L_230F	A10	CAGAGGATTCGGACAACAGAC	LNG1P2aF
L_230R	A10	GTCTTCATCATCCTCGTCCAA	LNG1P2aR

Appendix D

Tab. 22: Marker distortion and allele distribution of all polymorphic markers ordered according to the physical mapping position

Nr	Locus	a	h	b	X2	Signif.
1	SNP_0002A01	70	14	33	13.29	******
2	SNP_0003A01	56	17	40	2.67	-
3	SNP_0005A01	70	14	32	14.16	******
4	SNP_1074A01	20	46	38	5.59	**
5	L_110A01	68	11	26	18.77	*******
6	L_195A01	83	11	12	53.06	*******
7	SNP_0014A01	108	3	6	91.26	*******
8	SNP_0015A01	86	8	22	37.93	*******
9	SNP_1063A01	93	6	1	90.04	*******
10	SNP_1064A01	38	10	60	4.94	**
11	SNP_0017A01	83	22	4	71.74	*******
12	SNP_0029A01	33	10	64	9.91	****
13	SNP_0033A01	56	10	43	1.71	-
14	SNP_0037A01	101	6	9	76.95	*******
15	L_168A01	80	8	17	40.92	*******
16	SNP_0055A01	99	7	10	72.67	*******
17	SNP_0058A01	99	6	11	70.4	*******
18	SNP_0059A01	97	6	12	66.28	*******
19	SNP_0060A01	96	7	12	65.33	*******
20	SNP_1055A01	52	16	49	0.09	-
21	SNP_1056A01	37	18	45	0.78	-
22	SNP_0086A01	35	7	57	4.94	**
23	SNP_0089A01	35	0	39	0.22	-
24	SNP_0098A01	107	4	5	92.89	*******
25	SNP_0101A01	88	7	19	44.5	*******
26	SNP_0104A01	59	18	36	5.57	**
27	SNP_0108A01	44	16	56	1.44	-
28	SNP_0118A01	13	49	55	25.94	*******
29	SNP_0125A02	41	17	59	3.24	*
30	SNP_0128A02	43	18	55	1.47	-
31	SNP_1075A02	58	7	27	11.31	*****
32	SNP_0130A02	48	16	50	0.04	-
33	SNP_0131A02	71	4	33	13.88	******
34	SNP_1035A02	4	26	84	72.73	*******
35	L_031A02	76	1	34	16.04	*******
36	SNP_0145A02	43	20	52	0.85	-
37	SNP_0148A02	113	0	1	110.04	*******
38	SNP_0157A02	52	18	42	1.06	-
39	SNP_0158A02	63	16	37	6.76	***

40	L_154A02	82	7	16	44.45	*******
41	L_181A02	81	9	18	40.09	*******
42	SNP_0172A02	58	13	46	1.38	-
43	SNP_0177A02	108	1	1	105.04	*******
44	SNP_0182A02	96	0	1	93.04	*******
45	SNP_188A02	108	0	4	96.57	*******
46	SNP_191A02	100	0	13	66.98	*******
47	SNP_0193A02	68	14	33	12.13	******
48	SNP_0194A02	53	13	49	0.16	-
49	SNP_0201A02	65	1	42	4.94	**
50	SNP_0205A02	88	14	12	57.76	*******
51	SNP_0206A02	46	0	35	1.49	-
52	SNP_0207A02	18	10	89	47.11	*******
53	SNP_0217A02	102	0	10	75.57	*******
54	SNP_0222A03	56	4	45	1.2	-
55	SNP_1077A03	86	17	0	86	*******
56	SNP_0226A03	103	6	5	88.93	*******
57	L_083A03	97	2	11	68.48	*******
58	SNP_0243A03	108	0	7	88.7	*******
59	L_114A03	70	2	40	8.18	****
60	L_182A03	91	10	6	74.48	*******
61	SNP_0253A03	65	15	37	7.69	***
62	SNP_0255A03	2	0	98	92.16	*******
63	SNP_0264A03	67	18	31	13.22	******
64	SNP_0267A03	112	0	4	100.55	*******
65	SNP_1050A03	57	15	45	1.41	-
66	SNP_0284A03	52	15	50	0.04	-
67	SNP_0286A03	45	14	54	0.82	-
68	SNP_1080A03	70	6	22	25.04	*******
69	SNP_0297A03	45	19	52	0.51	-
70	SNP_0298A03	72	1	38	10.51	****
71	SNP_0309A03	54	16	45	0.82	-
72	SNP_1065A03	106	7	2	100.15	*******
73	SNP_0312A03	51	14	48	0.09	-
74	L_193A03	46	8	40	0.42	-
75	L_019A03	50	15	37	1.94	-
76	SNP_0319A04	36	25	56	4.35	**
77	SNP_0320A04	102	4	5	87.93	*******
78	SNP_0321A04	53	27	36	3.25	*
79	SNP_0323A04	53	27	37	2.84	*
80	SNP_0325A04	92	9	12	61.54	*******
81	SNP_1058A04	46	16	38	0.76	-
82	SNP_1059A04	59	18	23	15.8	*******

83	SNP_0331A04	26	11	80	27.51	*******
84	SNP_0335A04	98	0	16	58.98	*******
85	SNP_0342A04	111	4	2	105.14	*******
86	SNP_0343A04	74	0	34	14.81	******
87	SNP_0345A04	6	2	105	88.3	*******
88	SNP_0349A04	78	14	25	27.27	*******
89	SNP_0350A04	47	17	52	0.25	-
90	SNP_0351A04	93	3	14	58.33	*******
91	SNP_0353A04	44	16	38	0.44	-
92	SNP_0363A04	65	11	32	11.23	*****
93	SNP_0364A04	103	0	8	81.31	*******
94	SNP_0375A05	60	13	43	2.81	*
95	SNP_0381A05	29	0	67	15.04	******
96	L_187A05	87	12	9	63.38	*******
97	L_089A05	53	15	40	1.82	-
98	SNP_0393A05	49	12	53	0.16	-
99	SNP_0413A05	113	2	0	113	*******
100	SNP_0415A05	70	12	33	13.29	******
101	SNP_0424A05	54	3	56	0.04	-
102	SNP_0426A05	51	7	56	0.23	-
103	SNP_0430A05	93	5	17	52.51	*******
104	SNP_0448A05	91	4	20	45.41	*******
105	SNP_0476A05	43	14	56	1.71	-
106	SNP_1052A05	9	12	71	48.05	*******
107	SNP_1051A05	84	13	19	41.02	*******
108	SNP_0492A05	89	9	16	50.75	*******
109	SNP_0510A05	22	12	83	35.44	*******
110	SNP_0512A05	11	0	103	74.25	*******
111	L_042A05	88	5	14	53.69	*******
112	SNP_1043A06	0	1	114	114	*******
113	SNP_0519A06	33	18	64	9.91	****
114	SNP_0521A06	74	19	23	26.81	*******
115	SNP_0526A06	88	8	19	44.5	*******
116	SNP_0528A06	20	9	85	40.24	*******
117	SNP_1033A06	3	57	24	16.33	*******
118	SNP_1034A06	107	6	3	98.33	*******
119	L_048A06	87	5	13	54.76	*******
120	SNP_0534A06	100	3	13	66.98	*******
121	L_219A06	94	8	8	72.51	*******
122	SNP_0565A06	97	8	8	75.44	*******
123	SNP_0571A06	101	3	11	72.32	*******
124	SNP_0573A06	100	0	11	71.36	*******
125	SNP_0597A06	24	12	80	30.15	*******

126	SNP_0601A06	19	28	51	14.63	******
127	SNP_0602A06	104	1	1	101.04	*******
128	SNP_1047A07	2	0	115	109.14	*******
129	SNP_0607A07	113	0	2	107.14	*******
130	SNP_0609A07	57	16	41	2.61	-
131	SNP_0610A07	113	1	1	110.04	*******
132	SNP_0611A07	104	1	1	101.04	*******
133	SNP_0612A07	90	5	19	46.25	*******
134	SNP_0616A07	110	0	5	95.87	*******
135	SNP_0620A07	108	2	6	91.26	*******
136	SNP_0622A07	110	0	1	107.04	*******
137	L_057A07	91	6	13	58.5	*******
138	SNP_0630A07	92	3	19	48.01	*******
139	SNP_0637A07	53	12	37	2.84	*
140	SNP_0646A07	40	19	58	3.31	*
141	SNP_0649A07	57	1	54	0.08	-
142	SNP_0654A07	61	1	49	1.31	-
143	SNP_0660A07	89	3	21	42.04	*******
144	SNP_0662A07	89	7	13	56.63	*******
145	SNP_0663A07	71	15	29	17.64	*******
146	SNP_0666A07	107	5	2	101.15	*******
147	SNP_0672A07	98	2	12	67.24	*******
148	SNP_0676A07	91	3	19	47.13	*******
149	SNP_0677A07	5	7	86	72.1	*******
150	SNP_1046A07	50	19	48	0.04	-
151	SNP_0678A07	108	1	1	105.04	*******
152	SNP_0687A07	78	10	17	39.17	*******
153	SNP_0689A07	96	12	6	79.41	*******
154	SNP_0690A07	96	16	1	93.04	*******
155	SNP_0696A08	108	1	4	96.57	*******
156	SNP_0698A08	49	15	53	0.16	-
157	L_214A08	78	7	12	48.4	*******
158	L_005A08	88	6	10	62.08	*******
159	SNP_0714A08	36	0	37	0.01	-
160	SNP_0720A08	96	12	5	81.99	*******
161	SNP_0721A08	52	14	50	0.04	-
162	L_170A08	93	6	14	58.33	*******
163	SNP_0722A08	97	8	9	73.06	*******
164	SNP_1060A08	91	0	12	60.59	*******
165	SNP_0746A08	100	8	8	78.37	*******
166	SNP_0747A08	97	0	17	56.14	*******
167	SNP_0754A08	109	5	3	100.32	*******
168	L_206A08	63	19	24	17.48	*******

169	L_167A08	60	4	42	3.18	*
170	SNP_1062A08	37	18	61	5.88	**
171	SNP_0765A08	96	5	13	63.2	*******
172	SNP_0768A08	98	0	16	58.98	*******
173	SNP_0770A08	97	0	17	56.14	*******
174	SNP_0776A08	42	23	51	0.87	-
175	SNP_0778A08	13	10	94	61.32	*******
176	SNP_0779A08	58	19	40	3.31	*
177	SNP_0783A08	62	18	37	6.31	**
178	SNP_0786A08	89	2	23	38.89	*******
179	SNP_1041A08	41	9	66	5.84	**
180	L_139A08	96	2	11	67.52	*******
181	SNP_0790A08	7	4	106	86.73	*******
182	SNP_0795A09	49	14	36	1.99	-
183	SNP_0796A09	67	11	21	24.05	*******
184	SNP_0798A09	80	7	12	50.26	*******
185	SNP_0801A09	94	1	4	82.65	*******
186	SNP_1083A09	43	10	45	0.05	-
187	SNP_0808A09	87	0	12	56.82	*******
188	L_107A09	56	9	26	10.98	*****
189	SNP_0809A09	3	3	93	84.38	*******
190	SNP_0810A09	44	15	40	0.19	-
191	SNP_1071A09	94	3	2	88.17	*******
192	SNP_0812A09	5	0	84	70.12	*******
193	SNP_1049A09	9	17	19	3.57	*
194	L_130A09	86	1	11	57.99	*******
195	SNP_1044A09	47	17	35	1.76	-
196	SNP_0839A09	45	16	35	1.25	-
197	SNP_0842A09	79	1	19	36.73	*******
198	SNP_0843A09	82	0	17	42.68	*******
199	SNP_0844A09	93	0	5	79.02	*******
200	SNP_0857A09	81	1	17	41.8	*******
201	SNP_0861A09	95	0	4	83.65	*******
202	SNP_0862A09	95	0	4	83.65	*******
203	SNP_0865A09	41	17	41	0	-
204	L_065A09	59	5	20	19.25	*******
205	SNP_0867A09	87	6	6	70.55	*******
206	L_092A09	45	15	39	0.43	-
207	SNP_0868A09	60	20	3	51.57	*******
208	SNP_0879A09	90	5	4	78.68	*******
209	SNP_0882A09	40	16	43	0.11	-
210	SNP_0889A09	57	9	32	7.02	***
211	SNP_0896A09	88	0	11	59.89	*******
212	SNP_0898A09	68	5	22	23.51	*******

213	SNP_0902A09	58	12	29	9.67	****
214	L_039A09	46	9	30	3.37	*
215	SNP_0904A09	54	13	32	5.63	**
216	SNP_0906A09	60	11	28	11.64	*****
217	SNP_0907A09	62	11	25	15.74	*******
218	SNP_1038A09	37	11	51	2.23	-
219	SNP_0915A09	84	7	5	70.12	*******
220	SNP_0918A09	98	0	0	98	*******
221	SNP_0924A10	24	6	86	34.95	*******
222	SNP_1067A10	115	0	1	112.03	*******
223	SNP_1068A10	114	0	1	111.03	*******
224	SNP_0931A10	16	8	92	53.48	*******
225	L_033A10	88	5	19	44.5	*******
226	SNP_0939A10	94	13	10	67.85	*******
227	SNP_0951A10	36	36	31	0.37	-
228	SNP_0952A10	78	5	34	17.29	*******
229	SNP_0954A10	110	4	3	101.32	*******
230	SNP_0955A10	66	16	35	9.51	****
231	SNP_0959A10	50	8	41	0.8	-
232	SNP_0966A10	14	7	95	60.19	*******
233	SNP_0968A10	53	13	49	0.16	-
234	L_230A10	43	12	41	0.05	-
235	SNP_0970A10	46	16	55	0.8	-
236	SNP_0974A10	106	5	4	94.58	*******
237	SNP_0975A10	111	2	4	99.56	*******
238	SNP_0980A10	97	9	8	75.44	*******
239	L_050A10	63	12	30	11.71	*****
240	SNP_0987A10	101	7	9	76.95	*******
241	SNP_0989A10	63	0	6	47.09	*******
242	SNP_0996A10	110	5	1	107.04	*******
243	SNP_0998A10	48	18	51	0.09	-
244	SNP_0999A10	101	6	8	79.35	*******
245	SNP_1053	3	3	109	100.32	*******
246	SNP_1054	3	3	110	101.32	*******

Appendix E

R script for normal and multiple ANOVA marker trait association. The script was kindly provided by Ram Kumar and Dr. Ningwen Zhang:

```
rm(list=ls())

### geno = genotype data object ### column = marker, row = genotypes

geno<- read.table(file="genotype data file as txt file ", header = TRUE, row.names=1, sep="\t")

dim(geno)

geno[1:10,1:10]

mardata<- geno

### Log-ratio data

### logRatio = phenotype data ### column = phenotype, row = genotype

logRatio<- read.csv(file="phenotype data as csv excel file",header = TRUE, row.names =1)

dim(logRatio)

logRatio[1:10,1:6]

Pvalues<- matrix(NA,ncol=dim(mardata)[2],nrow=dim(logRatio)[2])

for(pheno in 1:dim(logRatio)[2]){

  for(marker in 1:ncol(mardata)){

genomodel<- lm(logRatio[,pheno] ~ as.factor(mardata[,marker]), na.action=na.omit)

Pvalues[pheno,marker] <- anova(genomodel)[1,5]

          }

        }


colnames(Pvalues) <- colnames(mardata)

rownames(Pvalues) <- colnames(logRatio)

head(Pvalues)

Pvalues<- t(Pvalues)

head(Pvalues)
```

```
LODvalue<- -1*log10(Pvalues)

head(LODvalue )

LODvalue[LODvalue[,5]>2,]

plot(seq(1:nrow(LODvalue)),LODvalue[,1],type="l")

par(mfrow=c(3,2))

for(i in 1:ncol(LODvalue)){

plot(seq(1:nrow(LODvalue)),LODvalue[,i],type="l",main=colnames(LODvalue)[i],xlab="marker",

                    ylab="-log10(Pvalue)")

abline(h=1.5,col="red")

abline(h=2,col="blue")

abline(h=3,col="green")

}

write.csv(Pvalues,file="name of the new file in csv excel format ")
```

```
##### Multiple testing
Qval<- matrix(NA,nrow=nrow(Pvalues),ncol=ncol(Pvalues))
for(i in 1:ncol(Pvalues)){
Qval[,i] <- p.adjust(Pvalues[,i], method = "BH")
}
head(Qval)
colnames(Qval) <- colnames(Pvalues)
rownames(Qval) <- rownames(Pvalues)
head(Qval)
LOD.qval<- -1*log10(Qval)
head(LOD.qval)
par(mfrow=c(3,2))
for(i in 1:ncol(LOD.qval)){
plot(seq(1:nrow(LOD.qval)),LOD.qval[,i],type="l",main=colnames(LOD.qval)[i],xlab="marker",
                ylab="-log10(Qvalue)")
abline(h=1.5,col="red")
abline(h=2,col="blue")
abline(h=3,col="green")
}
write.csv(LOD.qval, file="name of the new file in csv excel format")
```

Appendix F

Residuals plots greenhouse trial:

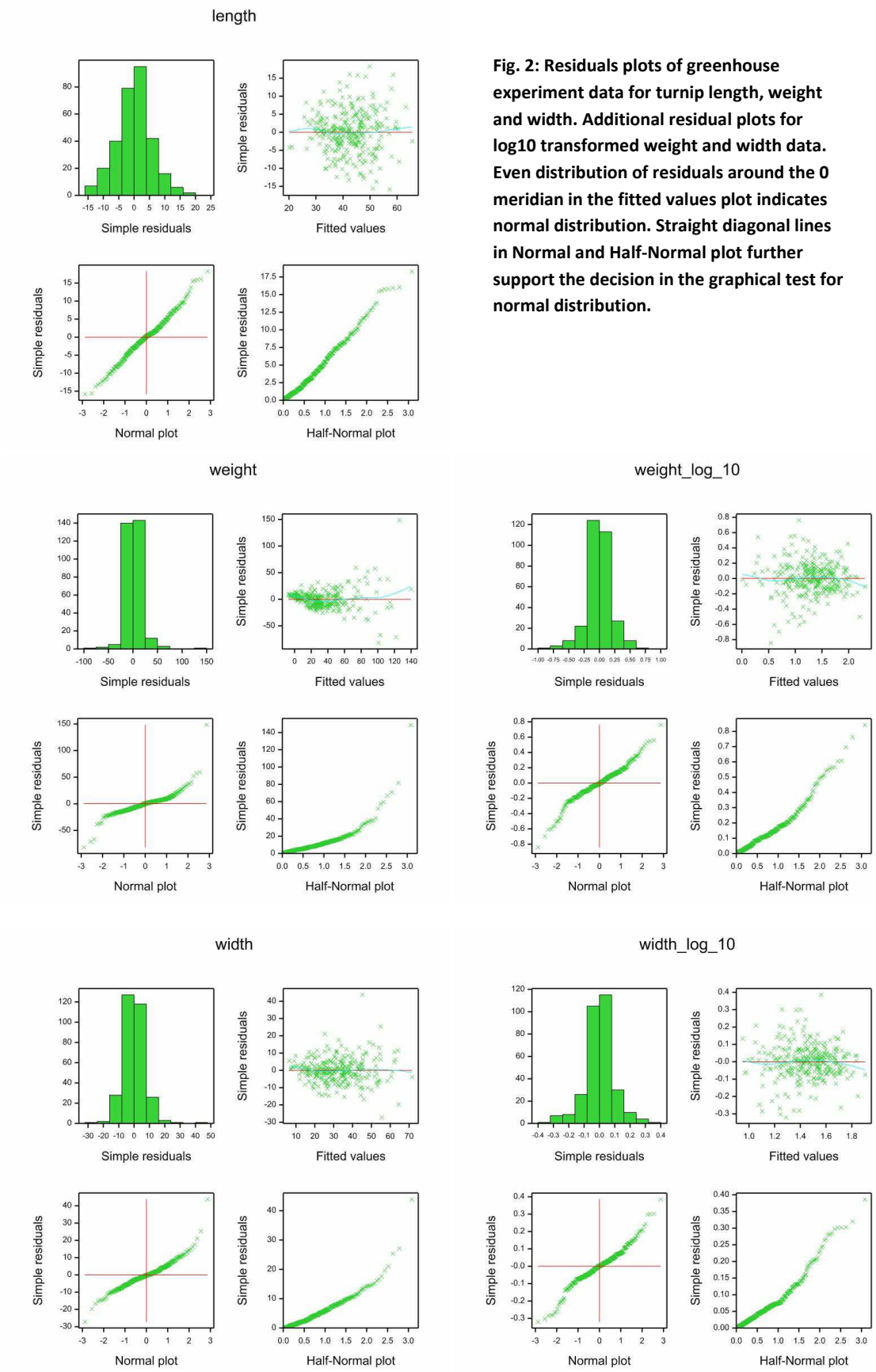

Fig. 2: Residuals plots of greenhouse experiment data for turnip length, weight and width. Additional residual plots for log10 transformed weight and width data. Even distribution of residuals around the 0 meridian in the fitted values plot indicates normal distribution. Straight diagonal lines in Normal and Half-Normal plot further support the decision in the graphical test for normal distribution.

Residuals plots field trial:

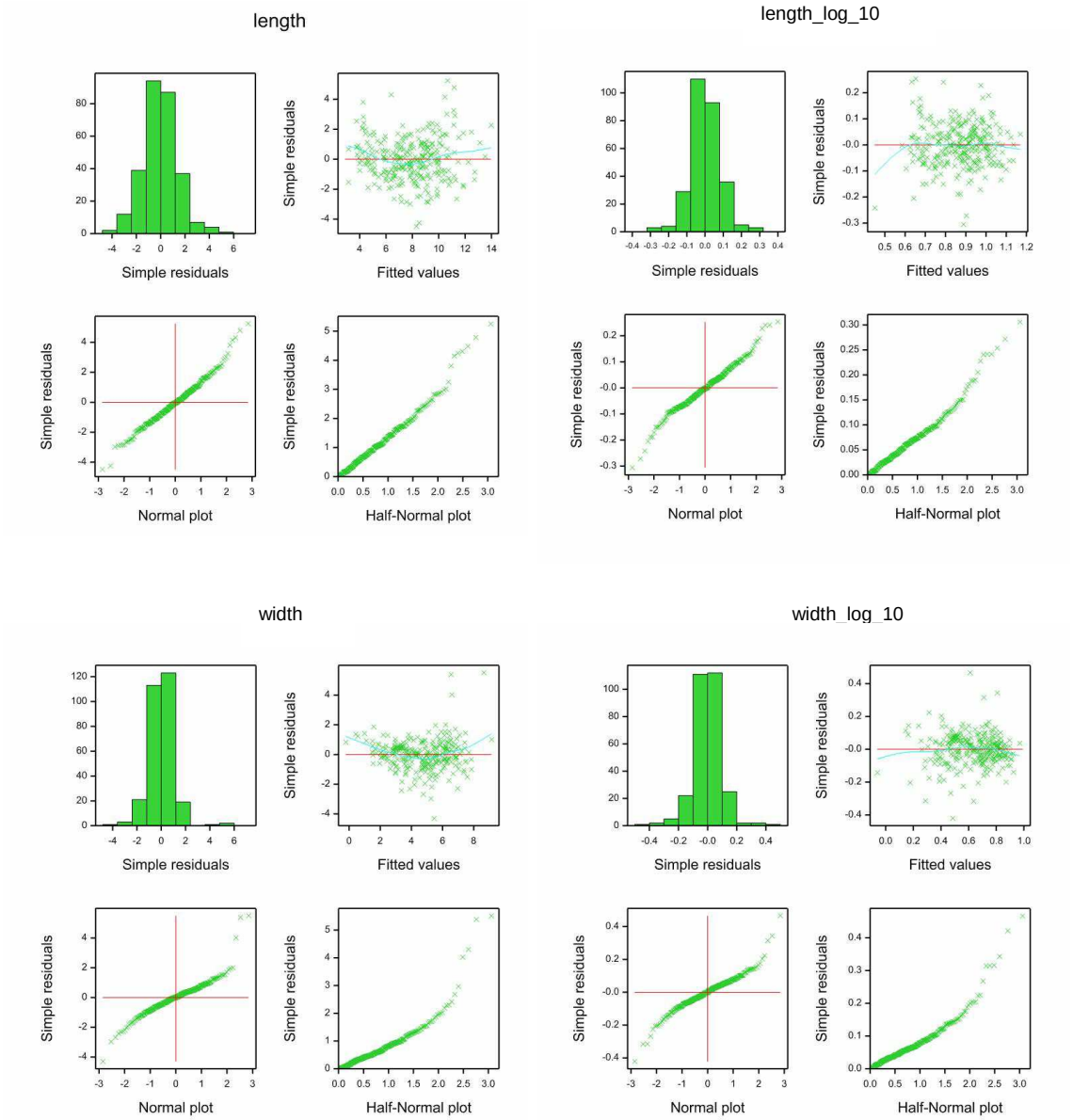

Fig.3: Residuals plots of field experiment data for turnip length and turnip width and log10 transformed data

Appendix G

ANOVA tables Greenhouse:

Tab. 23: ANOVA table greenhouse experiment for turnip length

Analysis of an unbalanced design using GenStat regression					
Variate: **length**					
Accumulated analysis of variance					
Change	d.f.	s.s.	m.s.	v.r.	F pr.
+ Block	2	2361.9	1180.95	23.7	<.001
+ Line	102	16293.57	159.74	3.21	<.001
Residual	202	10066.16	49.83		
Total	306	28721.63	93.86		

Minimum standard error of difference 5.764
Average standard error of difference 5.790
Maximum standard error of difference 7.059

Tab. 24: ANOVA table greenhouse experiment for weight log_{10} transformed

Analysis of an unbalanced design using GenStat regression					
Variate: **weight_log_10**					
Accumulated analysis of variance					
Change	d.f.	s.s.	m.s.	v.r.	F pr.
+ Block	2	8.18315	409.157	64.26	<.001
+ Line	102	42.00478	0.41181	6.47	<.001
Residual	202	12.86178	0.06367		
Total	306	63.04970	0.20604		

Minimum standard error of difference 0.2060
Average standard error of difference 0.2070
Maximum standard error of difference 0.2523

Tab. 25: ANOVA table greenhouse experiment for width log_{10} transformed

Analysis of an unbalanced design using GenStat regression					
Variate: **width_log_10**					
Accumulated analysis of variance					
Change	d.f.	s.s.	m.s.	v.r.	F pr.
+ Block	2	1.38967	0.69484	46.63	<.001
+ Line	102	8.65585	0.08486	5.69	<.001
Residual	202	3.01011	0.0149		
Total	306	13.05563	0.04267		

Tab. 26: ANOVA table for length ANOVA between blocks over all lines greenhouse experiment

Analysis of an unbalanced design using GenStat regression					
Variate: **length**					
Accumulated analysis of variance					
Change	d.f.	s.s.	m.s.	v.r.	F pr.
+ Block	2	2361.9	1180.95	13.62	<.001
Residual	304	26359.73	86.71		
Total	306	28721.63	93.86		
Predictions from regression model					
Response variate: length					
Block	Prediction				
B1	46.3				
B2	40.93				
B3	39.98				
Minimum standard error of difference				1.298	
Average standard error of difference				1.302	
Maximum standard error of difference				1.304	

Tab. 27: ANOVA table weight log_{10} transformed, ANOVA between blocks over all line in greenhouse experiment, additional regression prediction per block and mean value per block of non transformed data

Analysis of an unbalanced design using GenStat regression					
Variate: **weight_log_10**					
Accumulated analysis of variance					
Change	d.f.	s.s.	m.s.	v.r.	F pr.
+ Block	2	81.831	40.916	22.67	<.001
Residual	304	548.666	0.1805		
Total	306	630.497	0.206		
Predictions from regression model					
Response variate: weight_log_10					
Block	Prediction				
B1	1.535				
B2	1.227				
B3	1.158				
Minimum standard error of difference			0.0592		
Average standard error of difference			0.05939		
Maximum standard error of difference			0.05949		
untransformed means weight					
Block	mean of weight				
B1	48.45				
B2	24.83				
B3	22.81				

Tab. 28: ANOVA table width log_{10} transformed, ANOVA between blocks over all line greenhouse experiment, additional regression prediction per block and mean value per block of non transformed data

Analysis of an unbalanced design using GenStat regression					
Variate: **width_log_10**					
Accumulated analysis of variance					
Change	d.f.	s.s.	m.s.	v.r.	F pr.
+ Block	2	138.967	0.69484	18.11	<.001
Residual	304	1.166.596	0.03837		
Total	306	1.305.563	0.04267		
Predictions from regression model					
Response variate: width_log_10					
Block	Prediction				
B1	1.551				
B2	1.423				
B3	1.396				
Minimum standard error of difference			0.0273		
Average standard error of difference			0.02739		
Maximum standard error of difference			0.02743		
untransformed means width					
Block	mean width				
B1	38.46				
B2	28.85				
B3	27.95				

Field experiment:

Tab. 29: ANOVA table length log_{10} transformed, ANOVA between lines in field experiment

Analysis of an unbalanced design using GenStat regression					
Variate: **length_log10**					
Accumulated analysis of variance					
Change	d.f.	s.s.	m.s.	v.r.	F pr.
+ Block	2	289.697	144.848	139.47	<.001
+ Line	99	162.544	0.01642	1.58	0.004
Residual	181	187.982	0.01039		
Total	282	640.222	0.0227		

Minimum standard error of difference	0.08321
Average standard error of difference	0.08685
Maximum standard error of difference	0.12548

Tab. 30: ANOVA table width log10 transformed, ANOVA between lines in field experiment

Analysis of an unbalanced design using GenStat regression					
Variate: **width_log10**					
Accumulated analysis of variance					
Change	d.f.	s.s.	m.s.	v.r.	F pr.
+ Block	2	501.574	250.787	174.09	<.001
+ Line	99	379.005	0.03828	2.66	<.001
Residual	181	260.743	0.01441		
Total	282	1.141.322	0.04047		

Minimum standard error of difference	0.0980
Average standard error of difference	0.1023
Maximum standard error of difference	0.1478

Tab. 31: ANOVA table length log10 transformed, ANOVA between blocks over all lines field experiment, additional regression prediction per block and mean value per block of non transformed data

Analysis of an unbalanced design using GenStat regression					
Variate: **length_log_10**					
Accumulated analysis of variance					
Change	d.f.	s.s.	m.s.	v.r.	F pr.
+ Block	2	289.697	144.848	115.7	<.001
Residual	280	350.526	0		
Total	282	640.222	0		
Predictions from regression model					
Response variate: **length_log_10**					
Block	Prediction				
B1	0.7321				
B2	0.9105				
B3	0.9716				
Minimum standard error of difference			0.01619		
Average standard error of difference			0.0163		
Maximum standard error of difference			0.01641		
untransformed means length					
Block	mean length				
B1	5.608				
B2	8.331				
B3	9.691				

Tab. 32: ANOVA table width log10 transformed, ANOVA between blocks over all lines field experiment, additional regression prediction per block and mean value per block of non transformed data

Analysis of an unbalanced design using GenStat regression					
Variate: **width_log_10**					
Accumulated analysis of variance					
Change	d.f.	s.s.	m.s.	v.r.	F pr.
+ Block	2	501.574	250.787	109.76	<.001
Residual	280	639.748	0		
Total	282	1.141.322	0		
Predictions from regression model					
Response variate: **width_log_10**					
Block	Prediction				
B1	0.4481				
B2	0.674				
B3	0.7667				
Minimum standard error of difference			0.02188		
Average standard error of difference			0.02201		
Maximum standard error of difference			0.02217		
untransformed means width					
Block	mean width				
B1	2.996				
B2	4.928				
B3	6.160				

Tab. 33: ANOVA of Block 2 and 3 in greenhouse experiment for turnip length, width and weight

Analysis of an unbalanced design using GenStat regression					
Accumulated analysis of variance Variate: **length**					
Change	d.f.	s.s.	m.s.	v.r.	F pr.
+ block	1	45.89	45.89	0.58	0.447
Residual	204	16106.91	78.96		
Total	205	16152.8	78.79		
Accumulated analysis of variance Variate: **weight**					
Change	d.f.	s.s.	m.s.	v.r.	F pr.
+ block	1	211	211	0.37	0.544
Residual	204	116305.8	570.1		
Total	205	116516.8	568.4		
Accumulated analysis of variance Variate: **width**					
Change	d.f.	s.s.	m.s.	v.r.	F pr.
+ block	1	41.5	41.5	0.25	0.614
Residual	204	33252.2	163		
Total	205	33293.8	162.4		

Appendix H

Tab. 34: Genes corresponding to turnip QTL markers

Marker	Chromsome	Trait	SNP position	left border	right border	Gene				
greenhouse experiment										
SNP_1074	A01	weight	374356	369356	379356	Bra011821	Bra011820			
SNP_1056	A01	weight	15365196	15360196	15370196	Bra029965	Bra029966			
SNP_0898	A09	weight, length	30938434	30933434	30943434	Bra007799	Bra007800	Bra007801		
field experiment										
SNP_0191	A02	width	20728211	20723211	20733211	Bra023204	Bra023203			
SNP_0205	A02	width	23890520	23885520	23895520	Bra020660	Bra020659	Bra020658	Bra020657	
SNP_1065	A03	width	27864125	27859125	27869125	Bra024068	Bra024067			
SNP_0335	A04	width	7089683	7084683	7094683	Bra028246	Bra028247			
SNP_0364	A04	width	17743027	17738027	17748027	Bra016891	Bra016890	Bra016889	Bra016888	Bra016887
SNP_0602	A06	width	24995754	24990754	25000754	Bra025071	Bra025072	Bra025073		
SNP_0607	A07	width	286182	281182	291182	no annotation				
SNP_0610	A07	width	797437	792437	802437	Bra039016	Bra039015			
SNP_0612	A07	width	1316612	1311612	1321612	Bra038817	Bra038816			
SNP_0622	A07	width	5385095	5380095	5390095	Bra014903				
SNP_0678	A07	width	20498940	20493940	20503940	Bra015855	Bra015854	Bra015853		
SNP_0720	A08	width	5738432	5733432	5743432	Bra034882	Bra034881			
SNP_0747	A08	width	10090362	10085362	10095362	Bra021067				
SNP_0765	A08	width	14098893	14093893	14103893	Bra010360	Bra010361			
SNP_0768	A08	width	14692670	14687670	14697670	Bra010468	Bra010469			
SNP_0770	A08	width	14966189	14961189	14971189	Bra010519	Bra010520	Bra010521		
SNP_0786	A08	width	18758475	18753475	18763475	Bra016495	Bra016496	Bra016497		
SNP_0108	A01	length	25004591	24999591	25009591	Bra021490				

Tab. 35: Gene annotations in QTL intervalls

Marker	Gene name	Gene annotation
SNP_1074	Bra011821	electron carrier, iron ion binding protein
SNP_1074	Bra011820	DNA heat shock protein
SNP_1056	Bra029965	unknown protein
SNP_1056	Bra029966	branched-chain-amino-acid transaminase
SNP_0898	Bra007799	no annotation
SNP_0898	Bra007800	(ENTH) domain containing protein
SNP_0898	Bra007801	zinc finger protein-related
SNP_0191	Bra023204	actin protein
SNP_0191	Bra023203	unknown function
SNP_0205	Bra020660	"ATGEX2, GEX2; GEX2 (GAMETE EXPRESSED 2)"
SNP_0205	Bra020659	unknown protein
SNP_0205	Bra020658	pectinesterase family protein
SNP_0205	Bra020657	unknown protein
SNP_1065	Bra024068	"defense-related protein, putative"
SNP_1065	Bra024067	calmodulin binding / cyclic nucleotide binding
SNP_0335	Bra028246	F-box family protein
SNP_0335	Bra028247	F-box family protein
SNP_0364	Bra016891	promotor binding transcription factor
SNP_0364	Bra016890	mitochondrial import inner membrane translocase subunit
SNP_0364	Bra016889	rhodanese-like domain-containing protein
SNP_0364	Bra016888	electron carrier, iron ion binding protein
SNP_0364	Bra016887	calmodulin-binding protein
SNP_0602	Bra025071	transcription factor
SNP_0602	Bra025072	unknown protein
SNP_0602	Bra025073	unknown protein
SNP_0610	Bra039016	60S ribosomal protein L31 (RPL31A)
SNP_0610	Bra039015	"leucine-rich repeat protein kinase, putative"
SNP_0612	Bra038817	signal recognition particle binding
SNP_0612	Bra038816	diacylglycerol kinase
SNP_0622	Bra014903	unknown protein
SNP_0678	Bra015855	"oxidoreductase, acting on NADH or NADPH"
SNP_0678	Bra015854	"isoflavone reductase, putative"
SNP_0678	Bra015853	isoflavone reductase, putative"
SNP_0720	Bra034882	short-chain dehydrogenase/reductase (SDR) family protein
SNP_0720	Bra034881	unknown protein
SNP_0747	Bra021067	unknown function
SNP_0765	Bra010360	RabGAP/TBC domain-containing protein
SNP_0765	Bra010361	kinase/ protein serine/threonine kinase"
SNP_0768	Bra010468	linked to short hypocotyl phenotype in A. thaliana
SNP_0768	Bra010469	oxidoreductase, 2OG-Fe(II) oxygenase family protein"
SNP_0770	Bra010519	Rho-GTPase-activating protein-related
SNP_0770	Bra010520	ATP-dependent helicase/ helicase/ nucleic acid binding
SNP_0770	Bra010521	unknown function

SNP_0786	Bra016495	purine transmembrane transporter
SNP_0786	Bra016496	jacalin lectin family protein
SNP_0786	Bra016497	DNA binding / transcription factor
SNP_0108	Bra021490	oxygen binding / sterol 14-demethylase